Sparse Brain Network Recovery under Compressed Sensing

Hyekyoung Lee[1], Dong Soo Lee[1], Hyejin Kang[1], Boong-Nyun Kim[2], Moo K. Chung[3]

[1] Department of Nuclear Medicine

Seoul National University College of Medicine, Korea

`leehk@postech.ac.kr, dsl,hkang211@snu.ac.kr`

[2] Department of Neuropsychiatry

Seoul National University College of Medicine, Korea

`kbn1@snu.ac.kr`

[3] Department of Biostatistics and Medical Informatics

Waisman Laboratory for Brain Imaging and Behavior

`mkchung@wisc.edu`

August 17, 2010

Abstract

Partial correlation is a useful connectivity measure for brain networks, especially, when it is needed to remove the confounding effects in highly correlated networks. Since it is difficult to estimate the exact partial correlation under the small-n large-p situation, a sparseness constraint is generally introduced. In this paper, we consider the sparse linear regression model with a l_1-norm penalty, a.k.a., least absolute shrinkage and selection operator (LASSO), for estimating sparse brain connectivity. LASSO is a well-known decoding algorithm in the compressed sensing (CS). The CS theory states that LASSO can reconstruct the exact sparse signal even from a

small set of noisy measurements. We briefly show that the penalized linear regression for partial correlation estimation is related with CS. It opens a new possibility that the proposed framework can be used for a sparse brain network recovery. As an illustration, we construct sparse brain networks of 97 regions of interest (ROIs) obtained from FDG-PET data for the autism spectrum disorder (ASD) children and the pediatric control (PedCon) subjects. As a model validation, we check their reproducibilities by leave-one-out cross validation and compare the clustered structures derived from the brain networks of ASD and PedCon.

Keywords: Brain Connectivity, Compressed Sensing, Partial Correlation, LASSO.

1   Introduction

The functional and anatomical connectivity of human brain has known to exhibit large and complex network structures with nontrivial topological characteristics [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. By incorporating the graph theoretical approaches into connectivity analysis, we can gain a new understanding of the characteristics of human brain, from a microscale connectivity between single neurons to a macroscale connectivity between regions of interest (ROIs) in brain images. The brain connectivity has been usually categorized into well-known complex networks such as small-world [1, 3, 4, 5, 7, 11], scale-free [2, 5] or modular networks [8, 9, 10]. The human brain networks is formed from connectivity matrices defined between neuronal elements (single neurons for microscale and ROIs for macroscale network modeling). They are also known as human connectome [13].

The majority of previous brain network studies have been based on thresholding correlation in localizing the focal regions of high connectivity [14, 15, 2]. The correlation is used as a similarity measure of network connectivity between two regions. However, the main limitation of correlation-based connectivity analyses is that they fail to explicitly factor out the confounding effect of other regions. To remedy this shortcoming, partial correlation has been naturally introduced in factoring out the dependency of other regions [16, 7, 6] or eliminating the effect of experimental designs [17]. Unfortunately, this type of problem usually belongs to the small-n large-p setting, where the number of regions p are substantially larger than the number of samples n, so it is not feasible to estimate the exact partial correlation accurately [18, 19]. So far the majority of literature have used the penalized likelihood method in imposing the sparseness on the partial correlation estimation [20, 21, 22, 23, 24, 25]. Moreover, since the brain networks are known to be sparse and

highly clustered [26, 7], it is reasonable to incorporate the sparsity of network in estimating partial correlation. In this paper, we introduce a different approach based on the penalized linear regression for estimating sparse partial correlation [27, 28]. The penalized linear regression with l_1-norm, which is also known as the basis pursuit denoising in signal processing and least absolute shrinkage and selection operator (LASSO) in statistics, is usually formulated as the convex optimization to find the sparsest solution of the under-determined linear regression problem [29, 30].

LASSO is one of preferred decoding algorithms in the compressed sensing (CS) theory [31, 32, 33, 34]. The CS-theory states that if LASSO satisfies sparsity and incoherence, the exact recovery is guaranteed with the overwhelmingly high probability, even though the measurement data is not sufficient and contaminated with noise [35, 36]. Note that, in some sense, incoherence is a stronger condition than a uniform uncertainty principle (UUP) [37]. So, if our penalized linear regression for the partial correlation estimation satisfies sparseness and UUP, CS may provide a natural framework for modeling sparse brain networks, which has not been attempted before. It has been already examined that some measurement ensembles such as the Gaussian, Fourier or binary ensembles hold UUP [35]. We show that the proposed framework and the brain imaging data satisfy UUP by borrowing UUP for Gaussian ensemble.

In this paper, we focus on sparse model building of the macro-scale connectivity of human brain under CS. The proposed model is applied to the 97 ROIs extracted from FDG-PET data for autistim spectrum disorder (ASD) children and pediatric control (PedCon) subjects. It is generally known that ASD has the global underconnectivity and the local overconnectivity in the key brain regions [38, 39]. The differences between ASD and PedCon are mostly found in connectivities between lobes, especially, connection between secondary association cortices such as frontal and parietal regions [40, 41, 5]. Dense internal and sparse external linkages are properties of a cluster (also called a community or module). In particular, some studies suggest that the small-world network, which is one of famous characteristics of brain connectivity, induces a modular architecture (or community structure) [8, 9, 10]. Therefore, in this study, after estimating the partial correlation by the penalized linear regression, we seek the possible modular structures of ASD and PedCon brain network and observe their differences based on the lobe structures.

The main contributions of this paper are:

- to formulate the sparse brain connectivity based on correlation and partial correlation in the

penalized linear regression framework,

- to simply show that the penalized linear regression for partial correlation estimation can near-optimally recover the sparse brain connectivity by showing our study satisfies UUP of the Gaussian ensemble,

- to show the reproducibilities of the estimated networks by changing the data set using the leave-one-out cross-validation,

- and to show the modular structures of ASD and PedCon brain networks based on the existing lobe structures for the first time.

The organization of the paper is as follows. In Section 2, we provide notations that will be used through the paper and present the standard methods for calculating correlation and partial correlation. We formulate the problem of estimating correlation and the partial correlation under a sparsity constraint as the sparse linear regression in Section 3. In Section 4, after briefly introducing CS and checking the precondition for satisfying UUP of the proposed model, we further prove that our LASSO-based connectivity method satisfies UUP and the near-optimal brain connectivity recovery can be done. Numerical experiments are given in Section 5, where we use the 97 ROIs extracted from FDG-PET data for 26 autistic and 11 pediatric control subjects. We show that the proposed method consistently finds the brain networks which characterize the two groups well and have significant group differences in network connectivity.

2   Network Construction

A brain network is a graph $\mathcal{G} = (V, E)$ consisting of vertex (or node) set V and edge (or links) set E, which is a subset of $V \times V$. The graph can be uniquely represented as a square connectivity matrix, where the size of matrix is the number of nodes in the graph. The elements of connectivity matrix correspond to the weights of edge between two nodes. The connectivity matrix of an unweighted graph consists of entries 0 and 1, representing whether the edge exists or not. We call such a connectivity matrix an adjacency matrix, which is also obtained by thresholding the elements of connectivity matrix. In this section, we briefly introduce a connectivity matrix defined in terms of correlation and partial correlation.

2.1 Connectivity Matrix

Suppose $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_p\}$ is the n-dimensional data vector measured at the p selected ROIs on the FDG-PET images of n subjects. We assume $\boldsymbol{f}_i$ are centered and normalized, , i.e., $\boldsymbol{f}_i^\top \boldsymbol{1}_n = 0$ and $\boldsymbol{f}_i^\top \boldsymbol{f}_i = 1$ for all $i = 1, \ldots, p$. $\boldsymbol{1}_n \in \mathbb{R}^{n \times 1}$ is a vector of which all elements are equal to 1. The sample covariance matrix is then defined as

$$\boldsymbol{\Sigma} = [\sigma_{ij}] = \left[\frac{\boldsymbol{f}_i^\top \boldsymbol{f}_j}{n-1} \right] \in \mathbb{R}^{p \times p}. \tag{1}$$

The correlation coefficient between $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ is defined as

$$\begin{aligned}
\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} &= (\boldsymbol{f}_i^\top \boldsymbol{f}_j)/\sqrt{(\boldsymbol{f}_i^\top \boldsymbol{f}_i)(\boldsymbol{f}_j^\top \boldsymbol{f}_j)} \\
&= \boldsymbol{f}_i^\top \boldsymbol{f}_j.
\end{aligned} \tag{2}$$

If we note the inverse of covariance matrix, i.e., concentration or precision matrix, as $\boldsymbol{\Pi} = \boldsymbol{\Sigma}^{-1} = [\pi_{ij}] \in \mathbb{R}^{p \times p}$, the partial correlation between $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ is [42]

$$\theta_{ij} = -\pi_{ij}/\sqrt{\pi_{ii}\pi_{jj}}. \tag{3}$$

Usually the connectivity matrix is then constructed as a function of correlation or partial correlation.

