Dynamic Brain Connectivity Alternation Detection via Matrix-variate Differential Network Model

Jiadong Ji ¹, Yong He ^{1,*} and Lei Xie ²

¹ School of Statistics, Shandong University of Finance and Economics, Jinan, 250014, China

²Department of Computer Science, Hunter College, The City University of New York, New York, 10065, USA.

*To whom correspondence should be addressed.

Abstract

Motivation: Nowadays brain connectivity analysis has attracted tremendous attention and has been at the foreground of neuroscience research. Brain functional connectivity reveals the synchronization of brain systems through correlations in neurophysiological measures of brain activity. Growing evidence now suggests that the brain connectivity network experiences alternations with the presence of numerous neurological disorders, thus differential brain network analysis may provides new insights into disease pathologies. For the matrix-valued data in brain connectivity analysis, existing graphical model estimation methods assume a vector normal distribution that in essence requires the columns of the matrix data to be independent. It is obviously not true, they have limited applications. Among the few solutions on graphical model estimation under a matrix normal distribution, none of them tackle the estimation of differential graphs across different populations. This motivates us to consider the differential network for matrix-variate data to detect the brain connectivity alternation.

Results: The primary interest is to detect spatial locations where the connectivity, in terms of the spatial partial correlation, differ across the two groups. To detect the brain connectivity alternation, we innovatively propose a Matrix-Variate Differential Network (MVDN) model. MVDN assumes that the matrix-variate data follows a matrix-normal distribution. We exploit the D-trace loss function and a Lasso-type penalty to directly estimate the spatial differential partial correlation matrix where the temporal information is fully excavated. We propose an ADMM algorithm for the Lasso penalized D-trace loss optimization problem. We investigate theoretical properties of the estimator. We show that under mild and regular conditions, the proposed method can identify all differential edges accurately with probability tending to 1 in high-dimensional setting where dimensions of matrix-valued data p, q and sample size n are all allowed to go to infinity. Simulation studies demonstrate that MVDN provides more accurate differential network estimation than that achieved by other state-of-the-art methods. We apply MVDN to Electroencephalography (EEG) dataset, which consists of 77 alcoholic individuals and 45 controls. The hub genes and differential interaction patterns identified are consistent with existing experimental studies.

Contact: heyong@sdufe.edu.cn

Supplementary information: Supplementary data are available online.

1 Introduction

Nowadays brain connectivity analysis has attracted tremendous attention and has been at the foreground of neuroscience research. Brain functional

connectivity reveals the synchronization of brain systems through correlations in neurophysiological measures of brain activity. Varoquaux and Craddock (2013) pointed out that when measured during the resting state, it maps the intrinsic functional architecture of the brain. Growing evidence now suggests that the brain connectivity network experiences alternations

with the presence of numerous neurological disorders, including Alzheimer's disease, attention deficit hyperactivity disorder, autism spectrum disorder (Fox and Greicius, 2010; Daliri and Behroozi, 2013). In this paper, we focus on the problem of comparing brain functional connectivity patterns across the diseased versus the healthy control, which helps to provide new insights into disease pathologies.

Graphical model has been widely adopted to capture the interactions (connectivity) between variables, especially the Gaussian Graphical Model (GGM). Under GGM, two variables interact if and only if the corresponding entry of the precision matrix is nonzero, where the precision matrix is defined as the inverse of the covariance matrix. In recent years, numerous research on precision matrix estimation has appeared in the high dimensional setting, for example, Meinshausen and Bühlmann (2006); Yuan and Lin (2007); Cai *et al.* (2011); Zhang and Zou (2014). Some researchers studied a more flexible Gaussian copula graphical model which relax the restrictive Gaussian assumption of GGM in real application, for instance, Liu *et al.* (2009, 2012); He *et al.* (2017a). This research area is very active, as a result, this list of references here is illustrative rather than comprehensive.

In the brain connectivity study, it is often the case that the functional MRI data are collected for multiple subjects from the disease group and normal control. Compared to focusing on the brain region network from one particular group, it is of greater interest to investigate how the network of connected brain regions alters from disease group to the control group. It provides deeper insights on an underlying biological process. In other word, it is more appropriate to model the differential network of brain regions between two groups. Recently, differential network modelling has emerged as an important tool to analyze a set of changes in graph structure between two groups, and is typically modelled as the difference of the precision matrices for each group, for example, Zhao *et al.* (2014); Xia *et al.* (2015); Yuan *et al.* (2015) and so on. All these works assume that the vector of interacting variables follows a normal distribution. However, the observed functional MRI data in brain functional connectivity study is in the matrix form, whose row variables represents the brain regions while the column variables represents time points. Typically, the number of brain regions is of the order 10^2 and the number of time points is around 150 to 200. In fact, the matrix-valued data have been emerging rapidly in the current big data era. To our knowledge, for the type of matrix-variate observation data, there has not been any work on analyzing the differential network, which motivates us to propose a corresponding differential network framework.

