Journal of Machine Learning Research 17 (2016) 1-21 Submitted 1/16; Revised 5/16; Published 6/16

Equivalence of Graphical Lasso and Thresholding for Sparse Graphs

Somayeh Sojoudi, SOJOUDI@BERKELEY.EDU
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Editor: Michael Mahoney

Abstract

This paper is concerned with the problem of finding a sparse graph capturing the conditional dependence between the entries of a Gaussian random vector, where the only available information is a sample correlation matrix. A popular approach to address this problem is the graphical lasso technique, which employs a sparsity-promoting regularization term. This paper derives a simple condition under which the computationally-expensive graphical lasso behaves the same as the simple heuristic method of thresholding. This condition depends only on the solution of graphical lasso and makes no direct use of the sample correlation matrix or the regularization coefficient. It is proved that this condition is always satisfied if the solution of graphical lasso is close to its first-order Taylor approximation or equivalently the regularization term is relatively large. This condition is tested on several random problems, and it is shown that graphical lasso and the thresholding method lead to highly similar results in the case where a sparse graph is sought. We also conduct two case studies on brain connectivity networks of twenty subjects based on fMRI data and the topology identification of electrical circuits to support the findings of this work on the similarity of graphical lasso and thresholding.

Keywords: Graphical Lasso, Graphical Models, Sparse Graphs, Brain Connectivity Networks, Electrical Circuits

1. Introduction

In recent years, there has been a growing interest in developing techniques for estimating sparse undirected graphical models (Banerjee et al., 2008; Bruckstein et al., 2009; Chandrasekaran et al., 2010; Goldstein and Osher, 2009; Jalali et al., 2011; Schmidt et al., 2007). Finding sparse solutions have become essential to many applications, including signal processing, pattern recognition, and data mining. Many applications use L_1-regularized models such as the Lasso (Tibshirani, 1996). L_1 regularization aims to find sparse solutions, which is especially useful for high-dimensional problems with a large number of features (Bühlmann and Van De Geer, 2011; Fan and Lv, 2010; Meinshausen and Yu, 2009; Richtárik and Takáč, 2012; Wright et al., 2009; Zhang and Huang, 2008). Although Lasso-type algorithms are shown to be effective in recovering sparse solutions, they are computationally expensive for large-scale problems. In this work, we derive a simple condition under which the computationally-expensive graphical lasso behaves the same as the simple heuristic method of thresholding.

Consider a random vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ with a multivariate normal distribution. Let $\Sigma_* \in \mathbb{S}_+^n$ denote the correlation matrix associated with the vector $\mathbf{x}$. The inverse of the correlation matrix can be used to determine the conditional independence between the random variables $x_1, x_2, ..., x_n$. In particular, if the $(i,j)^{\text{th}}$ entry of Σ_*^{-1} is zero, then x_i and x_j are conditionally independent. The graph $\mathcal{G}(\Sigma_*^{-1})$ (i.e., the sparsity graph of Σ_*^{-1}) represents a graphical model capturing the conditional independence between the elements of $\mathbf{x}$. Assume that $\mathcal{G}(\Sigma_*^{-1})$ is a sparse graph. Finding this graph is nontrivial in practice because the exact correlation matrix Σ_* is rarely known. More precisely, $\mathcal{G}(\Sigma_*^{-1})$ should be constructed from a given sample correlation matrix as opposed to Σ_*. Let Σ denote an arbitrary $n \times n$ positive semidefinite matrix, which is provided as an estimate of Σ_*. In this paper, we do not impose any assumption on the error $\|\Sigma - \Sigma_*\|$. Consider the convex optimization problem

$$\underset{S \in \mathbb{S}_+^n}{\text{minimize}} \quad -\log\det(S) + \text{trace}(\Sigma S) \tag{1}$$

where $\mathbb{S}_+^n$ denotes the set of $n \times n$ positive semidefinite matrices. It is easy to verify that the optimal solution of the above problem is $S^{\text{opt}} = \Sigma^{-1}$. Hence, S^{opt} aims to estimate Σ_*^{-1}. On the other hand, although the inverse of Σ_* is assumed to be a sparse graph, a small perturbation of Σ_* would make its inverse a dense graph. This implies that the sparsity graph of S^{opt}, denoted as $\mathcal{G}(S^{\text{opt}})$, may not resemble the graphical model $\mathcal{G}(\Sigma_*^{-1})$ in general. Hence, the optimization problem (1) needs to be modified to indirectly enforce some sparsity on its solution. To this end, consider the problem

$$\underset{S \in \mathbb{S}_+^n}{\text{minimize}} \quad -\log\det(S) + \text{trace}(\Sigma S) + \lambda\|S\|_1 \tag{2}$$

where $\lambda \in \mathbb{R}_+$ is a regularization parameter and $\|S\|_1$ is defined as $\sum_{i=1}^n \sum_{j=1}^n |S_{ij}|$. This problem is referred to as *graphical lasso* (Banerjee et al., 2008; Friedman et al., 2008; Yuan and Lin, 2007). Intuitively, the penalty term $\lambda\|S\|_1$ with a large λ aims to decrease the off-diagonal entries of S in magnitude and enforce most of them to be zero. Henceforth, the notation S^{opt} is used to denote a solution of the graphical lasso instead of the unregularized optimization problem (1). There is a large body of literature suggesting that $\mathcal{G}(S^{\text{opt}})$ is a good estimate of the graphical model $\mathcal{G}(\Sigma_*^{-1})$ for a suitable choice of λ (Banerjee et al., 2008; Danaher et al., 2014; Friedman et al., 2008; Krämer et al., 2009; Liu et al., 2010; Witten et al., 2011; Yuan and Lin, 2007). Note that although graphical lasso is motivated by multivariate normal random variables, its application is beyond this class of random variables and it applies to all problems for which a *sparse* inverse correlation matrix is sought.