2.2 Relationship between Correlation and Partial Correlation

The elements π_{ij} $(i < j)$ of precision matrix $\boldsymbol{\Pi}$ can be rewritten as

$$\pi_{ij} = \frac{(-1)^{i+j}\boldsymbol{M}_{(i,j)}}{|\boldsymbol{\Sigma}|} = \sigma_{ji}\frac{\boldsymbol{M}_{(ij,ij)}}{|\boldsymbol{\Sigma}|} - \sum_{k \neq i,j}(-1)^{i+k}\sigma_{jk}\frac{\boldsymbol{M}_{(ij,jk)}}{|\boldsymbol{\Sigma}|}, \tag{4}$$

where $\boldsymbol{M}_{(ij,jk)}$ is the minor, i.e., the determinant of the matrix removing the ith and j^{th} rows and the jth and k^{th} columns from $\boldsymbol{\Sigma}$. By interchanging the indices, we also have

$$\pi_{ji} = \frac{(-1)^{i+j}\boldsymbol{M}_{(j,i)}}{|\boldsymbol{\Sigma}|} = \sigma_{ij}\frac{\boldsymbol{M}_{(ij,ij)}}{|\boldsymbol{\Sigma}|} - \sum_{k \neq i,j}(-1)^{j+k}\sigma_{ik}\frac{\boldsymbol{M}_{(ij,ik)}}{|\boldsymbol{\Sigma}|}. \tag{5}$$

Since the precision matrix is symmetric, an algebraic manipulation shows

$$\pi_{ij} = \sigma_{ij}\frac{\boldsymbol{M}_{(ij,ij)}}{|\boldsymbol{\Sigma}|} - \frac{1}{2}\sum_{k \neq i,j}\left(\sigma_{jk}\pi_{ki}|_{\sim j}\frac{\boldsymbol{M}_{(j,j)}}{|\boldsymbol{\Sigma}|} + \sigma_{ik}\pi_{kj}|_{\sim i}\frac{\boldsymbol{M}_{(i,i)}}{|\boldsymbol{\Sigma}|}\right),$$

where $\pi_{ki}|_{\sim j}$ is the (k, i) element of inverse of the matrix removing the jth row and column from $\boldsymbol{\Sigma}$. The above fomula shows that the partial correlation π_{ij} is the correlation between the ith and the jth data vectors (the first term) while excluding all correlations intermediated by other data vectors (the second term). The partial correlation has been often used for the brain network analysis and its applications due to the ability factoring out the influence of other regions on correlation [43]. However, for the small-n and large-p setting, it is difficult to estimate the partial correlation. So we introduce sparseness constraint in the linear regression model for estimating partial correlation.

3 Sparse Connectivity Estimation

In this section, we present the penalized linear regression for estimating sparse correlation and partial correlation.

3.1 Linear Regression for Correlation and Partial Correlation

Both correlation and partial correlation can be obtained by the linear regression as follows :

- Correlation ρ_{ij}:

$$\boldsymbol{f}_i = \alpha_{ij}\boldsymbol{f}_j + \boldsymbol{\epsilon}_i, \text{ for } i = 1, ..., p, \tag{6}$$

  where the regression parameters α_{ij} is equal to correlation ρ_{ij} in (2) when $\boldsymbol{f}_i$ is centered and normalized.

- Partial Correlation θ_{ij}:

$$\boldsymbol{f}_i = \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j + \boldsymbol{\epsilon}_i, \text{ for } i = 1, ..., p, \tag{7}$$

  where $\boldsymbol{\epsilon}_i$ is uncorrelated with all variables except $\boldsymbol{f}_i$ and β_{ij} is the measure of relationship between data vectors $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ given all other data vectors. When $\text{var}(\boldsymbol{\epsilon}_i) = (1/\pi_{ii})$ and $\text{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \pi_{ij}/(\pi_{ii}\pi_{jj})$, the partial correlation θ_{ij} is given by [27]

$$\theta_{ij} = -\pi_{ij}/\sqrt{\pi_{ii}\pi_{jj}} = \beta_{ij}\sqrt{(\pi_{ii}/\pi_{jj})}. \tag{8}$$

6

The proof of statements (6) and (8) are in Appendix 7.

Now, we change the linear regression model in (7) to a matrix form (See Fig. 1). If we denote $\boldsymbol{X} = [\boldsymbol{f}_1, \cdots, \boldsymbol{f}_p] \in \mathbb{R}^{n \times p}$ and $\boldsymbol{B} = [\beta_{ij}] \in \mathbb{R}^{p \times p}$ where $\boldsymbol{B}$ is a symmetric matrix with zero diagonal terms, we can rewrite the linear regression model for estimating partial correlation in (7) as

$$\boldsymbol{X} = \boldsymbol{X}\boldsymbol{B}. \tag{9}$$

Now we vectorize both sides in (9) as

$$\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{X}\boldsymbol{B}), \tag{10}$$

where $\text{vec}(\boldsymbol{X}) = [\boldsymbol{f}_1^\top \cdots \boldsymbol{f}_p^\top]^\top$ is the vectorization operator. Since $\text{vec}(\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}) = (\boldsymbol{Z}^\top \otimes \boldsymbol{X})\text{vec}(\boldsymbol{Y})$ with the Kronecker product $\otimes$ [44], we have

$$\text{vec}(\boldsymbol{X}\boldsymbol{B}) = (\boldsymbol{I} \otimes \boldsymbol{X})\text{vec}(\boldsymbol{B}),$$

where $\boldsymbol{I} \in \mathbb{R}^{p \times p}$ is a identity matrix. Then, (10) can be written in a matrix form

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{b}, \tag{11}$$

where $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$, $\boldsymbol{A} = (\boldsymbol{I} \otimes \boldsymbol{X}) \in \mathbb{R}^{np \times p^2}$ and $\boldsymbol{b} = \text{vec}(\boldsymbol{B}) \in \mathbb{R}^{p^2 \times 1}$. $\boldsymbol{A}$ is a block diagonal matrix, but not a square matrix, of which main diagonal blocks consist of $\boldsymbol{X}$s as shown in Fig. 1(b). Note that the elements of $\boldsymbol{X}$ follows Gaussian distribution with mean 0 and variance 1 because column vectors, $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_p$, are centered and normalized.

If we denote $N = np$ and $P = p^2$ ($n \ll p$), estimating the partial correlation in the linear model (11) fall under a high-dimension-small-sample-size situation. Thus, we should consider shrinkage methods to regularize the model parameters and one solution is to add the sparseness penalty such as l_1 norm to the model parameters.

3.2 Adding the Sparseness Constraint

The solution of linear data model in (11) is usually obtained by the least squares minimization:

$$\hat{\boldsymbol{b}} = \underset{\boldsymbol{b}}{\arg\min} \parallel \boldsymbol{x} - \boldsymbol{A}\boldsymbol{b} \parallel_2^2, \tag{12}$$
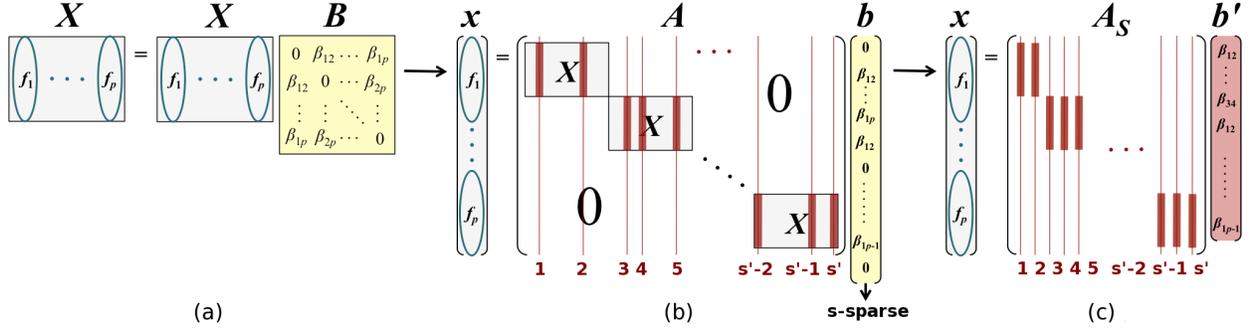
Figure 1: Linear regression model for the partial correlation estimation. Linear regression model in (7) is represented as (a) $\boldsymbol{X} = \boldsymbol{XB}$, where $\boldsymbol{X} = [\boldsymbol{f}_1, \cdots, \boldsymbol{f}_p] \in \mathbb{R}^{n \times p}$ and $\boldsymbol{B} = [\beta_{ij}] \in \mathbb{R}^{p \times p}$. $\boldsymbol{B}$ is a symmetric matrix with zero diagonal terms. By vectorization $\text{vec}(\cdot)$ and the property of Kronecker product $\otimes$, it can be changed to (b) $\boldsymbol{x} = \boldsymbol{Ab}$, where $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$, $\boldsymbol{A} = \boldsymbol{I} \otimes \boldsymbol{X} \in \mathbb{R}^{np \times p^2}$ and $\boldsymbol{b} = \text{vec}(\boldsymbol{B}) \in \mathbb{R}^{p^2 \times 1}$. $\boldsymbol{I} \in \mathbb{R}^{p \times p}$ is a identity matrix. If $\boldsymbol{b}$ is s-sparse, i.e. it has at most s number of nonzero elements, only the s' selected column vectors of $\boldsymbol{A}$ are used for calculating $\boldsymbol{x}$ ($s' \leq s$) as shown in (c).

where $\| \cdot \|_2$ is a l_2-norm. When the linear regression is under a small-n large-p problem, there exist possibly many solutions and we need to add the sparseness constraint to obtain one solution. The sparsest solution is obtained by l_0-norm penalty (number of non-zero elements) such as

$$\min_{\boldsymbol{b}} \| \boldsymbol{b} \|_0 \text{ subject to } \boldsymbol{x} = \boldsymbol{Ab}. \tag{13}$$

Since it is a combinatorial problem with NP-hard complexity, insted of l_0-norm, we employ l_1-norm penalty (sum of absolute values of elements) such as

$$\min_{\boldsymbol{b}} \| \boldsymbol{b} \|_1 \text{ subject to } \boldsymbol{x} = \boldsymbol{Ab}, \tag{14}$$

which is related with the linear programming, a.k.a., the basis pursuit denoising problem [29]. The discussion about the l_0 and l_1 equivalence can be found in [45]. For a noisy case, we can change (14) to the quadratic programming with a linear constraint

$$\min_{\boldsymbol{b}} \| \boldsymbol{x} - \boldsymbol{Ab} \|_2^2 \text{ subject to } \| \boldsymbol{b} \|_1 < \epsilon. \tag{15}$$

This is known as LASSO [30, 46]. We apply the LASSO framework to the least squares loss function of (6) and (7) for estimating sparse correlation and partial correlation as follows [27] :

- Model for sparse correlation estimation:

$$\hat{\alpha}_{ij} = \arg\min_{\alpha_{ij}} \sum_{i=1}^{p} \sum_{j \neq i} \parallel \boldsymbol{f}_i - \alpha_{ij}\boldsymbol{f}_j \parallel_2^2 + \lambda \sum_{i,j} |\alpha_{ij}|, \tag{16}$$

- Model for sparse partial correlation estimation:

$$\hat{\beta}_{ij} = \arg\min_{\beta_{ij}} \sum_{i=1}^{p} \parallel \boldsymbol{f}_i - \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j \parallel_2^2 + \lambda \sum_{i,j} |\beta_{ij}|. \tag{17}$$

To solve these problems, we exploited the coordinate descent learning and the active-set algorithm [27, 47]. The derivation and algorithms are outlined in Appendix 8.