For the matrix-valued data, matrix normal distribution is becoming increasingly popular in modelling such observations. In the matrix normal distribution framework, the brain connectivity network analysis can be viewed as a precision matrix inference problem. Specifically, let $\mathbf{X}_{p \times q}$ be the spatial-temporal matrix data from an image modality. We say $\mathbf{X}_{p \times q}$ follows a matrix normal distribution with the Kronecker product covariance structure $\Sigma = \Sigma_T \otimes \Sigma_S$, which we write as

$$\mathbf{X}_{p \times q} \sim N_{p,q}(\mathbf{M}_{p \times q}, \Sigma_T \otimes \Sigma_S),$$

if and only if $\text{Vec}(\mathbf{X}_{p \times q})$ follows a multivariate normal distribution with mean $\text{Vec}(\mathbf{M}_{p \times q})$ and covariance $\Sigma = \Sigma_T \otimes \Sigma_S$, where $\text{Vec}(\mathbf{X}_{p \times q})$ is formed by stacking the columns of $\text{Vec}(\mathbf{X}_{p \times q})$ into a vector in $\mathbb{R}^{pq}$, $\Sigma_S \in \mathbb{R}^{p \times p}$ and $\Sigma_T \in \mathbb{R}^{q \times q}$ denote the covariance matrices of p spatial locations and q times points, respectively. Correspondingly, we have

$$\text{Cov}^{-1}(\text{Vec}(\mathbf{X}_{p \times q})) = \Sigma_T^{-1} \otimes \Sigma_S^{-1} = \Omega_T \otimes \Omega_S,$$

where $\Omega_S \in \mathbb{R}^{p \times p}$ and $\Omega_T \in \mathbb{R}^{q \times q}$ denote the spatial and temporal precision matrix, respectively. Note that Σ_S and Σ_T are only identifiable up to a scaled factor, as $c\Sigma_T \otimes \frac{1}{c}\Sigma_S = \Sigma_T \otimes \Sigma_S$ for any $c > 0$. In fact,

in the brain connected network analysis, the partial correlation is a commonly adopted correlation measure (Peng *et al.*, 2009; Zhu and Li, 2018). Besides, the primary interest in brain connectivity analysis is to infer the the connectivity network characterized by the spatial precision matrix Ω_S while the temporal precision matrix Ω_T is of little interest and thus can be viewed as a nuisance parameter. In a word, under the matrix norm framework a region-by-region spatial partial correlation matrix characterizes the brain connectivity graph, in which nodes represent brain regions, and links measure conditional dependence between the brain regions. Brain connectivity analysis is equivalently transformed into the estimation of spatial partial correlation matrix. We remark here that the matrix normal distribution framework has been widely adopted in real application and is scientifically plausible in neuroimaging analysis. (see, for example, Yin and Li, 2012; Leng and Tang, 2012; Zhou, 2014; Xia and Li, 2017; Zhu and Li, 2018; Xia and Li, 2018)

None of existing graphical models under a matrix normal distribution tackle the differential network between the disease group and the control group directly, but instead focus on only a single graph (Zhu and Li (2018) focused on joint estimation of multiple graphs). In this paper, we for the first time propose a differential network model for dynamic matrix-variate data paralleled to static vector-variate data. In specific, assume $\mathbf{X}_{p \times q} \sim N_{p,q}(\mathbf{M}_{p \times q}, \Sigma_{T_X} \otimes \Sigma_{S_X})$ be the spatial-temporal matrix data from the disease group and $\mathbf{Y}_{p \times q} \sim N_{p,q}(\mathbf{M}_{p \times q}, \Sigma_{T_Y} \otimes \Sigma_{S_Y})$ from the control group. The primary interest in brain connectivity alternation detection is to infer the the differential connectivity network characterized by the difference of spatial precision matrices from two groups, namely, $\Omega_{S_X} - \Omega_{S_Y}$. For the matrix-valued data, directly applying the existing differential network estimation method, which assumes the row variables follow a vector normal distribution, would ignore the dependence between the columns of the matrix data and result in uncorrect conclusion. For example, the columns in functional magnetic resonance imaging (fMRI) study corresponds to times series of repeatedly measured brain activity and are highly correlated. fMRI is one of the mainstream imaging modalities to study brain functional connectivity nowadays. As an alternative, whitening can reduce the between-column correlation. However, Zhu and Li (2018) showed that their joint estimation method substantially outperforms the state of the art vector-normal-based differential network estimation methods such as Zhao *et al.* (2014) and Yuan *et al.* (2015), facilitated by whitening. In the simulation study, we show that the proposed method still outperforms Zhu and Li (2018)'s joint estimation methods, which further indicate that methods facilitated by whitening are inferior to the proposed method.

The main contribution of the current work lies in the following aspects. Firstly, we innovatively propose a Matrix-Variate Differential Network (MVDN) model which is particularly useful in modelling connectivity alternation for matrix-variate data such as in brain connectivity analysis. Secondly, we propose a computationally efficient algorithm for estimating the MVDN, which directly estimate the spatial partial correlation matrix difference without attempting to estimate the partial correlation matrices separately. Thirdly, the proposed approach fully excavates the temporal information without assuming the columns of the matrix data to be independent compared with the vector-normal graphical model methods. At last, we show that the proposed method can identify all differential edges accurately with probability tending to 1 under mild and regular conditions as p, q and sample size n go to infinity.

We denote some notations which are adopted throughout of the paper. For any vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$, let $|\boldsymbol{\mu}|_0 = \sum_{i=1}^d I\{\mu_i \neq 0\}$, $|\boldsymbol{\mu}|_1 = \sum_{i=1}^d |\mu_i|$, $|\boldsymbol{\mu}|_2 = (\sum_{i=1}^d \mu_i^2)^{1/2}$ and $|\boldsymbol{\mu}|_\infty = \max_i |\mu_i|$. Let $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{d \times d}$, define $\|\mathbf{A}\|_0 = \sum_{i=1}^d \sum_{j=1}^d I\{a_{ij} \neq 0\}$, $\|\mathbf{A}\|_F = (\sum_{i,j} a_{ij}^2)^{1/2}$, $\|\mathbf{A}\|_\infty = \max_{i,j} |a_{ij}|$ and $\|\mathbf{A}\|_1 = \sum_{i=1}^d \sum_{j=1}^d |a_{ij}|$. We use $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to denote the smallest and largest eigenvalues of $\mathbf{A}$ respectively, $\kappa(\mathbf{A})$ denote the conditional

number of $\mathbf{A}$ and denote the trace of $\mathbf{A}$ as $\text{Tr}(\mathbf{A})$ and $\det(\mathbf{A})$ be the determinant of $\mathbf{A}$. The notation $\otimes$ represents Kronecker product. For a set $\mathcal{H}$, denote by $|\mathcal{H}|$ the cardinality of $\mathcal{H}$. For a real number x , denote by $\lfloor x \rfloor$ the largest integer smaller than or equal to x and define $p \vee q = \max(p, q)$.