Suppose that it is known *a priori* that the true graph $\mathcal{G}(\Sigma_*^{-1})$ has k edges, for some given number k. Assume that the nonzero entries of the upper triangular part of Σ (excluding its diagonal) have different magnitudes (this assumption is satisfied both generically and under an infinitesimal perturbation of the nonzero entries of Σ). Two heuristic methods for finding an approximation of $\mathcal{G}(\Sigma_*^{-1})$ are as follows:

- *Graphical Lasso:* We solve the optimization problem (2) repeatedly for different values of λ until a solution S^{opt} with exactly $2k$ nonzero off-diagonal entries are found. Note that λ can be updated in the optimization problem using the bisection technique (Theorem 3 in Fattahi and Lavaei (2016) guarantees the existence of an appropriate interval for λ under generic conditions).

- *Thresholding:* Without solving any optimization problem, we simply identify those $2k$ entries of Σ that have the largest magnitudes among all off-diagonal entries of Σ. We then replace the remaining $n^2 - n - 2k$ off-diagonal entries of Σ with zero and denote the resulting matrix as Σ_k. Note that Σ and Σ_k have the same diagonal. Finally, we consider the sparsity graph of Σ_k, namely $\mathcal{G}(\Sigma_k)$, as an estimate of $\mathcal{G}(\Sigma_*^{-1})$.

The connection between graphical lasso and the thresholding technique is not well understood and there are only a few studies on this subject. The work Mazumder and Hastie (2012) has recently shown that if the sample covariance matrix is thresholded at λ and its corresponding graph is decomposed into connected components, then the vertex-partition induced by these components is equal to the one induced by the connected components of the estimated graph obtained from graphical lasso for the same λ. The paper Guillot and Rajaratnam (2011) obtains graph conditions that are required for preserving the positive definiteness of the sample correlation matrix after thresholding.

This paper is focused on the investigation of the connection between graphical lasso and thresholding. First, we derive a condition under which the heuristic thresholding method performs very similarly to the computationally-heavy graphical lasso. We then argue that this condition is satisfied as long as λ is large enough. Moreover, we demonstrate in numerical examples that graphical lasso and thresholding lead to the same approximate graph for $\mathcal{G}(\Sigma_*^{-1})$. Note that although the condition provided here depends on the solution of graphical lasso, it can be systematically expressed in terms of the sample correlation matrix Σ for certain types of graphs (see the technical report Sojoudi (2016) for the derivation of such conditions for acyclic graphs).

Recently, there has been a significant interest in studying the human brain functional connectivity networks using functional MRI (fMRI) data. Functional connectivity is measured as the temporal coherence or correlation between the activities of disjoint brain areas, where the direct statistical dependence between every two brain regions in the functional network can be obtained using partial correlation. In most fMRI studies, computing partial correlations is a daunting challenge due to the limitation on the number samples available from fMRI scans. Graphical lasso has become popular in the literature for the identification of the direct correlations between the activities of different parts of the brain using a small number of samples (Huang et al., 2010). In this work, we apply graphical lasso and thresholding to the resting-state fMRI data collected from twenty subjects and observe a high degree of similarity between the outcomes of these two techniques for each individual subject. Note that the matrix Σ is not invertible for the fMRI study conducted here. More precisely, this work makes no assumption on the invertibility of the sample correlation matrix Σ (although the true correlation matrix Σ_* needs to be invertible).

In Sojoudi and Doyle (2014), we have developed a new method for generating synthetic data based on sparse electrical circuit models, where certain nodes of each circuit are connected to one another through resistors and capacitors that are subject to thermal noise (to create stochasticity). The main property of this model is that the connectivity of the circuit model can be obtained from the sparsity pattern of the inverse covariance matrix associated with the nodal voltages. In other words, the sparsity pattern of the inverse covariance matrix has the same structure as the adjacency matrix of the circuit network. In this work, we use the above circuit model to verify the high similarity between graphical lasso and the thresholding technique for electrical networks.

This paper is organized as follows. The main results are presented in Section 2. Simulations on random systems are provided in Section 3. Two case studies on fMRI data and electrical circuits are conducted in Sections 4 and 5, respectively. Some concluding remarks are drawn in Section 6.

1.1 Notations and Definitions

Notations: $\mathbb{R}$, $\mathbb{S}^n$, and $\mathbb{S}^n_+$ denote the sets of real numbers, $n \times n$ (real) symmetric matrices, and $n \times n$ positive semidefinite matrices, respectively. $\text{trace}\{M\}$ and $\log\det\{M\}$ denote the trace and the logarithm of the determinant of a matrix M. The (i, j) entry of M is shown as M_{ij}. The notations $|x|$ and $\|M\|_1$ represent the absolute value of a scalar x and the sum of absolute values of a matrix M, respectively. The symbol $\text{sign}(\cdot)$ denotes the sign function (note that $\text{sign}(0) = 0$). The standard basis vectors in $\mathbb{R}^n$ are shown as $e_1, e_2, ..., e_n$. The optimal value of a matrix variable M is denoted as M^{opt}.