4 Relationship with Compressed Sensing

In this section, we show that the sparse linear model for estimating partial correlation is related with UUP. It opens a possibility that the the near-optimal recovery of sparse brain network is done by our propose method.

4.1 Uniform Uncertainty Principle

The coherence of $\boldsymbol{A}$, $\mu(\boldsymbol{A})$, is defined by the maximum correlation coefficient among all correlations between two different column vectors in (11). If all basis vectors are orthogonal, the coherence is minimized. The exact sparse signal recovery is guaranteed for s-sparse signal, i.e. at most s number of nonzero elements in a signal, if the coherence is bounded by $\mu(\boldsymbol{A}) \leq \mathcal{O}(1/s)$ [48]. However, since we assume that $N \ll P$, this condition is hard to satisfy. A looser condition, which guarantees the near-optimal sparse data recovery is a UUP, a.k.a., restricted isometry property (RIP) [35, 49].

Definition 1 (UUP) *A measurement matrix $\boldsymbol{A}$ satisfies the* uniform uncertainty principle *with oversampling factor λ if, for every sufficiently small $\gamma > 0$ and any s-sparse vector $\boldsymbol{b}$ such that*

$$s \leq \gamma \cdot N/\lambda,$$

$\boldsymbol{A}$ holds inequalities

$$\frac{1}{2} \cdot \frac{N}{P} \cdot \parallel \boldsymbol{b} \parallel_2^2 \ \leq \ \parallel \boldsymbol{A}\boldsymbol{b} \parallel_2^2 \ \leq \ \frac{3}{2} \cdot \frac{N}{P} \cdot \parallel \boldsymbol{b} \parallel_2^2, \tag{18}$$

with probability at least $1 - \mathcal{O}(p^{-\rho/\gamma})$ for some fixed constant $\rho > 0$.

Supposed that S is a subset of index set $\{1, \ldots, P\}$ indicating the indices of s nonzero elements in $\boldsymbol{b}$ and $\boldsymbol{A}_S \in \mathbb{R}^{N \times s}$ is a submatrix of $\boldsymbol{A}$ consisting of s column vectors selected by the index set S (see Fig. 1(c)). Then, (18) is equivalent to

$$\frac{1}{2} \cdot \frac{N}{P} \leq \lambda_{\min}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) \leq \lambda_{\max}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) \leq \frac{3}{2} \cdot \frac{N}{P}, \tag{19}$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are the largest and smallest eigenvalues.

Lemma 1 *The Gaussian ensemble $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, which are i.i.d. $\mathcal{N}(0, 1/p)$, holds the UUP with oversampling $\lambda = \log p$.*

Lemma 1 shows that, if we pick $\boldsymbol{A}$ in the sparse linear model $\boldsymbol{x} = \boldsymbol{Ab}$ as a Gaussian measurement ensemble $\boldsymbol{X}$ with $n \geq \gamma \cdot s \log p$, then, sparse recovery of $\boldsymbol{b}$ can be done with overwhelmingly large probability [35].

Now, we will show that the brain connectivity obtained by the penalized linear regression in (11) can be recovered under UUP for the first time. Since it is induced based on lemma 1, we will show the Gaussianess of data matrix $\boldsymbol{X}$ in (11) first.

4.2 Gaussianess of Data Matrix

Before discussing UUP of $\boldsymbol{A} = \boldsymbol{I} \otimes \boldsymbol{X}$ in the penalized linear regression model (11), we check its gaussianess of data matrix $\boldsymbol{X} = [\boldsymbol{f}_1, \cdots, \boldsymbol{f}_p] \in \mathbb{R}^{n \times p}$ using both Lilliefors test and quantile-quantile plots (QQ-plots) [50]. The QQ-plot visualizes how the empirical distribution follows the theoretical Gaussian distribution by plotting the empirical quantile in the vertical axis versus the expected theoretical quantile from a normal distribution in the horizontal axis. Because it is not possible to visualize QQ-plots for all ROIs here, we measured the correlation coefficients r of scatter points in the QQ-plots. If the empirical distribution follows Gaussian, r should asymptotically converge to 1. For the normally distributed random numbers from $\mathcal{N}(0, 1)$ whose sample sizes are same with the data matrix of ASD and PedCon, we obtained $r = 0.98 \pm 0.01$ and $r = 0.97 \pm 0.03$, respectively. For the measured data in ROIs of ASD and PedCon, we obtained $r = 0.97 \pm 0.02$ and $r = 0.96 \pm 0.03$ in Fig. 2 left panels. So the high correlation guarantees that our data matrix come from a normal distribution. Using Lilliefors statistic, we statistically tested the Gaussian assumption. Since the
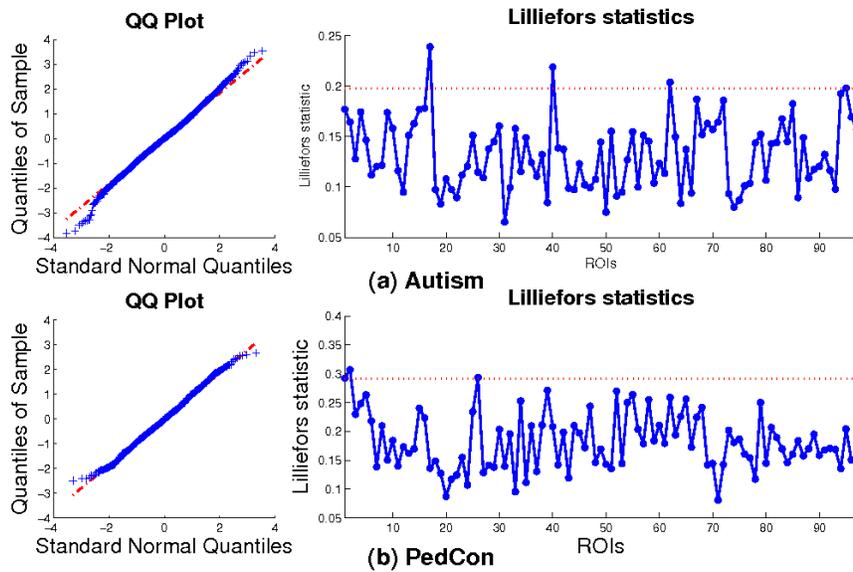
Figure 2: Checking Gaussian assumption of data matrix measured in ROIs for (a) ASD and (b) PedCon. In the left panels, the horizontal axises indicate the quantiles of a normal distribution while the vertical axises indiate the quantiles of an input sample. If the input data satisfies the Gaussian assumption, the blue dots closely lie along the straight red line. In the right panels, the horizontal axises display the index of ROIs while the vertical axes displays Lilliefors statistic which measures the maximum difference between empirical and theoretical Gaussian distributions. Most of ROIs (blue solid lines) shows value less than the cutoff values (red dotted lines), (a) 0.19 and (b) 0.29 at the α level of 1%.

Lilliefors statistics of data matrix are mostly smaller than the cutoff values of 0.19 (ASD) and 0.29 (PedCon) at 1% level in Fig. 2 right panels, there is no reason not to assume normality for the given data matrix, i.e. i.e., $X_{ij} \sim$ i.i.d. $\mathcal{N}(0,1)$.

4.3 Sparse Brain Connectivity Recovery under CS

For notational convenience, we replace $\boldsymbol{X}$ with $\frac{1}{\sqrt{p}}\boldsymbol{X}$. Then, the elements of $\boldsymbol{X}$ follows $X_{ij} \sim \mathcal{N}(0, 1/p)$ and $\boldsymbol{X}$ holds UUP with oversampling $\lambda = \log p$. Now, based on the gaussianess of data matrix $\boldsymbol{X}$ and lemma 1, we show that the measurement matrix $\boldsymbol{A} = (\boldsymbol{I} \otimes \boldsymbol{X}) \in \mathbb{R}^{N \times P}$ in (11) holds UUP.