The remainder of the article is organized as follows. In Section 2, we present our matrix-variate differential network modeling framework and the ADMM algorithm for the Lasso penalized D-trace loss optimization problem. We also introduce the simulation designs and real data sets in this section. In section 3, we provide the theoretical results, with the detailed assumption and proofs delegated to the supplementary material. We also present the simulation results and the real data application results in this section. At last, we give a brief discussion on possible future directions in Section 4.

2 Methods

Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{p \times q}$ be denote the spatial-temporal matrices of the diseased and the healthy control groups. We assume $\mathbf{X}, \mathbf{Y}$ follows a matrix normal distribution with the Kronecker product covariance structure, and the mean matrices, without loss of generality, are assumed to be zero. In other words, $\mathbf{X} \sim N_{p,q}(\mathbf{0}_{p \times q}, \Sigma_{T_X} \otimes \Sigma_{S_X})$, and $\mathbf{Y} \sim N_{p,q}(\mathbf{0}_{p \times q}, \Sigma_{T_Y} \otimes \Sigma_{S_Y})$. Under the matrix normal framework, let

$$\Omega_{S_X} = \Sigma_{S_X}^{-1} \text{ and } \mathbf{R}_{S_X} = \mathbf{D}_{S_X}^{-1/2} \Omega_{S_X} \mathbf{D}_{S_X}^{-1/2},$$

where $\mathbf{R}_{S_X}$ is the partial correlation matrix of the spatial locations in diseased group and $\mathbf{D}_{S_X}$ is the diagonal matrix of Ω_{S_X} . Similarly, for the healthy control group, we can define Ω_{S_Y} and $\mathbf{R}_{S_Y}$ parallelly. The primary interest of this paper is to detect spatial locations where the connectivity, in terms of the spatial partial correlation, differ across the two groups. In this paper, we define the differential network between two groups as the difference between the two spatial partial correlation matrices, denoted by $\Delta = \mathbf{R}_{S_Y} - \mathbf{R}_{S_X}$. The $\Sigma_{T_X}, \Sigma_{T_Y}$ are viewed as nuisance parameters.

Let $\{\mathbf{X}_1, \dots, \mathbf{X}_{n_1}\}$ and $\{\mathbf{Y}_1, \dots, \mathbf{Y}_{n_2}\}$, each a matrix with dimension $p \times q$, be two sets of i.i.d. random samples from the independent matrix normal populations $\mathbf{X}$ and $\mathbf{Y}$, respectively. Motivated by the D-trace loss function in Yuan *et al.* (2015), we propose to estimate Δ by minimizing the following loss function:

$$L(\Delta, \Gamma_{S_X}, \Gamma_{S_Y}) = \frac{1}{4} (\text{Tr}(\Gamma_{S_X} \Delta \Gamma_{S_Y} \Delta) + \text{Tr}(\Gamma_{S_Y} \Delta \Gamma_{S_X} \Delta)) - \text{Tr}(\Delta(\Gamma_{S_X} - \Gamma_{S_Y})),$$

where Γ_{S_X} and Γ_{S_Y} denote the correlation versions of Σ_{S_X} and Σ_{S_Y} , respectively. It can be shown that the Hessian matrix with respect to Δ of the loss function is $(\Gamma_{S_X} \otimes \Gamma_{S_X} + \Gamma_{S_Y} \otimes \Gamma_{S_Y})/2$, which indicates that the loss function $L(\Delta, \Gamma_{S_X}, \Gamma_{S_Y})$ is convex with respect to Δ and has a unique minimizer at $\Delta = \mathbf{R}_{S_Y} - \mathbf{R}_{S_X}$. In brain connectivity alternation detection analysis, it is often the case that the altered connectivity in the spatial networks of two groups is far less compared with the dimensionality, which motivates us to add a penalty to the D-trace loss function. Generally, we can add a decomposable non-convex penalty function which has the form $P_\lambda(\delta) = \sum_{j,k} p_\lambda(\Delta_{j,k})$ such as smoothly clipped absolute deviation (SCAD) penalty. For simplicity, we consider a simple lasso penalty here. Finally Δ is estimated by solving the following optimization problem:

$$\hat{\Delta} = \arg \min_{\Delta} \left\{ L(\Delta, \hat{\Gamma}_{S_X}, \hat{\Gamma}_{S_Y}) + \lambda \|\Delta\|_1 \right\}, \quad (1)$$

where λ is a tuning parameter and $\hat{\Gamma}_{S_X}, \hat{\Gamma}_{S_Y}$ are sample correlation matrices defined as

$$\begin{aligned} \hat{\Gamma}_{S_X} &= \text{DiagScale} \left\{ \sum_{i=1}^{n_1} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top \right\}, \\ \hat{\Gamma}_{S_Y} &= \text{DiagScale} \left\{ \sum_{i=1}^{n_2} (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^\top \right\}, \end{aligned} \quad (2)$$