Definition 1 *Given a symmetric matrix $S \in \mathbb{S}^n$, the support (sparsity) graph of S is defined as a graph with the vertex set $\mathcal{V} := \{1, 2, ..., n\}$ and the edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ such that $(i, j) \in \mathcal{V}$ if and only if $S_{ij} \neq 0$, for every two different vertices $i, j \in \mathcal{V}$. The support graph of S captures the sparsity of the matrix S and is denoted as $\mathcal{G}(S)$.*

Definition 2 *Given two graphs $\mathcal{G}$ and $\mathcal{G}'$ with the same vertex set, define $\mathcal{G} - \mathcal{G}'$ as a graph obtained from $\mathcal{G}$ by removing the common edges of $\mathcal{G}$ and $\mathcal{G}'$.*

2. Main Results

In this section, we study the connection between thresholding and graphical lasso. To simplify the presentation, we assume that the nonzero entries of the upper triangular part of Σ (excluding its diagonal) have different magnitudes. Assume also that the number of such nonzero entries is greater than or equal to k (in order to guarantee that the thresholding method is able to obtain a graph with k edges). For notational convenience, we denote the $(i, j)^{\text{th}}$ entry of the matrix $(S^{\text{opt}})^{-1}$ as $(S^{\text{opt}})^{-1}_{ij}$ throughout this work.

2.1 Optimality Conditions

Consider the convex optimization problem (2) together with an optimal solution S^{opt}. First, we aim to obtain necessary and sufficient optimality conditions for graphical lasso.

Lemma 3 *S^{opt} is an optimal solution of graphical lasso if and only if the conditions*

$$(S^{opt})^{-1}_{ij} = \Sigma_{ij} + \lambda \qquad \text{if} \quad i = j \tag{3a}$$
$$(S^{opt})^{-1}_{ij} = \Sigma_{ij} + \lambda \times \operatorname{sign}(S^{opt}_{ij}) \qquad \text{if} \quad S^{opt}_{ij} \neq 0 \tag{3b}$$
$$(S^{opt})^{-1}_{ij} \leq \Sigma_{ij} + \lambda \qquad \text{if} \quad S^{opt}_{ij} = 0 \tag{3c}$$
$$(S^{opt})^{-1}_{ij} \geq \Sigma_{ij} - \lambda \qquad \text{if} \quad S^{opt}_{ij} = 0 \tag{3d}$$

are satisfied for all indices $i, j \in \{1, ..., n\}$.

Proof Due to the convexity of graphical lasso, a locally optimal solution of this problem is a global solution and, therefore, a local perturbation analysis can be used to prove the lemma. To this end, notice that $-\log(0) = +\infty$, $\operatorname{trace}(\Sigma S) \geq 0$, and $\|S\|_1$ is finite only when all entries of S are finite. It follows from these properties that S^{opt} has bounded entries and a nonzero determinant, i.e., $S^{\text{opt}} \succ 0$ (note that a zero determinant for S^{opt} makes the objective function of graphical lasso equal to $+\infty$). This means that $S^{\text{opt}}(\varepsilon; i, j)$ defined as $S^{\text{opt}} + \varepsilon(e_i e_j^T + e_j e_i^T)$ is a feasible solution of the optimization problem (2) for small values of ε (note that the matrix $e_i e_j^T + e_j e_i^T$ is sign indefinite). On the other hand, for every $i, j \in \{1, ..., n\}$, one can write:

$$\begin{aligned} &\big[-\log\det(S^{\text{opt}}) + \operatorname{trace}(\Sigma S^{\text{opt}}) + \lambda\|S^{\text{opt}}\|_1\big] \\ &- \big[-\log\det(S^{\text{opt}}(\varepsilon; i, j)) + \operatorname{trace}(\Sigma S^{\text{opt}}(\varepsilon; i, j)) + \lambda\|S^{\text{opt}}(\varepsilon; i, j)\|_1\big] \\ &\qquad = 2\left((S^{\text{opt}})^{-1}_{ij} - \Sigma_{ij}\right)\varepsilon + 2\left(|S^{\text{opt}}_{ij}| - |S^{\text{opt}}_{ij} + \varepsilon|\right)\lambda + \tau\varepsilon^2 + O(\varepsilon^3) \end{aligned} \tag{4}$$

where τ is a positive number due to the strict convexity of the objective of the optimization problem (2). The above equation is derived based on the Taylor series expansion of $\log\det(\cdot)$ and the fact that the derivative of $\log\det(S)$ with respect to S is equal to S^{-1}. Now, recall that S^{opt} is an optimal solution of (2), whereas $S^{\text{opt}} + \varepsilon(e_i e_j^T + e_j e_i^T)$ is only a feasible solution. Hence, the left side of the equality (4) must always be non-positive for all sufficiently small values of ε. A simple analysis of this equation leads to the conditions provided in (3). ■

Lemma 3 offers a set of necessary and sufficient conditions for the matrix S^{opt} to be an optimal solution of graphical lasso. These optimality conditions can be summarized as:

- The diagonal of $(S^{\text{opt}})^{-1}$ is obtained from that of Σ after a shift by the number λ.
- Each off-diagonal element (i, j) of the matrix $(S^{\text{opt}})^{-1}$ is in the interval $[\Sigma_{ij} - \lambda, \Sigma_{ij} + \lambda]$, and is located at one of the endpoints of this interval if (i, j) is an edge of the graphical model $\mathcal{G}(S^{\text{opt}})$.

2.2 First Condition for Equivalence

In this part, we derive a condition under which graphical lasso and thresholding result in the same estimate graphical model for the random vector $\mathbf{x}$.