Denote $\boldsymbol{A}_S$ and $\boldsymbol{A}_{S'}$ as submatrices formed by taking s and s' column vectors from $\boldsymbol{A}$ ($s' \leq s$), respectively. Submatrices $\boldsymbol{X}_K$ and $\boldsymbol{X}_{K'}$ of $\boldsymbol{X}$ ($k' \leq k$) are defined by the same way. $\boldsymbol{A}_S$ is a block diagonal matrix (not a square matrix) of which block matrices are $\boldsymbol{X}_{K_1}, \ldots, \boldsymbol{X}_{K_p}$ with the number of column vectors $k_1, \ldots k_p$. The column vectors of all block matrices of $\boldsymbol{A}_S$ are selected from $\boldsymbol{X}_K$ with replacement ($k_1, \ldots, k_p \leq k \leq s$). To clarify the explanation, $\boldsymbol{A}_S$ is illustrated in Fig. 1(c). $\boldsymbol{A}_{S'}$, a submatrix of $\boldsymbol{A}_S$, is constructed from $\boldsymbol{X}_{K'}$ in the same way. Note that $\boldsymbol{A} = \boldsymbol{I} \otimes \boldsymbol{X}$ is a block diagonal matrix of which diagonal term is repeatedly filled with $\boldsymbol{X}$ (see Fig. 1(c)). Subsequently $\lambda_{\min}(\boldsymbol{A}^\top \boldsymbol{A}) = \lambda_{\min}(\boldsymbol{X}^\top \boldsymbol{X})$ and $\lambda_{\max}(\boldsymbol{A}^\top \boldsymbol{A}) = \lambda_{\max}(\boldsymbol{X}^\top \boldsymbol{X})$ hold [44].

We use $\begin{pmatrix} P \\ s' \end{pmatrix}$ to indicate possible combinations of s' objects from P objects. To see whether $\boldsymbol{A}$ holds UUP, we should exhibit that all $\begin{pmatrix} P \\ s' \end{pmatrix}$ submatrices $\boldsymbol{A}_{S'}$ for all s' ($s' \leq s$) satisfy (19) with overwhelming probability. The smallest and largest eigenvalues of $\boldsymbol{A}_S^\top \boldsymbol{A}_S$ are

$$\lambda_{\min}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) = \min\left\{ \lambda_{\min}(X_{K_1}^\top X_{K_1}), \ldots, \lambda_{\min}(X_{K_p}^\top X_{K_p}) \right\},$$
$$\lambda_{\max}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) = \max\left\{ \lambda_{\max}(X_{K_1}^\top X_{K_1}), \ldots, \lambda_{\max}(X_{K_p}^\top X_{K_p}) \right\}.$$

Since the Gaussian ensemble $\boldsymbol{X}$ satisfies UUP according to lemma 1, for all subsets $K' \subset K$, we obviously have

$$\begin{aligned} \frac{1}{2} \cdot \frac{n}{p} &\leq \lambda_{\min}(\boldsymbol{X}_K^\top \boldsymbol{X}_K) \leq \lambda_{\min}(\boldsymbol{X}_{K'}^\top \boldsymbol{X}_{K'}) \\ &\leq \lambda_{\min}(\boldsymbol{X}_{K'}^\top \boldsymbol{X}_{K'}) \leq \lambda_{\max}(\boldsymbol{X}_K^\top \boldsymbol{X}_K) \leq \frac{3}{2} \cdot \frac{n}{p} \end{aligned}$$

with overwhelming probability. Thus, $\lambda_{\min}(\boldsymbol{A}_S^\top \boldsymbol{A}_S)$ and $\lambda_{\max}(\boldsymbol{A}_S^\top \boldsymbol{A}_S)$ satisfy that

$$
\begin{aligned}
\lambda_{\min}(\boldsymbol{X}_K^\top \boldsymbol{X}_K) &\leq \lambda_{\min}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) \\
&\leq \lambda_{\max}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) \leq \lambda_{\max}(\boldsymbol{X}_K^\top \boldsymbol{X}_K).
\end{aligned}
$$

Then, for all subsets $S' \subset S$, the block matrices of $\boldsymbol{A}_{S'}$ come from $\boldsymbol{X}_{K'}$ $(K' \subset K)$ and it holds that

$$
\begin{aligned}
\frac{1}{2} \cdot \frac{N}{P} &\leq \lambda_{\min}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) \leq \lambda_{\min}(\boldsymbol{A}_{S'}^\top \boldsymbol{A}_{S'}) \\
&\leq \lambda_{\min}(\boldsymbol{A}_{S'}^\top \boldsymbol{A}_{S'}) \leq \lambda_{\max}(\boldsymbol{A}_S^\top \boldsymbol{A}_S) \leq \frac{3}{2} \cdot \frac{N}{P},
\end{aligned}
$$

where $N = np$ and $P = p^2$. Therefore, $\boldsymbol{A}$ holds UUP with same probability of $\boldsymbol{X}$.

For this reason, we can tell that the sparse partial correlation obtained by LASSO is a near-optimal solution under CS and the brain network recovery based on the partial correlation can be theoretically guaranteed.

5 Numerical Implementation

5.1 Data Description

5.1.1 Subjects

There are twenty six children with ASD (24 boys, mean age: 6.0 ± 1.8 years) and eleven children with PedCon (8 boys, mean age: 9.73 ± 2.55 years). The ASD group, who was diagnosed by the Korean version of Autism Diagnostic Interview-Revised (K-ADI-R) [51], was recruited from Child and Adolescent Psychiatric Outpatient Clinic of Seoul National University Hospital, Seoul, South Korea. The pediatric controls comprised as children who were failed to meet the criteria of any psychiatric disorder and visited the clinic for IQ evaluation.

5.1.2 PET Image Data Acquisition

All PET scans were obtained from ECAT EXACT 47 (Siemens-CTI, Knoxville, USA) PET scanner with an intrinsic resolution of 5.2 mm FWHM. PET images were 47 contiguous planes with a thickness of 3.4 mm. After transmission scan measured by 68Ge rod sources for attenuation correction,

emission scan was administered. All participants were scanned under normal environmental noise of scanner room. Image reconstruction was performed using a filtered back-projection algorithm (Shepp-Logan filter at a cutoff frequency of 0.3 cycles/pixel as $128 \times 128 \times 47$ matrices of size 2.1 $\times$ 2.1 $\times$ 3.4 mm).

5.1.3 Preprocessing

All PET data were preprocessed using Statistical Parametric Mapping (SPM 2, University College of London, UK), implemented in the Matlab 6.5 (Mathworks Inc., USA) environment. After spatial normalization to Korean Standard template space, mean FDG uptake within ROIs was extracted using Statistical Probabilistic Anatomical Map-Korean version (SPAM-K) [52]. The values of FDG uptake were globally normalized to the individual's total gray matter mean count.

5.2 Reproducibility of Brain Network

We applied the brain network recovery method to the 97 ROIs extracted from FDG-PET data. To validate the recovered network, we checked the reproducibility of network obtained from different datasets using the leave-one-out cross-validation. Since the number of ASD and PedCon data was 26 and 11, respectively, the leave-one-out cross-validation produced 26 and 11 brain networks (partial correlation matrices) of ASD and PedCon. Then, we estimated the standard deviation of each partial correlation element in each group in order to observe its variation during the cross-validation process. If the standard deviation is less than 0.1, we say that the partial correlation is *reproducible* because it means that the partial correlation is consistent even though the given data set in the same group is varied [53]. The reproducible partial correlation was divided into nonzero and zero elements. So all elements of a partial correlation matrix in a group is classified into three classes : irreproducible (purple), zero (yellow) and nonzero (red) as illustrated in Fig. 3. When λ increases, the number of zero elements increases, i.e., the sparseness increases, but irreproducible elements decrease (for $\lambda = 1, 3$, the number of nonzero elements are less than 1% and the number of irreducible elements are near zero). The reproducible element can be thought as meaningful connectivity because it appears consistently for varying data sets under certain conditions such as ASD or PedCon. We conclude that the sparseness constraint helps to find the relevant connectivity and our penalized linear regression method for partial correlation estimation recovers the near-
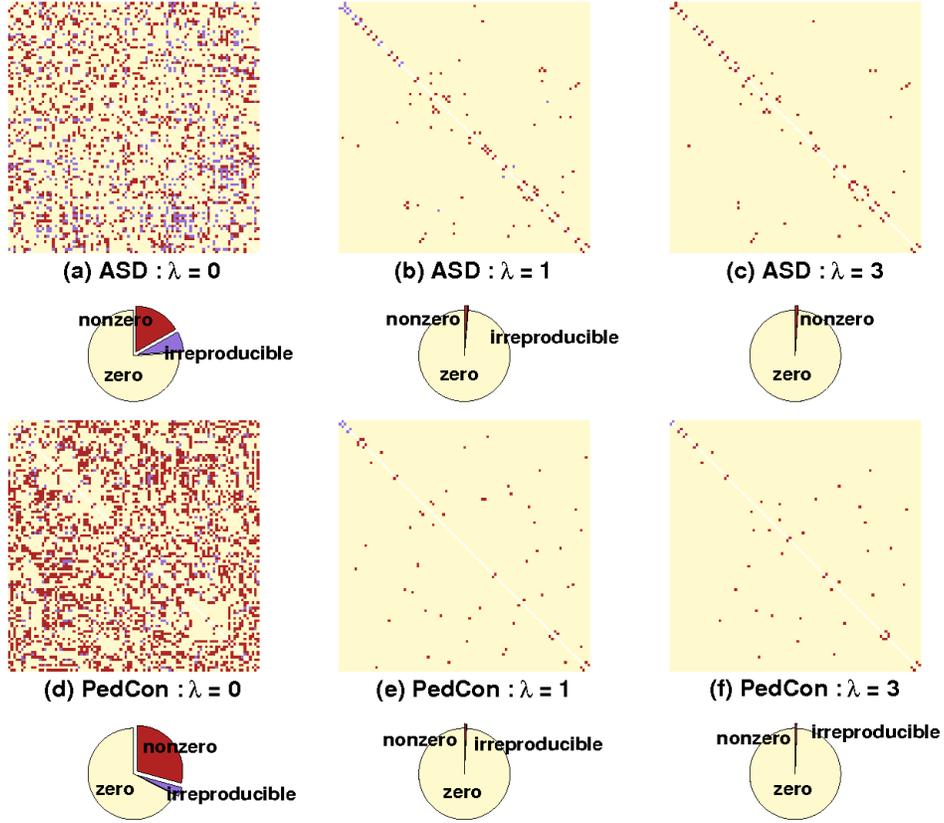
Figure 3: Partial correlation matrices of (a)-(c) ASD and (d)-(f) PedCon varying $\lambda = 0, 1, 3$ from left to right. We checked the reproducibility and sparseness (zero elements) of partial correlation during the leave-one-out cross validation. If the standard deviation of partial correlation obtained by the cross validation is more than 0.1, we call it irreproducible and represent it by the purple dots. We divided the reproducible partial correlation into zero and nonzero elements, which are represented by the yellow and red dots. The ratio of 3 kinds of elements is shown as the pie chart.