where for a square matrix $\mathbf{C}$, $\text{DiagScale}(\mathbf{C}) = \text{diag}(\mathbf{C})^{-\frac{1}{2}} \mathbf{C} \text{diag}(\mathbf{C})^{-\frac{1}{2}}$. Note that the sample correlation matrices in (2) are based on the matrix-valued observations. The optimization problem in (1) can be solved by the Alternating Direction method of Multipliers (ADMM) algorithm. To present the detailed procedure, we first introduce two multivariate functions $G(\cdot, \cdot, \cdot, \cdot)$ and $H(\cdot, \cdot)$. For symmetric $p \times p$ matrices $\mathbf{A}$ and $\mathbf{B}$, let $\mathbf{A} = \mathbf{U}_A \Lambda_A \mathbf{U}_A^\top$ and $\mathbf{B} = \mathbf{U}_B \Lambda_B \mathbf{U}_B^\top$ be the corresponding eigenvalue decompositions respectively, where $\Lambda_A = \text{diag}(\lambda_1^A, \dots, \lambda_p^A)$ and $\Lambda_B = \text{diag}(\lambda_1^B, \dots, \lambda_p^B)$. Define function $G(\cdot, \cdot, \cdot, \cdot)$ as

$$G(\mathbf{A}, \mathbf{B}, \mathbf{C}, \gamma) = \mathbf{U}_A \left\{ \mathbf{D} \circ (\mathbf{U}_A^\top \mathbf{C} \mathbf{U}_B^\top) \right\} \mathbf{U}_B,$$

where $\mathbf{D} = (D_{ij})$ with $D_{i,j} = (\lambda_j^A \lambda_i^B + \gamma)^{-1}$ and $\circ$ denotes the Hadamard product of two matrices. For a positive λ , define a $\mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{p \times p}$ map $H(\cdot, \cdot)$ as follows:

$$H(\mathbf{A}, \lambda)_{i,j} = \begin{cases} A_{ij} - \lambda, & A_{ij} > \lambda, \\ A_{ij} + \lambda, & A_{ij} < -\lambda, \\ 0, & -\lambda \leq A_{ij} \leq \lambda. \end{cases}$$

In detail, we summarize the procedure in Algorithm 1.

Algorithm 1 ADMM Algorithm for Lasso Penalized D-trace Loss Function

Input: $\mathcal{D} = \{\mathbf{X}_i, i = 1, \dots, n_1, \mathbf{Y}_i, i = 1, \dots, n_2\}$

Output: Estimate of the precision matrix difference $\hat{\Delta}$

- 1: Initialization: $k = 0$; Set $\Delta_1^0, \Delta_2^0, \Delta_3^0 = \{\text{diag}(\hat{\Gamma}_{S_Y}) + \mathbf{I}\}^{-1} - \{\text{diag}(\hat{\Gamma}_{S_X}) + \mathbf{I}\}^{-1}$; Set $\Lambda_1^0, \Lambda_2^0, \Lambda_3^0 = \mathbf{0}$.
 - 2: Given $\Delta_1^k, \Delta_2^k, \Delta_3^k, \Lambda_1^k, \Lambda_2^k$ and Λ_3^k at the k -th step, update:
 - 3: $\Delta_1^{k+1} = G(\hat{\Gamma}_{S_X}, \hat{\Gamma}_{S_Y}, 2\rho\Delta_3^k + 2\rho\Delta_2^k + \hat{\Gamma}_{S_X} - \hat{\Gamma}_{S_Y} + 2\Lambda_1^k - 2\Lambda_3^k, 4\rho)$;
 - 4: $\Delta_2^{k+1} = G(\hat{\Gamma}_{S_Y}, \hat{\Gamma}_{S_X}, 2\rho\Delta_3^k + 2\rho\Delta_1^k + \hat{\Gamma}_{S_X} - \hat{\Gamma}_{S_Y} + 2\Lambda_2^k - 2\Lambda_3^k, 4\rho)$;
 - 5: $\Delta_3^{k+1} = H((\rho\Delta_1^{k+1} + \rho\Delta_2^{k+1} - \Lambda_1^k + \Lambda_2^k)/(2\rho), \lambda/2\rho)$;
 - 6: $\Lambda_1^{k+1} = \Lambda_1^k + \rho(\Delta_3^{k+1} - \Delta_1^{k+1})$;
 - 7: $\Lambda_2^{k+1} = \Lambda_2^k + \rho(\Delta_3^{k+1} - \Delta_2^{k+1})$;
 - 8: $\Lambda_3^{k+1} = \Lambda_3^k + \rho(\Delta_1^{k+1} - \Delta_2^{k+1})$;
 - 9: Repeat steps 2-8 until convergence. Finally output the Δ_3^{k+1} as the estimate of the precision matrix difference.
-

In the simulation study and real data analysis, we set the step size $\rho = 1$ and terminate the algorithm if in the $k + 1$ -th step, we have

$$\|\Delta_j^{k+1} - \Delta_j^k\|_F < 10^{-3} \max\{1, \|\Delta_j^k\|_F, \|\Delta_j^{k+1}\|_F\}, \quad j = 1, 2, 3.$$