Definition 4 *A symmetric matrix $\hat{S}$ is said to be equivalent to S^{opt} if $\hat{S}$ can be obtained from S^{opt} through two operations: (i) permutation of its off-diagonal entries, and (ii) flipping the sign of some of its off-diagonal entries. We use the notation $\hat{S} \sim S^{opt}$ to show this equivalence.*

Theorem 5 *Let k denote the number of edges of the graph $\mathcal{G}(S^{opt})$. Consider the optimization problems*

$$\underset{S\in\mathbb{S}^n}{\text{minimize}} \quad \text{trace}(\Sigma S) + \lambda\|S\|_1 \quad \text{subject to} \quad \mathrm{S} \sim \mathrm{S}^{opt} \tag{5}$$

and

$$\underset{S\in\mathbb{S}^n_+}{\text{minimize}} \quad -\log\det(S) + \text{trace}(\Sigma S) + \lambda\|S\|_1 \quad \text{subject to} \quad \mathrm{S} \sim \mathrm{S}^{opt} \tag{6}$$

If these two problems possess the same solution, then S^{opt} and Σ_k will have the same support graph.

Proof Let $\hat{S}^{\text{opt}}$ denote an optimal solution of the optimization problem (5). Our first goal is to show that $\mathcal{G}(\hat{S}^{\text{opt}}) = \mathcal{G}(\Sigma_k)$. To this end, notice that every feasible solution S of (5) satisfies the equality $\|S\|_1 = \|S^{\text{opt}}\|_1$ due to Definition 4. This implies that the additive term $\lambda\|S\|_1$ can be eliminated from the objective function of the optimization problem (5). On the other hand, the first part of the objective function can be expressed as $\text{trace}(\Sigma S) = \sum_{i,j=1}^{n} (\Sigma_{ij} S_{ij})$. By investigating this sum, it is straightforward to show that the optimization problem (5) has a unique solution $\hat{S}^{\text{opt}}$ that can be obtained as follows:

- First, we focus on the upper triangular part of Σ (excluding the diagonal) and identify those k entries with the largest absolute values, which are denoted as $\Sigma_{i_1 j_1}, \Sigma_{i_2 j_2}, ...,$ $\Sigma_{i_k j_k}$ such that
$$|\Sigma_{i_1 j_1}| > |\Sigma_{i_2 j_2}| > ... > |\Sigma_{i_k j_k}| \tag{7}$$
(note that the nonzero entries of the upper triangular part of Σ have different magnitudes, by assumption).

- Second, we repeat the above procedure on the matrix S^{opt} and identify those k entries with the greatest absolute values, which are denoted as $(p_1, q_1), ..., (p_k, q_k)$ such that
$$|S^{\text{opt}}_{p_1 q_1}| \geq |S^{\text{opt}}_{p_2 q_2}| \geq ... \geq |S^{\text{opt}}_{p_k q_k}| \tag{8}$$

- For every $l \in \{1, ..., k\}$, the $(i_l, j_l)^{\text{th}}$ entry of $\hat{S}^{\text{opt}}$ is equal to -sign$(\Sigma_{i_l j_l})|S^{\text{opt}}_{p_l q_l}|$.

Now, it is easy to verify that

$$\mathcal{G}(\hat{S}^{\text{opt}}) = \mathcal{G}(\Sigma_k) \tag{9}$$

On the other hand, since the objective of the optimization problem (2) is strictly convex, $\mathcal{S}^{\text{opt}}$ is a unique solution of this problem. Hence, due to the fact the feasible set of the problem (6) is contained in that of (2), the optimization problem (6) has the unique solution $\mathcal{S}^{\text{opt}}$. Therefore, the two solutions $\hat{S}^{\text{opt}}$ and $\mathcal{S}^{\text{opt}}$ of the problem (6) are identical. Now, the proof follows from the relation (9). ∎

As verified by the author, the condition given in Theorem 5 is satisfied for many numerical examples, leading to the equivalence of thresholding and graphical lasso. The main intuition behind the satisfaction of the above condition is as follows:

- Consider a small number k (or a large number λ) for which the matrix S^{opt} is highly sparse.
- Due to Lemma 3, the diagonal entries of S^{opt} would be relatively much larger than the nonzero off-diagonal entries of S^{opt}.
- Hence, the permutation of the (small) off-diagonal entries of the positive semidefinite S^{opt} would not make the matrix sign indefinite and also has a negligible effect on the $\log\det$ of the matrix.
- Under such circumstances, the condition derived in Theorem 5 would be satisfied (note that (5) is obtained from (6) by dropping the sign-definite condition and the $\log\det$ term).

To strengthen the above argument, an easy-to-check condition will be provided next to guarantee the equivalence of thresholding and graphical lasso.

2.3 Second Condition for Equivalence

Consider the solution S^{opt}. The objective of this part is to derive a condition of equivalence that depends only on the entries of S^{opt}, without using λ or Σ explicitly. Recall that this equivalence does not requite that the matrices S^{opt} and Σ_k be identical (which is unlikely to occur in practice), and is only concerned with the sparsity patterns of these matrices.

Theorem 6 *Let k denote the number of edges of the graph $\mathcal{G}(S^{opt})$. Assume that the inequalities*

$$\operatorname{sign}\left(S^{opt}_{ij}\right) \times \operatorname{sign}\left((S^{opt})^{-1}_{ij}\right) \leq 0 \tag{10a}$$

$$|(S^{opt})^{-1}_{ij}| \geq |(S^{opt})^{-1}_{pq}| \tag{10b}$$

hold for every two pairs (i,j) and (p,q) satisfying

$$(i,j) \in \mathcal{G}(S^{opt}) \tag{11a}$$

$$(p,q) \in \mathcal{G}((S^{opt})^{-1}) - \mathcal{G}(S^{opt}) \tag{11b}$$

Then, graphical lasso and thresholding produce the same graph, i.e., $\mathcal{G}(S^{opt}) = \mathcal{G}(\Sigma_k)$.