15

optimal brain network.

5.3 Visualization of Brain Network

A 3-dimensional brain network is difficult to visualize in 2 dimension. Therefore we embeded ROIs in the 3-dimensional image in Fig. 4(a) into the 2-dimensional space in Fig. 4(b) by ISOMAP, which is one of low dimensional embedding methods preserving the local structure such as the distance between neighbors [54]. The color of node in Fig. 4 (a) and (b) represents the lobe where the node belongs.

The well-known characteristics of ASD is local overconnectivity and global underconnectivity. ASD has been mostly compared with the normal control by connectivities between or within lobes [40, 38, 39, 55]. Hence, we hypothesized that (1) finding the modular structure of the estimated brain network helps to compare between ASD and PedCon and (2) the modular structure of ASD network is more similar to the lobe structure than one of PedCon network. In other words, the number of edges between lobes in ASD network is smaller than that of PedCon, while the number of edges in a lobe of ASD network is larger than that of PedCon. After calculating the threshold which maximizes the number of clusters and obtaining the thresholded partial correlation matrix, we partitioned the ROIs (nodes) of graph by the agglomerative hierarchical clustering [56]. Fig. 4 (c) and (d) show that the clustered brain networks of ASD and PedCon. The color of nodes and edge in Fig. 4 (c) and (d) represents the cluster where the nodes belong. The color of cluster is selected by blending the node colors, which are determined by their located lobes in Fig. 4 (a) and (b), in the cluster. So, if the cluster consists of nodes in the same lobe, the node color is identical to one of Fig. 4 (b). We can see that the clustered brain network of ASD is more similar to the lobe structure than the network of PedCon. In the next section, we seek the differences between ASD and PedCon networks via a statistical inference.

5.4 Significance Test for Finding Group Differences

After constructing the thresholded partial correlation matrices of ASD and PedCon, we estimated the number of (a) edges, (b) clusters, (c) edges connected between two ROIs in different lobes, and (d) edges connected between two ROIs in same lobe and (e) the sum of anatomical distances of edges

(a) ROIs in 3-dimensional space

(b) 2-dimensional embedding of ROIs
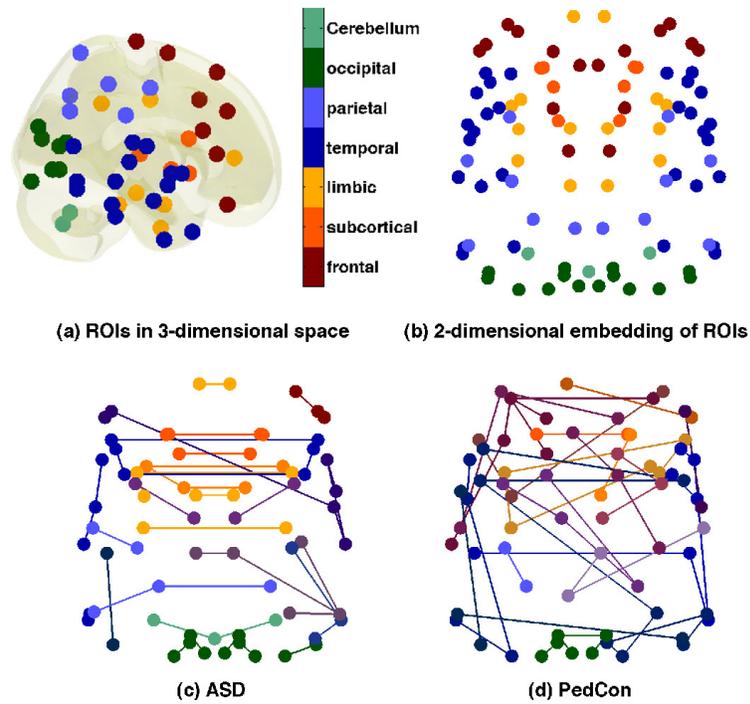
(c) ASD

(d) PedCon

Figure 4: Visualization of ROIs in (a) 3-dimensional and (b) 2-dimensional space. ROIs in the 3-dimensional space are embedded to the 2-dimensional space by ISOMAP, which preserves the distance between nodes during the low dimensional embedding. Each lobe is represented by different color as shown in colorbar (a). Clustered brain networks for (c) ASD and (d) PedCon. In these figures, the color represents the cluster.

in each network. Then, we compared them using the Wilcoxon rank sum test in Fig. 5. In each panel of Fig. 5, the red box is for ASD and the blue box is for pedCon. When they were different at significance level 0.01, we marked the asterisk (*) in each panel. Fig. 5 supports the hypothesis the the autistic brain network has local overconnectivity and long-range underconnectivity compared to the normal control [57]. Although the number of edges and clusters of ASD and PedCon were not different in (a) and (b), the number of edges connected between lobes and the sum of anatomical distances of edges were significantly larger than one of ASD in (c) and (e) ($p < 0.001$), while the number of edges connected within a lobe of PedCon was significantly smaller than one of ASD in (d) ($p < 0.001$).

We also estimated the number of edges connected between each pair of 7 lobes, frontal, subcortical, limbic, temporal, parietal, occipital and limbic lobes, and the number of edges connected in each lobe. Then, we compared them using the Wilcoxon rank sum test in Fig. 6. In each panel, the vertical axis represents the number of edges and the horizontal axis represents groups, ASD and PedCon. The red box and the blue box are for ASD and PedCon, respectively. The asterisk (*) was marked when $p < 0.01$. The smaller the p-value under each panel is, the more different the connectivity between ASD and PedCon is. According to Fig. 6, the local overconnectivity was found in ASD children than control children, especially within frontal, occipital, subcortical,temporal, and limbic lobes ($p < 0.01$). Also, long-range underconnectivity patterns between lobes were observed in ASD children as follows: frontal-parietal ($p < 0.001$), frontal-limbic ($p < 0.001$), parietal-occipital ($p < 0.05$), parietal-limbic ($p < 0.05$), parietal-subcortical ($p < 0.001$), occipital-temporal ($p < 0.001$), occipital-subcortical ($p < 0.01$), cerebellum-temporal ($p < 0.005$), temporal-limbic ($p < 0.001$), temporal-subcortical ($p < 0.001$), and limbic-subcortical ($p < 0.001$). Functional underconnectivity between frontal and parietal regions in ASD was quite consistent with other studies, because it is associated with deficits of planning and problem solving in ASD [41]. Frontal mirror neuron system was suggested to mediate understanding of the other's emotional states in concert with limbic center, such as amygdala. This emotional dysfunction in ASD children might be explained through the abnormal connectivity between frontal and limbic system [58]. The occipital regions showed reduced functional connectivity with the temporal regions, which was associated with impairment of mentalizing in ASD [59]. Abnormal behavioral phenotypes in ASD could be involved in these long-range dysconnectivities.
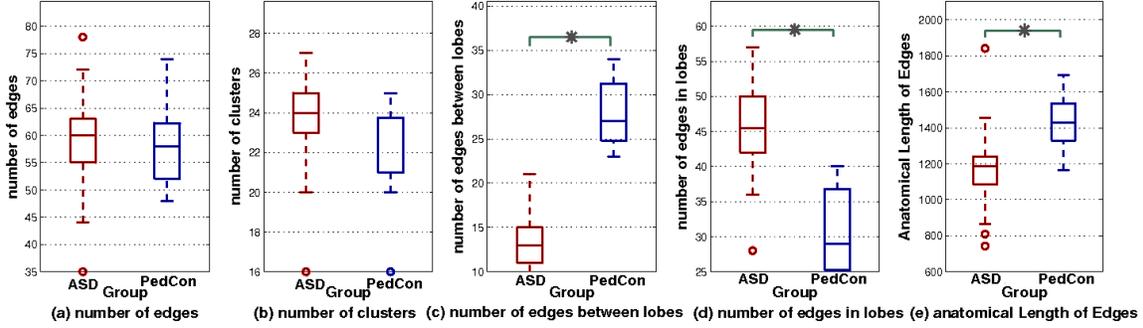
Figure 5: Significance tests on (a) the number of edges ($p = 0.5489$), (b) the number of clusters ($p = 0.0406$), (c) the number of edges connected between lobes ($p < 0.001$), (d) the number of edges connected in lobe ($p < 0.001$), and (e) sum of anatomical length of edges ($p < 0.001$). The red box in the left is for autism and the blue box in the right is for PedCon. The asterisk (*) represents the significant difference obtained by the Wilcoxon rank sum test with the significance level $p < 0.01$.