The tuning parameter λ is selected by minimizing the Bayesian information criterion, i.e., λ is chosen to minimize

$$(n_1 + n_2)L(\lambda) + \log(n_1 + n_2)\|\Delta\|_0,$$

where $L(\lambda)$ represents the loss function based on either L_∞ or L_F norm which are defined by

$$L_\infty(\lambda_n) = \left\| \frac{1}{2} (\hat{\Gamma}_{S_X} \hat{\Delta}(\lambda) \hat{\Gamma}_{S_Y} + \hat{\Gamma}_{S_Y} \hat{\Delta}(\lambda) \hat{\Gamma}_{S_X}) - \hat{\Gamma}_{S_X} + \hat{\Gamma}_{S_Y} \right\|_\infty,$$

Fig. 1. Three types of graph used in our simulation studies: (A) Hub graph; (B) Scale-free graph; (C) Small-world graph

$$L_F(\lambda_n) = \left\| \frac{1}{2} \left(\hat{\Gamma}_{S_X} \hat{\Delta}(\lambda) \hat{\Gamma}_{S_Y} + \hat{\Gamma}_{S_Y} \hat{\Delta}(\lambda) \hat{\Gamma}_{S_X} \right) - \hat{\Gamma}_{S_X} + \hat{\Gamma}_{S_Y} \right\|_F.$$

2.1 Simulation studies

In this section, we introduce simulation designs for evaluating the performance of the proposed matrix-variate differential network estimation method. For the temporal covariance matrices, we considered the following two types of structures. Type 1 is an autoregressive model, $\Sigma_{T_X} = (\sigma_{T_X, i, j})$ with elements $\sigma_{T_X, i, j} = 0.4^{|i-j|}$ and $\Sigma_{T_Y} = (\sigma_{T_Y, i, j})$ with elements $\sigma_{T_Y, i, j} = 0.5^{|i-j|}$, $1 \leq i, j \leq q$. Type 2 is a moving average model, $\Sigma_{T_X} = (\sigma_{T_X, i, j})$ with nonzero elements $\sigma_{T_X, i, j} = 1/(|i-j|+1)$ for $|i-j| \leq 3$ and $\Sigma_{T_Y} = (\sigma_{T_Y, i, j})$ with nonzero elements $\sigma_{T_Y, i, j} = 1/(|i-j|+1)$ for $|i-j| \leq 4$. Note that Σ_{T_X} and Σ_{T_Y} are both nuisance parameters. We mainly focus on the spatial graphs of two groups encoded in Ω_{S_X} and Ω_{S_Y} respectively. We first generate the edge set E_{S_X} for the group $\mathbf{X}$. To this end, we consider the following three graph structures: a hub graph, a scale-free graph, and a small-world graph, as shown in Figure 1.

- **Scenario 1: Hub graph** We partition p features into 5 equally-sized and non-overlapping sets: $C_1 \cup C_2 \cdots \cup C_5 = \{1, \dots, p\}$, $|C_k| = p/5$, $C_i \cap C_j = \emptyset$. For the smallest $i \in C_k$, we set $(i, j) \in E_{S_X}$ for all $\{j \neq i : j \in C_k\}$.
- **Scenario 2: Scale-free graph** Initially only one edge exists in E_{S_X} . Then the graph grows such that each new node is connected to one existed node with the probability proportional to the degree of the node. In practice, we use R package “huge” to generate such a graph structure.
- **Scenario 3: Small-world graph** The graph is generated by the R package “rags2ridges” with 5 starting neighbors and 5% probability of rewiring (for further detail, one may refer to Wieringen and Peeters (2016)).

The non-zero entries of Ω_{S_X} is then determined by the edge set E_{S_X} . The value of each nonzero entry of Ω_{S_X} was generated from a uniform distribution with support $[-0.5, -0.1] \cup [0.1, 0.5]$. To ensure positive definiteness of Ω^{S_X} , let $\Omega_{S_X} = \Omega_{S_X} + (1 + |\lambda_{\min}(\Omega_{S_X})|)\mathbf{I}$. We then proceed to generate the differential matrix Δ_0 . For Scenario 1, we randomly select two hub nodes from the 5 equally-sized and non-overlapping sets and the differential matrix Δ_0 is generated such that the connections of these two hub nodes change sign between Ω_{S_X} and Ω_{S_Y} . For Scenario 2 and Scenario 3, we randomly select 40% of the edges in

E_{S_X} and the differential matrix Δ_0 is generated such that the corresponding connections change sign between Ω_{S_X} and Ω_{S_Y} . The covariance matrix Σ_{S_X} and Σ_{S_Y} are generated by $(\Omega_{S_X})^{-1}$ and $(\Omega_{S_Y})^{-1}$, respectively. Finally we generate n_1 i.i.d observations of $\mathbf{X}$ from the $N_{p \times q}(\mathbf{0}, \Sigma_{T_X} \otimes \Sigma_{S_X})$ distribution and n_2 i.i.d observations of $\mathbf{Y}$ from the $N_{p \times q}(\mathbf{0}, \Sigma_{T_Y} \otimes \Sigma_{S_Y})$ distribution.

In the simulation study, we examined a range of spatial and temporal dimensions and the sample sizes: $p = \{50, 100, 200\}$, $q = \{50, 100\}$ and $n_1 = n_2 = \{20, 50\}$, which are consistent with the usual setup in functional connectivity analysis. All the simulation results are based on 100 replications.