Proof Consider two arbitrary pairs (i,j) and (p,q) satisfying (11). It follows from (10) and Lemma 3 that

$$|\Sigma_{ij}| \geq \lambda \tag{12}$$

and that

$$|\Sigma_{ij}| - \lambda \geq |(S^{\text{opt}})^{-1}_{pq}| \tag{13}$$

Similar to the proof of Theorem 5, let the entries of the upper triangular part of Σ_k be ordered as

$$|\Sigma_{i_1 j_1}| > ... > |\Sigma_{i_k j_k}| > |\Sigma_{i_{k+1} j_{k+1}}| \geq \cdots \geq |\Sigma_{i_m j_m}| \tag{14}$$

where $m = \frac{n^2-n}{2}$. To prove the theorem by contradiction, assume that $\mathcal{G}(S^{\text{opt}}) \neq \mathcal{G}(\Sigma_k)$. Recall that $\mathcal{G}(\Sigma_k)$ is a graph with n vertices and the edges $(i_1, j_1), (i_2, j_2), ..., (i_k, j_k)$. Since $\mathcal{G}(S^{\text{opt}})$ has exactly k edges, the above assumption implies that there exist two numbers s and r such that

$$s \in \{1, ..., k\} \qquad \text{and} \qquad (i_s, j_s) \notin \mathcal{G}(S^{\text{opt}}) \tag{15a}$$

$$r \in \{s+1, ..., m\} \qquad \text{and} \qquad (i_r, j_r) \in \mathcal{G}(S^{\text{opt}}) \tag{15b}$$

Therefore, it follows from (13) that

$$|(S^{\text{opt}})^{-1}_{i_s j_s}| \leq |\Sigma_{i_r j_r}| - \lambda < |\Sigma_{i_s j_s}| - \lambda \tag{16}$$

However, the inequality

$$|(S^{\text{opt}})^{-1}_{i_s j_s}| < |\Sigma_{i_s j_s}| - \lambda \tag{17}$$

is in contradiction with the relation

$$(S^{\text{opt}})^{-1}_{i_s j_s} \in [\Sigma_{i_s j_s} - \lambda, \Sigma_{i_2 j_s} + \lambda] \tag{18}$$

that is given in (3). This completes the proof. ■

Theorem 6 states that graphical lasso and thresholding are equivalent if two conditions are satisfied:

- *Condition 1:* The sign of every nonzero off-diagonal entry of S^{opt} is different from that of its corresponding entry in the inverse of S^{opt}.
- *Condition 2:* If a zero off-diagonal entry of S^{opt} takes a nonzero value in the inverse of S^{opt}, then its magnitude is not larger than the magnitude of any off-diagonal element of $(S^{\text{opt}})^{-1}$ corresponding to a nonzero entry of S^{opt}.

Note that the above conditions only depend on S^{opt} and are not directly related to Σ. To better understand these conditions, we decompose S^{opt} as

$$S^{\text{opt}} = D^{\text{opt}} + O^{\text{opt}} \tag{19}$$

where D^{opt} is a diagonal matrix and O^{opt} has a zero diagonal. If the norm of $(D^{\text{opt}})^{-1}O^{\text{opt}}$ is less than 1, one can write

$$(S^{\text{opt}})^{-1} = \left(\sum_{t=0}^{\infty} \left(-(D^{\text{opt}})^{-1} O^{\text{opt}} \right)^t \right) (D^{\text{opt}})^{-1} \tag{20}$$

In general, we have

$$(S^{\text{opt}})^{-1} = (D^{\text{opt}})^{-1} - (D^{\text{opt}})^{-1} O^{\text{opt}} (D^{\text{opt}})^{-1} + h.o.t. \tag{21}$$

where *h.o.t* stands for *higher order terms* in the Taylor series expansion if the norm of $(D^{\text{opt}})^{-1}O^{\text{opt}}$ is less than 1, and otherwise is equal to $(S^{\text{opt}})^{-1}O^{\text{opt}}(D^{\text{opt}})^{-1}O^{\text{opt}}(D^{\text{opt}})^{-1}$ (as a general formula). We refer to

$$E^{\text{opt}} := (D^{\text{opt}})^{-1} - (D^{\text{opt}})^{-1}O^{\text{opt}}(D^{\text{opt}})^{-1} \tag{22}$$

as the first-order approximation of $(S^{\text{opt}})^{-1}$. Note that if λ is relatively large, it is expected that O^{opt} will be small compared to D^{opt}, which will lead to small higher order terms.