6 Conclusion and Discussion

In this paper, we showed that correlation and partial correlation can be formulated in the linear regression framework. Partial correlation is widely used for connectivity estimation, especially, for the highly correlated network, because it estimates true relationship between two nodes by factoring out the redundant dependences on other nodes. However, the main difficulty of brain network modeling based on partial correlation is that, because of the insufficient number of scans (small n) compared to the complexity of network (large p), the partial correlation is unreliably estimated. To remedy the small-n large-p problem, the penalty term for the sparseness constraint is usually introduced and added to the linear regression model. The penalized linear regression, a.k.a. LASSO, can naturally lead to the sparse brain network construction as we have shown. Under the CS defined in the linear model framework, UUP guarantees that the exact sparse signal recovery even from small number of noisy measurements. By appropriating UUP of Gaussian ensemble, we demonstrated that the proposed method can estimate partial correlations reliably and recover the sparse brain network. The numerical experiments show that the obtained sparse brain network is consistent for different datasets in the same group and our ASD network has local overconnectivity and long-range underconnectivity as already known in the autism-related clinical studies [38, 39, 55].
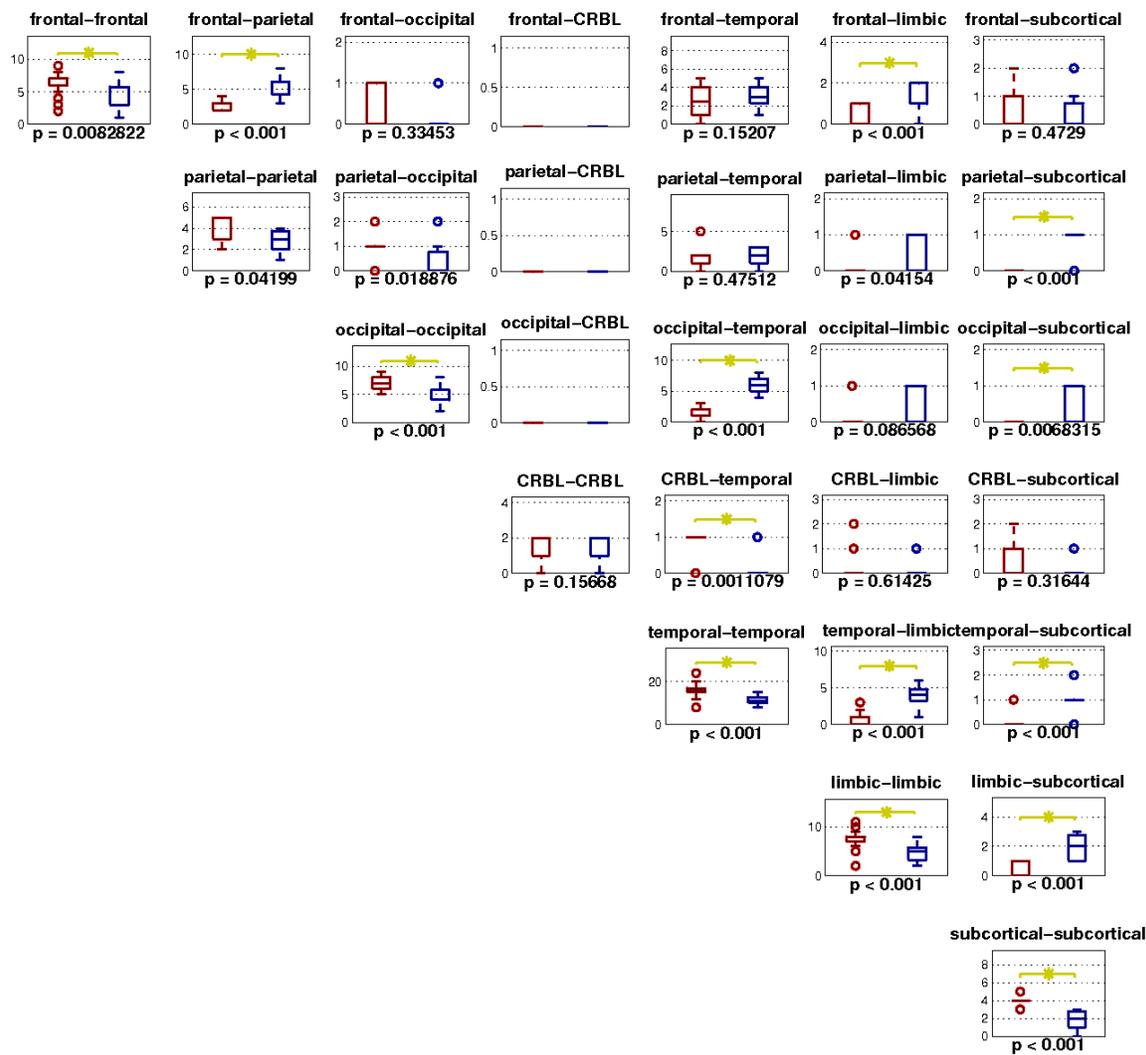
Figure 6: Significance on the number of edges within a lobe and between lobes. The lobes consist of frontal, parietal, limbic, occipital, subcortical and temporal lobes. In each panel, the vertical axis indicates the number of edges and the horizontal axis indicates groups (ASD and PedCon). The red box is for ASD and the blue box is for PedCon. Under each panel, there is the p-value obtained by the Wilcoxon rank sum test for measuring the differences between two groups. If the difference is significant, we indicated it with the asterisk (*).

Appendix

7 Proof of Correlation and Partial Correlation in linear model (6) and (8)

The loss function of correlation coefficient in (6) is given by

$$
\begin{aligned}
\boldsymbol{L}_{corr}(\alpha_{ij}, \boldsymbol{f}_i, \boldsymbol{f}_j) &= \parallel \boldsymbol{f}_i - \alpha_{ij}\boldsymbol{f}_j \parallel^2 \\
&= (\boldsymbol{f}_i - \alpha_{ij}\boldsymbol{f}_j)^\top (\boldsymbol{f}_i - \alpha_{ij}\boldsymbol{f}_j) \\
&= (\boldsymbol{f}_j^\top \boldsymbol{f}_j)\alpha_{ij}^2 - 2(\boldsymbol{f}_i^\top \boldsymbol{f}_j)\alpha_{ij} + (\boldsymbol{f}_i^\top \boldsymbol{f}_i).
\end{aligned}
\tag{20}
$$

The optimal solution minimizing (20) is easily obtained by

$$
\alpha_{ij} = \frac{\boldsymbol{f}_i^\top \boldsymbol{f}_j}{\boldsymbol{f}_j^\top \boldsymbol{f}_j} = \sqrt{\frac{\boldsymbol{f}_i^\top \boldsymbol{f}_i}{\boldsymbol{f}_j^\top \boldsymbol{f}_j}} \frac{\boldsymbol{f}_i^\top \boldsymbol{f}_j}{\sqrt{\boldsymbol{f}_i^\top \boldsymbol{f}_i \, \boldsymbol{f}_j^\top \boldsymbol{f}_j}} = \sqrt{\frac{\boldsymbol{f}_i^\top \boldsymbol{f}_i}{\boldsymbol{f}_j^\top \boldsymbol{f}_j}} \rho_{ij}.
$$

We assume that $\boldsymbol{f}_i^\top \boldsymbol{f}_i = 1$ for all $i = 1, \ldots, p$. Thus, α_{ij} is equal to ρ_{ij} in (2).

Minimizing the loss function of partial correlation in (7) is equal to minimizing the correlation between residuals and data vectors except the i-th data vector as follows :

$$
\min_{\boldsymbol{\beta}_i} \parallel \boldsymbol{f}_i - \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j \parallel^2 \Leftrightarrow
\begin{cases}
\mathrm{cov}(\boldsymbol{f}_i - \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j, \boldsymbol{f}_1) = 0 \\
\qquad\qquad \vdots \\
\mathrm{cov}(\boldsymbol{f}_i - \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j, \boldsymbol{f}_{i-1}) = 0 \\
\mathrm{cov}(\boldsymbol{f}_i - \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j, \boldsymbol{f}_{i+1}) = 0 \\
\qquad\qquad \vdots \\
\mathrm{cov}(\boldsymbol{f}_i - \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j, \boldsymbol{f}_p) = 0
\end{cases}
\tag{21}
$$

We can rewrite (21) to a matrix form in (6).

The covariance matrix $\boldsymbol{\Sigma}$ of data matrix $\boldsymbol{X} = [\boldsymbol{f}_1 \ldots \boldsymbol{f}_p]^\top$ is $\boldsymbol{\Sigma} = [\sigma_{ij}] = [\mathrm{cov}(\boldsymbol{f}_i, \boldsymbol{f}_j)]$. Then, the leftmost matrix in (6) is the matrix eliminating the i-th column and row vectors from $\boldsymbol{\Sigma}$, which is denoted by $\boldsymbol{\Sigma}_{\sim i}$. Exchanging the first row and column with the i-th row and column, we can