We evaluate the performance of the estimation methods from the view of support recovery. The support recovery results are evaluated by true positive rate (TPR), true discovery rate TDR and true negative rate (TNR) along a range of tuning parameter λ . Suppose the true difference matrix Δ has the support $S_0 = \{(j, k) : \delta_{jk}^0 \neq 0, \text{ and } j \neq k\}$ and its estimator $\hat{\Delta}$ has the support set $\hat{S}$. TPR, TDR and TNR are defined as follows:

$$\text{TPR} = \frac{\text{TP}}{|S_0|}, \quad \text{TNR} = \frac{\text{TN}}{p(p-1) - |S_0|}, \quad \text{TDR} = \text{TP}/|\hat{S}_0|$$

where TP and TN are the numbers of true positives and true negatives respectively, which are defined as

$$\text{TP} = \#\{(i, j) : (i, j) \in S_0 \cap \hat{S}\},$$

$$\text{TN} = \#\{(i, j) : (i, j) \in S_0^c \cap \hat{S}^c\}.$$

We compare the proposed Matrix-Variate Differential Model (denoted as MVDN) estimation method with the joint multiple matrix Gaussian graphs estimation method in Zhu and Li (2018). Note that in Zhu and Li (2018), non-convex and convex penalty are proposed to jointly estimate matrix graphs and we simply denote the corresponding method as Non-convex and Convex. Moreover, Zhu and Li (2018) shows the advantage of directly working with the matrix data rather than working with the vector-valued data after whitening in their simulation study, so there is no need to compare the matrix-variate methods with vector-valued methods such as Zhao *et al.* (2014), Cai *et al.* (2016) and He *et al.* (2017b).

2.2 Evaluation of computational complexity

We carried out simulations for Scenario 1 to compare the computing speed of different methods with $n = \{20, 50\}$, $q = \{50, 100\}$ and $p = \{50, 100, 200, 300\}$. Timings (in seconds) are computed using a single core, Intel(R) Xeon(R) CPU E5-2697 v3 at 2.60GHz and 128 Gbytes memory, which is shown in Table 1.

Fig. 2. ROC curve (upper panels) and Precision-Recall curve (lower panels) for Scenario 1 with $p = 50, 100, 300$ and autoregressive temporal covariance structure. Solid line represents MVDN, dotted line represents convex method, and dashed line represents non-convex method. The sample size $n_1 = n_2 = 20$ and $q = 50$.

Fig. 4. ROC curve (upper panels) and Precision-Recall curve (lower panels) for for Scenario 3 with $p = 50, 100, 300$ and autoregressive temporal covariance structure. Solid line represents MVDN, dotted line represents convex method, and dashed line represents non-convex method. The sample size $n_1 = n_2 = 20$ and $q = 50$.

Fig. 3. ROC curve (upper panels) and Precision-Recall curve (lower panels) for Scenario 2 with $p = 50, 100, 300$ and autoregressive temporal covariance structure. Solid line represents MVDN, dotted line represents convex method, and dashed line represents non-convex method. The sample size $n_1 = n_2 = 20$ and $q = 50$.

2.3 Application to Electroencephalography (EEG) data

The EEG data was collected in a study examining EEG correlates of genetic predisposition to alcoholism. It is available at website <http://kdd.ics.uci.edu/datasets/eeg/eeg.data.html>, which consists of 77 alcoholic individuals and 45 controls. Each subject was fitted with a 61-lead electrode cap and was recorded at 256 Hz for 1 second. There were in addition a ground and two bipolar deviation electrodes, which are excluded from the analysis. The electrode positions were located at standard sites (Standard Electrode Position Nomenclature, American Electroencephalographic Association 1990), and were organized into frontal, central, parietal, occipital, left temporal, and right temporal regions (Acharya *et al.* (2016)). Each subject performed 120 trials under three types of stimuli. For more details on data collection, one may refer to Zhang *et al.* (1995). Similar to Xia and Li (2017), we first preprocessed the data by first averaging all trials under a single stimulus condition following Li *et al.* (2010) and then performed an α -band filtering on the signals following Hayden *et al.* (2006). Finally, we obtain 61×256 matrix data for each subject.

3 Results

3.1 Theoretical results

We show that the proposed estimator $\hat{\Delta}$ in (1) shares exactly the same support as the true differential partial correlation matrix Δ with probability tending to 1 under mild and regular conditions, indicating that the proposed method can identify all differential edges accurately as dimensions of matrix-valued data p, q and sample size n go to infinity. In fact, we show that as long as the signal strength is larger than $c\sqrt{\log(p \vee q)/(\min(n_1, n_2)q)}$ for some constant c , then we establish a theorem that $\hat{\Delta}$ can recover not only the support of Δ but also the signs of its nonzero entries.

3.2 Simulation Results

In this section we present the simulation results for the experimental design introduced in Section 2.1. To evaluate the support recovery performance of the methods, we plot the ROC curve and the Precision-Recall (PR) curve for various scenarios. Figure 2 depicts the ROC curve and PR curve for Scenarios 1 (Hub Graph) with $p = 50, 100, 200$ and autoregressive temporal covariance structure, in which the solid line represents MVDN, dotted line and dashed line represents Convex method and Non-convex method in Zhu and Li (2018) respectively. The top panels in Figure 2 are the ROC curves for $p = 50, 100, 200$, from which we can see that the MVDN performs uniformly better than the Convex and Non-convex methods. It also can be seen that the Convex and Non-convex methods performs nearly the same and both worse than the MVDN method. The advantage of the MVDN method is obviously shown from the Precision-Recall curves in the bottom panels of Figure 2. When the true positive rate (TPR) does not exceed 80%, the true discovery rate (TDR) of MVDN method is always 100% while that of Convex and Non-convex methods do not exceed 50%. From Figure 2, we can also see that as dimension p , increases, the performances of all methods get worse, which is expected as the graph becomes large. What we want to emphasize is that the MVDN method generally performs robustly as p varies. Similar conclusions above from Figure 2 can be drawn from Figure 3 and Figure 4, which correspond to Scenario 2 (Scale-free graph) and Scenario 3 (Small-world graph) respectively. From Figure 2 to Figure 4, we can see that all methods performs similarly for different types of graph structures. To save space, we put all the other simulation results in the Supplementary material. From the

Figures in the Supplementary material, we can also draw conclusion that for the same Scenarios, as sample size n_1, n_2 increases (with p, q fixed), the performances of all methods get better. Besides, with n, p fixed, and q increases, the performances of all methods also get better as we collect more temporal data, thus have more information excavated.