Theorem 7 *The condition* (10) *given in Theorem 6 to guarantee the equivalence of graphical lasso and thresholding is satisfied if* $(S^{opt})^{-1}$ *is replaced by its first-order approximation* E^{opt} *in the condition.*

Proof Equation (22) yields that

$$E^{\text{opt}}_{ij} = \begin{cases} (D^{\text{opt}})^{-1}_{ii} & \text{if} \quad i=j \\ -(D^{\text{opt}})^{-1}_{ii}O^{\text{opt}}_{ij}(D^{\text{opt}})^{-1}_{jj} & \text{if} \quad (i,j)\in\mathcal{G}(S^{\text{opt}}) \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

for every $i,j\in\{1,...,n\}$. Note that $D^{\text{opt}}>0$ due to the positive definiteness of S^{opt}. Hence, given a pair $(i,j)\in\mathcal{G}(S^{\text{opt}})$, one can write:

$$\text{sign}\left(S^{\text{opt}}_{ij}\right)\text{sign}\left(E^{\text{opt}}_{ij}\right) = -\text{sign}\left(O^{\text{opt}}_{ij}\right)^2(D^{\text{opt}})^{-1}_{ii}(D^{\text{opt}})^{-1}_{jj}\leq 0 \tag{24}$$

Moreover, given arbitrary pairs (i,j) and (p,q) satisfying (11), it follows from (23) that $E^{\text{opt}}_{pq}=0$. Therefore,

$$|E^{\text{opt}}_{ij}|\geq|E^{\text{opt}}_{pq}| \tag{25}$$

The proof is completed by combining (24) and (25). ■

Before further simplifying the conditions of Theorem 6 based on Theorem 7, it is desirable to offer a more general condition measuring the "similarity" of graphical lasso and thresholding (as opposed to their "equivalence").

Definition 8 *Define* $\mathbb{I}_k$ *as the set of indices (locations) of those* k *entries of the upper triangular part of* $(\mathcal{S}^{opt})^{-1}$ *that have the largest magnitudes.*

Note that if multiple entries of $(\mathcal{S}^{\text{opt}})^{-1}$ have the same value, then $\mathbb{I}_k$ may not be uniquely defined. In that case, $\mathbb{I}_k$ can be considered as any of the sets satisfying the properties given in Definition 8.

Theorem 9 *Let* k *denote the number of edges of the graph* $\mathcal{G}(S^{opt})$*, and* h *be the number of indices* (i,j)*'s in the set* $\mathbb{I}_k$ *for which the relation*

$$\text{sign}\left(S^{opt}_{ij}\right)\times\text{sign}\left((S^{opt})^{-1}_{ij}\right)<0 \tag{26}$$

holds. Then, the graphs $\mathcal{G}(\Sigma_k)$ *and* $\mathcal{G}(\mathcal{S}^{opt})$ *have at least* h *edges in common. In particular, graphical lasso and thresholding are equivalent if* $h=k$*.*

Proof The proof of Theorem 6 can be adopted to prove this theorem. The details are omitted for brevity. ■

Theorem 9 states that the graph $\mathcal{G}(\mathcal{S}^{\text{opt}})$ (obtained using graphical lasso) and the graph $\mathcal{G}(\Sigma_k)$ (obtained using thresholding) share at least h edges. Hence, this theorem provides a simple mechanism to check the similarity between graphical lasso and thresholding through a simple test on $\mathcal{S}^{\text{opt}}$. This mechanism will be further studied below.

Definition 10 *Given a matrix $H \in \mathbb{S}^n$ and a positive integer $t \in \{1, 2, ..., k\}$, define $f(H, t)$ as the magnitude of the t^{th} largest entry (in magnitude) of the upper triangular part of H (excluding its diagonal). For example, $f(H, 1)$ is equal to the absolute value of an off-diagonal entry of H with the largest magnitude.*

We define ΔE^{opt} as the difference between the matrix $(S^{\text{opt}})^{-1}$ and its first-order approximation E^{opt}.

Theorem 11 *Given a positive integer $t \in \{1, 2, ..., k\}$, the graphs $G(\Sigma_k)$ and $G(S^{opt})$ have at least t edges in common if*

$$2\ f(\Delta E^{opt}, 1) < f(E^{opt}, t) \tag{27}$$

In particular, graphical lasso and thresholding lead to the same approximate graph if the above inequality is satisfied for $t = k$.

Proof The proof follows from Theorems 7 and 9. The main idea will be sketched for $t = k$ below. Consider arbitrary pairs (i, j) and (p, q) satisfying (11). It follows from (27) that

$$|E_{ij}^{\text{opt}}| \geq f(E^{\text{opt}}, k) \geq 2f(\Delta E^{\text{opt}}, 1) \geq\ 2|\Delta E_{ij}^{\text{opt}}| \tag{28}$$

Similarly,

$$|E_{ij}^{\text{opt}}| \geq f(E^{\text{opt}}, k) \geq 2f(\Delta E^{\text{opt}}, 1) \geq\ 2|\Delta E_{pq}^{\text{opt}}| \tag{29}$$

On the other hand, $(S^{\text{opt}})_{ij}^{-1} = E_{ij}^{\text{opt}} + \Delta E_{ij}^{\text{opt}}$ and $\text{sign}(S_{ij}^{\text{opt}}) = \text{sign}(O_{ij}^{\text{opt}}) = -\text{sign}(E_{ij}^{\text{opt}})$. The above relations together with (23) imply the condition (10). This completes the proof. ■

Roughly speaking, $f(\Delta E^{\text{opt}}, 1)$ is small for a relatively large number λ, and $f(E^{\text{opt}}, t)$ would stay away from zero due to Lemma 3. Theorem 11 explains that the relationship between E^{opt} and ΔE^{opt} determines the degree of similarity between graphical lasso and thresholding. Note that Theorem 11 is more conservative than Theorem 9, but its condition is more insightful.