$$
\begin{bmatrix} \mathrm{cov}(\boldsymbol{f}_i, \boldsymbol{f}_1) \\ \vdots \\ \mathrm{cov}(\boldsymbol{f}_i, \boldsymbol{f}_{i-1}) \\ \mathrm{cov}(\boldsymbol{f}_i, \boldsymbol{f}_{i+1}) \\ \vdots \\ \mathrm{cov}(\boldsymbol{f}_i, \boldsymbol{f}_p) \end{bmatrix}^\top = \begin{bmatrix} \beta_{i1} \\ \vdots \\ \beta_{i,i-1} \\ \beta_{i,i+1} \\ \vdots \\ \beta_{ip} \end{bmatrix}^\top \begin{bmatrix} \mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_1) & \cdots & \mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_{i-1}) & \mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_{i+1}) & \cdots & \mathrm{cov}(\boldsymbol{f}_1, \boldsymbol{f}_p) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\boldsymbol{f}_{i-1}, \boldsymbol{f}_1) & \cdots & \mathrm{cov}(\boldsymbol{f}_{i-1}, \boldsymbol{f}_{i-1}) & \mathrm{cov}(\boldsymbol{f}_{i-1}, \boldsymbol{f}_{i+1}) & \cdots & \mathrm{cov}(\boldsymbol{f}_{i-1}, \boldsymbol{f}_p) \\ \mathrm{cov}(\boldsymbol{f}_{i+1}, \boldsymbol{f}_1) & \cdots & \mathrm{cov}(\boldsymbol{f}_{i+1}, \boldsymbol{f}_{i-1}) & \mathrm{cov}(\boldsymbol{f}_{i+1}, \boldsymbol{f}_{i+1}) & \cdots & \mathrm{cov}(\boldsymbol{f}_{i+1}, \boldsymbol{f}_p) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\boldsymbol{f}_p, \boldsymbol{f}_1) & \cdots & \mathrm{cov}(\boldsymbol{f}_p, \boldsymbol{f}_{i-1}) & \mathrm{cov}(\boldsymbol{f}_p, \boldsymbol{f}_{i+1}) & \cdots & \mathrm{cov}(\boldsymbol{f}_p, \boldsymbol{f}_p) \end{bmatrix}
\tag{6}
$$

rewrite the covariance matrix $\boldsymbol{\Sigma}$ as

$$
\boldsymbol{\Sigma} = \left[\begin{array}{c|ccccccc} \sigma_{ii} & \sigma_{i1} & \cdots & \sigma_{i,i-1} & \sigma_{i,i+1} & \cdots & \sigma_{ip} \\ \hline \sigma_{1i} & \sigma_{11} & \cdots & \sigma_{1,i-1} & \sigma_{1,i+1} & \cdots & \sigma_{1p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{i-1,i} & \sigma_{i-1,1} & \cdots & \sigma_{i-1,i-1} & \sigma_{i-1,i+1} & \cdots & \sigma_{i-1,p} \\ \sigma_{i+1,i} & \sigma_{i+1,1} & \cdots & \sigma_{i+1,i-1} & \sigma_{i+1,i+1} & \cdots & \sigma_{i+1,p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{pi} & \sigma_{p1} & \cdots & \sigma_{p,i-1} & \sigma_{p,i+1} & \cdots & \sigma_{pp} \end{array}\right]
$$

$$
= \left[\begin{array}{c|c} \sigma_{ii} & \boldsymbol{\sigma}_{i,\sim i} \\ \hline \boldsymbol{\sigma}_{\sim i,i} & \boldsymbol{\Sigma}_{\sim i} \end{array}\right],
$$

where $\boldsymbol{\sigma}_{i,\sim i} = [\sigma_{1i}\ \sigma_{11}\ \cdots\ \sigma_{1,i-1}\ \sigma_{1,i+1}\ \cdots\ \sigma_{1p}]$ and $\boldsymbol{\sigma}_{\sim i,i} = \boldsymbol{\sigma}_{i,\sim i}^\top$. The inverse of covariance matrix $\boldsymbol{\Sigma}$ is denoted by

$$
\boldsymbol{\Pi} = \boldsymbol{\Sigma}^{-1} = \left[\begin{array}{c|c} \pi_{ii} & \boldsymbol{\pi}_{i,\sim i} \\ \hline \boldsymbol{\pi}_{\sim i,i} & \boldsymbol{\Pi}_{\ i} \end{array}\right].
$$

According to the rule of inverse of a partitioned matrix,

$$
\begin{aligned}
\boldsymbol{\pi}_{i,\sim i} &= -(1 - \boldsymbol{\sigma}_{i,\sim i}\boldsymbol{\Sigma}_{\sim i}^{-1}\boldsymbol{\sigma}_{\sim i,i})^{-1}\boldsymbol{\sigma}_{i,\sim i}\boldsymbol{\Sigma}_{\sim i}^{-1} \\
&= -\pi_{ii}\boldsymbol{\sigma}_{i,\sim i}\boldsymbol{\Sigma}_{\sim i}^{-1}.
\end{aligned}
$$

Revisiting (6), we can write down in the matrix form :

$$
\boldsymbol{\sigma}_{i,\sim i} = \boldsymbol{\beta}_{i,\sim i}\boldsymbol{\Sigma}_{\sim i}.
$$

Then,

$$
\begin{aligned}
\boldsymbol{\beta}_{i,\sim i} &= \boldsymbol{\sigma}_{i,\sim i}\boldsymbol{\Sigma}_{\sim i}^{-1} \\
&= -\boldsymbol{\pi}_{i,\sim i}/\pi_{ii}.
\end{aligned}
$$

Therefore, (8) holds.

8 Algorithms for Estimating Sparse Correlation Coefficient and Partial Correlation in the Penalized Linear Model (16) and (17)

The coordinate descent learning and the active-set algorithm are applied to determine solutions to minimize (16) and (17), [27, 47].

Derivatives of objectivite function (16) with respect to the correlation coefficient ρ_i are given by

$$\frac{\partial(16)}{\partial\rho_{ij}} = -\boldsymbol{f}_i^\top \boldsymbol{f}_j + \rho_i + \lambda, \quad \text{for } \rho_{ij} > 0,$$

$$\frac{\partial(16)}{\partial\rho_{ij}} = -\boldsymbol{f}_i^\top \boldsymbol{f}_j + \rho_i - \lambda, \quad \text{for } \rho_{ij} < 0.$$

Then,

$$\rho_{ij} = \alpha_{ij} = \left[\boldsymbol{f}_i^\top \boldsymbol{f}_j, \lambda\right]_+,$$

where

$$[a, b]_+ = \begin{cases} a - b & \text{if } a > 0 \text{ and } |a| > b \\ a + b & \text{if } a < 0 \text{ and } |a| > b \\ 0 & \text{if } |a| \leq b \end{cases}.$$

Actually, the sparse correlation is same with the thresholded correlation with the thresholding constant λ. λ can be obtained by the piecewise linear solution paths [60].

Derivatives of objective function (17) for the partial correlation with respect to θ_{ij} fixed π_{ii} are given by

$$\frac{\partial(17)}{\partial\theta_{ij}} = \frac{1}{n}\sum_{t=1}^{n}\left(f_{ti} - \sum_{k\neq i,j}\theta_{ik}\sqrt{\frac{\pi_{kk}}{\pi_{ii}}}f_{tk} - \theta_{ij}\sqrt{\frac{\pi_{jj}}{\pi_{ii}}}f_{tj}\right)\left(-\sqrt{\frac{\pi_{jj}}{\pi_{ii}}}f_{tj}\right) + \lambda$$

$$= \left(\frac{\pi_{jj}}{n\pi_{ii}}\boldsymbol{f}_j^\top\boldsymbol{f}_j\right)\theta_{ij} - \frac{1}{n}\sqrt{\frac{\pi_{jj}}{\pi_{ii}}}\boldsymbol{f}_i^\top\boldsymbol{f}_j + \frac{w_i}{n}\sum_{k\neq i,j}\theta_{ik}\frac{\sqrt{\pi_{kk}\pi_{jj}}}{\pi_{ii}}\boldsymbol{f}_k^\top\boldsymbol{f}_j + \lambda$$

for $\theta_{ij} > 0$, and

$$\frac{\partial(17)}{\partial\theta_{ij}} = \frac{1}{n}\sum_{t=1}^{n}\left(f_{ti} - \sum_{k\neq i,j}\theta_{ik}\sqrt{\frac{\pi_{kk}}{\pi_{ii}}}f_{tk} - \theta_{ij}\sqrt{\frac{\pi_{jj}}{\pi_{ii}}}f_{tj}\right)\left(-\sqrt{\frac{\pi_{jj}}{\pi_{ii}}}f_{tj}\right) - \lambda$$

$$= \left(\frac{\pi_{jj}}{n\pi_{ii}}\boldsymbol{f}_j^\top\boldsymbol{f}_j\right)\theta_{ij} - \frac{1}{n}\sqrt{\frac{\pi_{jj}}{\pi_{ii}}}\boldsymbol{f}_i^\top\boldsymbol{f}_j + \frac{1}{n}\sum_{k\neq i,j}\theta_{ik}\frac{\sqrt{\pi_{kk}\pi_{jj}}}{\pi_{ii}}\boldsymbol{f}_k^\top\boldsymbol{f}_j - \lambda.$$

for $\theta_{ij} < 0$. Then,

$$\theta_{ij} = \frac{\left[\frac{1}{n}\left(\boldsymbol{f}_i^\top \boldsymbol{f}_j - \sum_{k \neq i,j} \theta_{ik} \frac{\sqrt{\pi_{kk}\pi_{jj}}}{\pi_{ii}} \boldsymbol{f}_k^\top \boldsymbol{f}_j\right), \lambda\right]_+}{\frac{\pi_{jj}}{n\pi_{ii}} \boldsymbol{f}_j^\top \boldsymbol{f}_j}. \tag{22}$$

The outline of algorithm for the sparse inverse covariance estimation using the linear regression is as follows :

Input : $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{f}_1 ... \boldsymbol{f}_p \end{bmatrix} \in \mathbb{R}^{n \times p}$, λ

Output : $\boldsymbol{\theta} \in \mathbb{R}^{p \times p}$, $\boldsymbol{\pi} \in \mathbb{R}^p$

Step 1 Normalize the data matrix $\boldsymbol{X}$.

   Initialize $[\theta_{ij}]_{i,j=1,...,p(i<j)}$ and $[\pi_{ii}]_{i=1,...,p}$.

Step 2 For $i = 1, ..., p$ and $j = i + 1, ..., p$,

   Update θ_{ij} in (22).

   Repeat until convergence.