Table 1. Average timings (in seconds) over 20 replications.

$N = 20$				
p	$q = 50$		$q = 100$	
	MVDN	Non-convex	MVDN	Non-convex
50	0.13	28.73	0.13	26.74
100	0.61	141.30	0.59	114.18
200	3.34	1079.23	2.98	676.28
300	10.27	4418.82	9.97	2262.24
$N = 50$				
p	$q = 50$		$q = 100$	
	MVDN	Non-convex	MVDN	Non-convex
50	0.14	25.79	0.15	26.01
100	0.60	108.45	0.60	112.76
200	2.86	574.25	2.76	616.70
300	8.40	1624.75	7.87	1913.39

In summary, the simulation results show that the MVDN method always outperforms its competitors Non-convex and Convex method in all the scenarios, illustrating the advantages of the MVDN method. Zhu and Li (2018) showed that the non-convex and convex methods substantially outperforms the state of the art vector-normal-based methods. It further indicates the advantage of MVDN method against the vector-normal-based methods.

3.3 Application Results

Figure 5 illustrate the differential network estimated by different methods, in which orange edges show an increase in partial correlation dependency from control group to alcoholic group; grey edges show a decrease and red points stand for hub nodes whose number of edges ranks the top 5 for each method. Specifically, the tuning parameter λ for method MVDN is selected by minimizing the Bayesian information criterion, i.e., λ is chosen to minimize $(n_1 + n_2)L(\lambda) + \log(n_1 + n_2)\|\Delta\|_0$ where $L(\lambda)$ represents the loss function based on L_F norm introduced in section 2. For the Non-convex and Convex method, the λ is selected by minimizing a prediction criterion by using fivefold cross-validation as described in Zhu and Li (2018), i.e., first divide the data set into five parts $\mathcal{D}_1, \dots, \mathcal{D}_5$ for the control (denoted as group 1) and the alcoholic group (denoted as group 2). For each group $k, k = 1, 2$, define $\hat{\Gamma}_k^l$ and $\hat{\Gamma}_k^{-l}$ to be the sample correlation matrices based on samples in $\mathcal{D}_l$ and $\{\mathcal{D}_1, \dots, \mathcal{D}_5\} \setminus \mathcal{D}_l, l = 1, \dots, 5$. Let $\hat{\Omega}_k^{-l}(\lambda)$ be the partial correlation matrices calculated based on $\hat{\Gamma}_k^{-l}$ under the tuning parameter λ . Then λ is chosen to minimize $CV(\lambda)$ which is defines as

$$CV(\lambda) = \sum_{l=1}^5 \sum_{k=1}^2 \left(-\log \left(\det \left(\hat{\Omega}_k^{-l}(\lambda) \right) \right) + \text{Tr} \left(\hat{\Gamma}_k^l \hat{\Omega}_k^{-l}(\lambda) \right) - p \right).$$

From Figure 5, we can see that the differential network structure estimated by MVDN method is much sparse compared with those estimated by Non-convex and Convex methods. The differential network structure estimated by Non-convex and Convex methods are so dense that it is impossible to draw any useful conclusions from the graphs. In contrast, we can draw interesting conclusions from the differential graph estimated by MVDN. Firstly, we can see that those electrodes in the frontal region and temporal region (denoted by symbols FP, AF, F,TC and TP)

show a decrease in connections and some asymmetry between the left and right frontal regions in the alcoholic group compared to the control, which agrees with the findings in the literature, see for example, Hayden *et al.* (2006). Secondly, those electrodes in the central region and parietal region mainly show a increase in connections in the alcoholic group compared to the control. Thirdly, about 56.1% of the identified differential edges by MVDN show a increase in connection from the control group to the alcoholic group, which indicates that alcoholic would strengthen the connection activities in brain region.

4 Discussion

Brain functional connectivity reveals the synchronization of brain systems through correlations in neurophysiological measures of brain activity. Growing evidence shows that the brain connectivity network experiences alternations with the presence of numerous neurological disorders. Thus it entails the analysis of differential network. The most typical feature of brain connectivity data is its matrix-form while the primary interest is to infer the spatial connectivity network. To our knowledge, no literatures focus on the differential network analysis for matrix-form type data. This motivates us to consider such an important topic. In this paper, we establish a matrix-variate differential network Model assuming that the matrix-variate data follows a matrix-normal distribution, and propose to exploit the D-trace loss function and a Lasso-type penalty to estimate the spatial differential partial correlation matrix directly. Theoretically we show that the proposed method can identify all differential edges accurately with probability tending to 1 in high-dimensional setting under mild and regular conditions.

The work could be extended in the following aspects. First, in this paper, we exploit the D-trace loss function and a Lasso-type penalty to estimate the spatial differential partial correlation matrix. In fact we can consider the constrained ℓ_1 minimization approach as in Zhao *et al.* (2014), that is,

$$\arg \min |\Delta|_1, \quad \text{subject to}$$

$$|\hat{\Gamma}_{S_X} \Delta \hat{\Gamma}_{S_Y} - \hat{\Gamma}_{S_X} + \hat{\Gamma}_{S_Y}|_\infty \leq \lambda_n,$$

where λ_n is a tuning parameter. The theoretical properties of the estimator are still needed to be investigated, but we expect that the empirical performance should be similar.