Note that the definition of graphical lasso in (2) is based on the regularization term $\lambda\|S\|_1$. Consider a second version of graphical lasso where only the off-diagonal entries of S are penalized in the regularization term. This is realized by replacing the term $\|S\|_1$ with $2\sum_{i=1}^{n}\sum_{j=i+1}^{n}|S_{ij}|$. The optimality conditions given in Lemma 3 for graphical lasso are valid for the second version of graphical lasso after dropping λ from the right side of (3a). As a result, the original and second version of graphical lasso are very similar, and the only

$$\Sigma = \begin{bmatrix} 1 & 0.5448 & 0.4980 & 0.2045 & -0.2818 & -0.1452 \\ 0.5448 & 1 & -0.1327 & -0.0604 & -0.6860 & -0.0457 \\ 0.4980 & -0.1327 & 1 & 0.1283 & -0.1859 & -0.5174 \\ 0.2045 & -0.0604 & 0.1283 & 1 & 0.4019 & 0.6238 \\ -0.2818 & -0.6860 & -0.1859 & 0.4019 & 1 & 0.5139 \\ -0.1452 & -0.0457 & -0.5174 & 0.6238 & 0.5139 & 1 \end{bmatrix} \tag{30}$$

$$\mathcal{S}^{\text{opt}} = \begin{bmatrix} 0.6934 & -0.0453 & -0.0229 & 0 & 0 & 0 \\ -0.0453 & 0.7114 & 0 & 0 & 0.1153 & 0 \\ -0.0229 & 0 & 0.6919 & 0 & 0 & 0.0321 \\ 0 & 0 & 0 & 0.6997 & 0 & -0.0839 \\ 0 & 0.1153 & 0 & 0 & 0.7098 & -0.0305 \\ 0 & 0 & 0.0321 & -0.0839 & -0.0305 & 0.7025 \end{bmatrix} \tag{31}$$

$$(\mathcal{S}^{\text{opt}})^{-1} = \begin{bmatrix} 1.4500 & 0.0949 & 0.0480 & -0.0003 & -0.0155 & -0.0029 \\ 0.0949 & 1.4500 & 0.0036 & -0.0013 & -0.2360 & -0.0106 \\ 0.0480 & 0.0036 & 1.4500 & -0.0081 & -0.0035 & -0.0674 \\ -0.0003 & -0.0013 & -0.0081 & 1.4500 & 0.0077 & 0.1738 \\ -0.0155 & -0.2360 & -0.0035 & 0.0077 & 1.4500 & 0.0640 \\ -0.0029 & -0.0106 & -0.0674 & 0.1738 & 0.0640 & 1.4500 \end{bmatrix} \tag{32}$$

difference is in the diagonal of $(S^{\text{opt}})^{-1}$. It is easy to verify that the theorems developed in this work are all valid for the second version of graphical lasso as well. However, note that in order to obtain a graph with k edges, the value of the regularization term λ may not be the same for the original and the second version of graphical lasso.

3. Numerical Examples

Example 1: Consider Σ as the randomly generated matrix given in (30). The solution $\mathcal{S}^{\text{opt}}$ of graphical lasso with $\lambda = 0.45$ is provided in (31) (this value of λ guarantees that $\mathcal{G}(\mathcal{S}^{\text{opt}})$ will have n edges). It can be deduced from this solution that the graph $\mathcal{G}(\mathcal{S}^{\text{opt}})$ consists of the vertex set $\mathcal{V} := \{1, 2, ..., 6\}$ and the edge set $\mathcal{E} := \{(1, 2), (1, 3), (2, 5), (3, 6), (4, 6), (5, 6)\}$. On the other hand, it follows from a simple inspection of the matrix Σ that $\mathcal{E}$ coincides with the index set of the 6 largest absolute values of the upper triangular part of Σ. Hence, graphical lasso and thresholding are equivalent in this example, meaning that $\mathcal{G}(\mathcal{S}^{\text{opt}}) = \mathcal{G}(\Sigma_6)$.

It is desirable to find out whether the simple conditions proposed in Theorem 9 can be used to detect this equivalence. Recall that these conditions are expressed in terms of the matrix $\mathcal{S}^{\text{opt}}$, without using λ and Σ explicitly. To check these conditions, the matrix $(\mathcal{S}^{\text{opt}})^{-1}$ can be obtained as the one given in (32). Since the upper triangular part of $\mathcal{S}^{\text{opt}}$ has 6 nonzero entries, the index set $\mathbb{I}_6$ needs to be found. After the identification of the

$$\Sigma = \begin{bmatrix} 1 & 0.9342 & 0.8156 & 0.8609 & 0.6994 & 0.8457 \\ 0.9342 & 1 & 0.7110 & 0.7736 & 0.7283 & 0.8532 \\ 0.8156 & 0.7110 & 1 & 0.8593 & 0.8905 & 0.7958 \\ 0.8609 & 0.7736 & 0.8593 & 1 & 0.7876 & 0.7793 \\ 0.6994 & 0.7283 & 0.8905 & 0.7876 & 1 & 0.7040 \\ 0.8457 & 0.8532 & 0.7958 & 0.7793 & 0.7040 & 1 \end{bmatrix} \tag{35}$$

$$\Sigma_5 = \begin{bmatrix} 1 & 0.9342 & 0 & 0.8609 & 0 & 0 \\ 0.9342 & 1 & 0 & 0 & 0 & 0.8532 \\ 0 & 0 & 1 & 0.8593 & 0.8905 & 0 \\ 0.8609 & 0 & 0.8593 & 1 & 0 & 0 \\ 0 & 0 & 0.8905 & 0 & 1 & 0 \\ 0 & 0.8532 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{36}$$

$$\mathcal{S}^{\text{opt}} = \begin{bmatrix} 0.5417 & -0.0247 & 0 & -0.0032 & 0 & 0 \\ -0.0247 & 0.5417 & 0 & 0 & 0 & -0.0009 \\ 0 & 0 & 0.5408 & -0.0027 & -0.0118 & 0 \\ -0.0032 & 0 & -0.0027 & 0.5406 & 0 & 0 \\ 0 & 0 & -0.0118 & 0 & 0.5408 & 0 \\ 0 & -0.0009 & 0 & 0 & 0 & 0.5405 \end{bmatrix} \tag{37}$$