Step 3 For $i = 1, ..., p$,

   Estimate $\boldsymbol{\epsilon}_i = \boldsymbol{f}_i - \sum_{j \neq i} \beta_{ij} \boldsymbol{f}_j$ and $\pi_{ii} = 1/\mathrm{var}(\boldsymbol{\epsilon}_i)$.

Step 4 Repeat step 2 and 3 until convergence.

We assume that $\pi_{ii} = 1$ $(i = 1, ..., p)$ for computational simplicity. To reduce computational complexity, Friedman and Peng exploited the active set algorithm [27, 47]. Step 2 of above algorithm is changed to as follows :

Step 2* Active set algorithm

2.1 Construct the current active set $\boldsymbol{\Lambda} = \{(i,j) |$ current $\theta_{ij} \neq 0\}$.

   2.1.1 Update θ_{ij} $(i,j) \in \boldsymbol{\Lambda}$ in (22).

   2.1.2 Repeat updating until convergence.

2.2 For $i = 1, ..., p$ and $j = i + 1, ..., p$, update θ_{ij} in (22).

2.3 Repeat Step 2* until convergence.

Acknowledgment

References

[1] O. Sporn and J. Zwi, "The small world of the cerebral cortex," *Neuroinformatics*, vol. 2, pp. 145–162, 2004.

[2] V. Eguiluz, D. Chialvo, G. Cecchi, M. Baliki, and A. Apkarian, "Scale-free brain functional networks," *Physical Review Letters*, vol. 94, p. 018102, 2005.

[3] C. Stam, "Functional connectivity patterns of human magnetoencephalographic recordings : a small-world network?" *Neuroscience Letters*, vol. 355, pp. 25–28, 2004.

[4] C. Stam, B. Jones, G. Nolte, M. Breakspear, and P. Scheltens, "Small-world networks and functional connectivity in alzheimers disease," *Cerebral Cortex*, vol. 17, pp. 92–99, 2007.

[5] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *Journal of Neuroscience*, vol. 26, pp. 63–72, 2006.

[6] R. Salvador, J. Suckling, M. Coleman, J. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, pp. 1332–1342, 2005.

[7] Y. He, Z. Chen, and A. Evans, "Small-world anatomical networks in the human brain revealed by cortical thickness from MRI," *Cerebral Cortex*, vol. 170, pp. 2407–2419, 2007.

[8] Z. Chen, Y. He, P. Rosa-Neto, J. Germann, and E. A.C., "Revealing modular architecture of human brain structural networks by using cortical thickness from MRI," *Cerebral Cortex*, pp. 2374–2381, 2008.

[9] M. Valencia, M. Pastor, M. Fernandez-Seara, J. Artieda, J. Martinerie1, and M. Chavez, "Complex modular structure of large-scale brain networks," *Chaos*, vol. 19, 2009.

[10] P. Laurienti, C. Hugenschmidt, and S. Hayasaka, "Modularity maps reveal community structure in the resting human brain," *Nature Precedings*, 2009.

[11] D. Bassett, "Small-world brain networks," *The Neuroscientist*, vol. 12, pp. 512–523, 2006.

[12] M. Rubinov and S. O., "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, p. in press, 2009.

[13] O. Sporns, G. Tononi, and R. Kotter, "The human connectome: a structural description of the human brain," *PLoS Computational Biology*, vol. 1, 2005.

[14] J. Cao and K. Worsley, "The geometry of correlation fields with an application to functional connectivity of the brain," *Annals of Applied Probability*, vol. 9, pp. 1021–1057, 1999.

[15] M. Koch, D. Norris, and M. Hund-Georgiadis, "An investigation of functional and anatomical connectivity using magnetic resonance imaging," *NeuroImage*, vol. 16, pp. 241–250, 2002.

[16] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, and H. Benalia, "Partial correlation for functional brain interactivity investigation in functional MRI," *NeuroImage*, vol. 32, 2006.

[17] A. McIntosh, F. Bookstein, J. Haxby, and C. Grady, "Spatial pattern analysis of functional brain images using partial least squares," *NeuroImage*, vol. 3, pp. 143–157, 1996.

[18] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *Annals of Statistics*, vol. 34, pp. 1436–1462, 2006.

[19] J. Schäfer and K. Strimmer, "An empirical bayes approach to inferring large-scale gene association networks," *Bioinformatics*, vol. 21, pp. 754–764, 2005.

[20] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *Journal of Machine Learning Research*, vol. 9, pp. 485–516, 2008.

[21] P. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Annals of Statistics*, vol. 36, pp. 199–227, 2008.

[22] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, pp. 432–441, 2008.

[23] L. Sun, R. Patel, J. Liu, K. Chen, T. Wu, J. Li, E. Reiman, and Y. Ye, "Mining brain region connectivity for alzheimer's disease study via sparse inverse covariance estimation," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2009.

[24] G. Cecchi, I. Rish, B. Thyreau, B. Thirion, M. Plaze, M. Paillere-Martinot, C. Martelli, J. Martinot, and J. Poline, "Discriminative network models of schizophrenia," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.

[25] S. Huang, J. Li, L. Sun, J. Liu, T. Wu, K. Chen, A. Fleisher, E. Reiman, and J. Ye, "Learning brain connectivity of alzheimer's disease from neuroimaging data," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.

[26] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS Computational Biology*, vol. 3, 2007.

[27] J. Peng, P. Wang, N. Zhou, and J. Zhu, "Partial correlation estimation by joint sparse regression models," *Journal of the American Statistical Association*, vol. 104, pp. 735–746, 2009.

[28] O. Banerjee, L. Ghaoui, A. d'Aspremont, and G. Natsoulis, "Convex optimization techniques for fitting sparse gaussian graphical models," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2006.

[29] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1999.

[30] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society B*, vol. 58, pp. 267–288, 1996.

[31] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, pp. 21–30, 2008.

[32] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, 2006.

[33] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, pp. 489–509, 2006.

[34] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, 2007.

[35] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Information Theory*, pp. 5406–5425, 2006.

[36] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, pp. 1207–1223, 2006.

[37] N. Meinshausen and B. Yu, "Lasso-type recovery of sparse representations for high-dimensional data," Department of Statistics, UC Berkeley, Tech. Rep. 720, 2006.

[38] M. Murias, S. Webb, J. Greenson, and D. G., "Resting state cortical connectivity reflected in EEG coherence in individuals with autism," *Biological Psychiatry*, vol. 62, pp. 270–273, 2007.

[39] D. Williams, G. Goldstein, and N. Minshew, "Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing," *Child Neuropsychol.*, pp. 279–298, 2006.

[40] E. Courchesne, K. Pierce, C. Schumann, E. Redcay, J. Buckwalter, D. Kennedy, and J. Morgan, "Mapping early brain development in autism," *Neuron*, vol. 56, pp. 399–413, 2007.

[41] M. Just, V. Cherkassky, T. Keller, and N. Minshew, "Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity," *Brain*, vol. 127, pp. 1811–1821, 2004.

[42] W. L., *All of Statistics*. Springer, 2004.

[43] D. Bassett and E. Bullmore, "Human brain networks in health and disease," *Current Opinion in Neurology*, vol. 22, pp. 340–347, 2009.

[44] A. Laub, *Matrix Analysis for Scientists and Engineers*. SIAM, 2004.

[45] Y. Sharon, J. Wright, and Y. Ma, "Computation and relaxation of conditions for equivalence between l_1 and l_0 minimization," UIUC, Tech. Rep. UILU-ENG-07-2008, 2007.

[46] E. Candès, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," 2005, pp. 295–308.

[47] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Stanford University, Tech. Rep., 2008.

[48] D. Hsu, S. Kakade, J. Langford, and T. Zhang, "Multi-label prediction via compressed sensing," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.

[49] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, pp. 589–592, 2008.

[50] M. Chung, K. Worsley, S. Robbins, T. Paus, J. Taylor, J. Giedd, J. Rapoport, and A. Evans, "Deformation-based surface morphometry applied to gray matter deformation," *NeuroImage*, vol. 18, pp. 198–213, 2003.

[51] H. Yoo, *Korean version of Autism Diagnostic Interview-Revised.* Seoul: Hakjisa, 2007.

[52] J. Lee, D. Lee, Y. Kim, J. Kim, J. Lee, B. Koo, J. Kim, J. Kwon, T. Yoo, K. Chang, S. Kim, H. Kang, E. Kang, J. Chung, and M. Lee, "Quantification of brain images using korean standard templates and structural and cytoarchitectonic probabilistic maps," *Korean Journal of Nuclear Medicine*, vol. 38, pp. 241–252, 2004.

[53] "IUPAC compendium of chemical terminology – the gold book," 2009. [Online]. Available: http://www.iupac.org/goldbook

[54] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," pp. 2319–2323, 2000. [Online]. Available: http://isomap.stanford.edu/

[55] N. Minshew and D. Williams, "The new neurobiology of autism: Cortex, connectivity, and neuronal organization," *Arch Neurol.*, pp. 945–950, 2007.

[56] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning.* New York: Springer, 2009.

[57] M. Belmonte, G. Allen, A. Beckel-Mitchener, L. Boulanger, R. Carper, and S. Webb, "Autism and abnormal development of brain connectivity," *Journal of Neuroscience*, vol. 24, pp. 9228–9231, 2004.

[58] M. Dapretto, M. Davies, J. Pfeifer, A. Scott, M. Sigman, S. Bookheimer, and M. Iacoboni, "Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders," *Nature Neuroscience*, pp. 28–30, 2005.

[59] F. Castelli, C. Frith, F. Happé, and U. Frith, "Autism, asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes," *Brain*, pp. 1839–1849, 2002.

[60] S. Rosset and J. Zhu, "Piecewise linear regularized solution paths," *Annals of Statistics*, vol. 35, pp. 1012–1030, 2007.