Second, the matrix-normal distribution assumption for the brain connectivity data could be somewhat restrictive. Ning and Liu (2013) propose a semiparametric extension of the matrix-normal distribution, called the matrix-nonparanormal distribution. Thus we can assume that the brain connectivity data follows matrix-nonparanormal distribution and establish the differential network model for matrix-nonparanormal data. This extension is more challenging because there exists diverging number of unknown transformation functions needed to be carefully dealt. We leave this problem as a promising future work as there is a lot of work still needed to be done.

The proposed matrix-variate differential network model is very flexible and may provide deeper understanding of the brain connectivity alternation mechanism. It is demonstrated the matrix-variate differential network models enjoy great advantages over existing models and thus are highly recommended in the brain connectivity analysis.

Acknowledgements

We would like to thank professor Yunzhang Zhu for providing the simulation codes of their Non-convex and Convex method.

Fig. 5. Differential network estimated by different methods. Orange edges show an increase in partial correlation dependency from control group to alcoholic group; grey edges show a decrease. Red points stand for hub nodes whose number of edges ranks the top 5 for each method.

Funding

This work was supported by grants from the National Natural Science Foundation of China (grant number 81803336 and 11801316), Natural Science Foundation of Shandong Province (ZR2018BH033) and National Statistical Scientific Research Project (2018LY63). The funding body played no role in the design, writing or decision to publish this manuscript. *Conflict of Interest:* none declared

References

- Acharya, J. N., Hani, A. J., R., J. C., CNIM, P. T., and Tsuchida, T. N. (2016). American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature. *The Neurodiagnostic Journal*, **56**(4), 245–252.
- Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **106**(494), 594–607.
- Cai, T., Li, H., Liu, W., and Xie, J. (2016). Joint estimation of multiple high-dimensional precision matrices. *Statistica Sinica*, **26**, 445–464.
- Daliri, M. R. and Behroozi, M. (2013). Advantages and disadvantages of resting state functional connectivity magnetic resonance imaging for clinical applications. *Omics Journal of Radiology*, **3**(1).
- Fox, M. D. and Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience*, **4**(19), 19.
- Hayden, E., Wiegand, Remeyer, E., Bauer, L., O'Connor, S., Chorlian, D., Porjesz, B., and Begleiter, H. (2006). Patterns of regional brain activity in alcohol-dependent subjects. *Alcoholism Clinical & Experimental Research*, **30**(12), 1986–1991.
- He, Y., Zhang, X., Wang, P., and Zhang, L. (2017a). High dimensional gaussian copula graphical model with FDR control. *Computational Statistics & Data Analysis*, **113**, 457–474.
- He, Y., Zhang, X., Ji, J., and Liu, B. (2017b). Joint estimation of multiple high-dimensional gaussian copula graphical models. *Australian & New Zealand Journal of Statistics*, **59**(3), 289–310.
- Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *Journal of the American Statistical Association*, **107**(499), 1187–1200.
- Li, B., Kim, M. K., and A., N. (2010). On dimension folding of matrix- or array-valued statistical objects. *Annals of Statistics*, **38**(2), 1094–1121.
- Liu, H., Lafferty, J., and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine*

Learning Research, **10**, 2295–2328.

- Liu, H., Han, F., Yuan, M., Lafferty, J., and Wasserman, L. (2012). High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, **40**(4), 2293–2326.
- Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, **34**, 1436–1462.
- Ning, Y. and Liu, H. (2013). High-dimensional semiparametric bigraphical models. *Biometrika*, **100**(3), 655–670.
- Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Publications of the American Statistical Association*, **104**(486), 735–746.
- Varoquaux, G. and Craddock, R. C. (2013). Learning and comparing functional connectomes across subjects. *Neuroimage*, **80**(8), 405–415.
- Wieringen, W. N. V. and Peeters, C. F. W. (2016). Ridge estimation of inverse covariance matrices from high-dimensional data. *Computational Statistics & Data Analysis*, **103**(C), 284–303.
- Xia, Y. and Li, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics*, **73**(3).
- Xia, Y. and Li, L. (2018). Matrix graph hypothesis testing and application in brain connectivity alteration detection. *Statistica Sinica*, to appear.
- Xia, Y., Cai, T., and Cai, T. T. (2015). Testing differential networks with applications to detecting gene-by-gene interactions. *Biometrika*, **102**(2), 247–266.
- Yin, J. and Li, H. (2012). Model selection and estimation in the matrix normal graphical model. *Journal of Multivariate Analysis*, **107**(3), 119.
- Yuan, H., Xi, R., and Deng, M. (2015). Differential network analysis via the lasso penalized d-trace loss. *Biometrika*, **104**(4).
- Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, **94**(1), 19–35.
- Zhang, T. and Zou, H. (2014). Sparse precision matrix estimation via lasso penalized D-trace loss. *Biometrika*, **101**(1), 103–20.
- Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W., and Litke, A. (1995). Event related potentials during object recognition tasks. *Brain Research Bulletin*, **38**(6), 531–8.
- Zhao, S. D., Cai, T. T., and Li, H. (2014). Direct estimation of differential networks. *Biometrika*, **101**(2), 253–268.
- Zhou, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Annals of Statistics*, **42**(2), 532–562.
- Zhu, Y. and Li, L. (2018). Multiple matrix gaussian graphs estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, to appear.