$$(\mathcal{S}^{\text{opt}})^{-1} = \begin{bmatrix} 1.8500 & 0.0842 & 0.0001 & 0.0109 & 0.0000 & 0.0002 \\ 0.0842 & 1.8500 & 0.0000 & 0.0005 & 0.0000 & 0.0032 \\ 0.0001 & 0.0000 & 1.8500 & 0.0093 & 0.0405 & 0.0000 \\ 0.0109 & 0.0005 & 0.0093 & 1.8500 & 0.0002 & 0.0000 \\ 0.0000 & 0.0000 & 0.0405 & 0.0002 & 1.8500 & 0.0000 \\ 0.0002 & 0.0032 & 0.0000 & 0.0000 & 0.0000 & 1.8500 \end{bmatrix} \tag{38}$$

largest absolute values of the matrix $(\mathcal{S}^{\text{opt}})^{-1}$, it turns out that

$$\mathbb{I}_6 = \{(1,2),(1,3),(2,5),(3,6),(4,6),(5,6)\} \tag{33}$$

In addition, the relation

$$\text{sign}\left(S_{ij}^{\text{opt}}\right) \times \text{sign}\left((S^{\text{opt}})_{ij}^{-1}\right) < 0, \quad \forall (i,j) \in \mathbb{I}_6 \tag{34}$$

holds. Therefore, it follows from Theorem 9 that graphical lasso and thresholding lead to the same result.

Example 2: Consider the randomly generated matrix in (35). Thresholding this matrix at the level of $k = 5$ yields the solution given in (36). On the other hand, solving the graphical lasso for $\lambda = 0.85$ leads to the solution $\mathcal{S}^{\text{opt}}$ provided in (37), with the inverse given in (38) (this value of λ guarantees that $\mathcal{G}(\mathcal{S}^{\text{opt}})$ will have $n-1$ edges). By analyzing the matrix $\mathcal{S}^{\text{opt}}$ and its inverse, it can be verified that the conditions provided in Theorem 9 are satisfied. Hence, thresholding and graphical lasso are equivalent. An interesting observation is that the matrix Σ has two close entries 0.8457 and 0.8532 such that only one of them is removed through thresholding, but graphical lasso recognizes this fact and selects the entry with a slightly higher value. In other words, even in the case where the matrix Σ have entries with similar values (which makes it hard to decide what entries should be included in the graphical model), graphical lasso may still behave the same as thresholding.

Example 3: We construct a matrix NN^T, where the entries of $N \in \mathbb{R}^{30\times 30}$ are chosen at random according to some probably distribution. Define Σ as a matrix obtained from NN^T through a normalization to make its diagonal entries all equal to 1. We order the entries of Σ according to (14) and consider λ as $\frac{|\Sigma_{i_{30}j_{30}}|+|\Sigma_{i_{29}j_{29}}|}{2}$ (i.e., the average of the 29^{th} and 30^{th} largest absolute values of the upper triangular part of Σ). Two experiments will be concluded below.

Experiment I: By choosing every entry of the matrix N from a normal probability distribution, we generated 100 random matrices Σ's. In Figure 1(a), the number of edges of $\mathcal{G}(\mathcal{S}^{\text{opt}})$ is shown for each of the 100 trials (the trials are reordered to make the curve increasing). In Figure 1(b), the number of edges of the difference graph $\mathcal{G}(\mathcal{S}^{\text{opt}}) - \mathcal{G}(\Sigma_k)$ is depicted, where k is considered as the number of edges of $\mathcal{G}(\mathcal{S}^{\text{opt}})$. It can be seen that graphical lasso and thresholding are equivalent in more than 75 trials and are different by at most 2 edges in the remaining trials (note that the orderings of the trials for Figures 1(a) and 1(b) are different to ensure that each curve changes smoothly). This supports the claim of the paper that graphical lasso and thresholding would behave highly similarly. As opposed to computing the graph $\mathcal{G}(\mathcal{S}^{\text{opt}}) - \mathcal{G}(\Sigma_k)$ directly, Theorem 9 offers a simple insightful condition to find a subset of the common edges of $\mathcal{G}(\mathcal{S}^{\text{opt}})$ and $\mathcal{G}(\Sigma_k)$. This simple condition is tested on the 100 trials and the results are summarized in Figure 1(c). This figure shows the percentage of the common edges of $\mathcal{G}(\mathcal{S}^{\text{opt}})$ and $\mathcal{G}(\Sigma_k)$ that are detected by Theorem 9. It can be observed that the condition provided in the paper is able to detect a similarity degree on the order of 80% for more than 90% of the trials.

Experiment II: This study is the same as the previous experiment with the only difference that every entry of the matrix N was chosen from the interval $[0, 1]$ with a uniform probability distribution. The results are shown in Figures 1(d), (e) and (f). It can be seen that graphical lasso and thresholding are similar with the probability of at least 95%.

4. Case Study on Brain Networks

Brain functional connectivity is defined as the statistical dependencies between the activities of disjoint brain regions. A brain functional network can be represented by a set of nodes and edges, where each node corresponds to a brain region and each edge shows the presence of correlated activities (namely, a nonzero partial correlation) between two disjoint regions. We apply graphical lasso and thresholding techniques to the fMRI data collected from twenty subjects to obtain their associated brain functional networks. These fMRI data

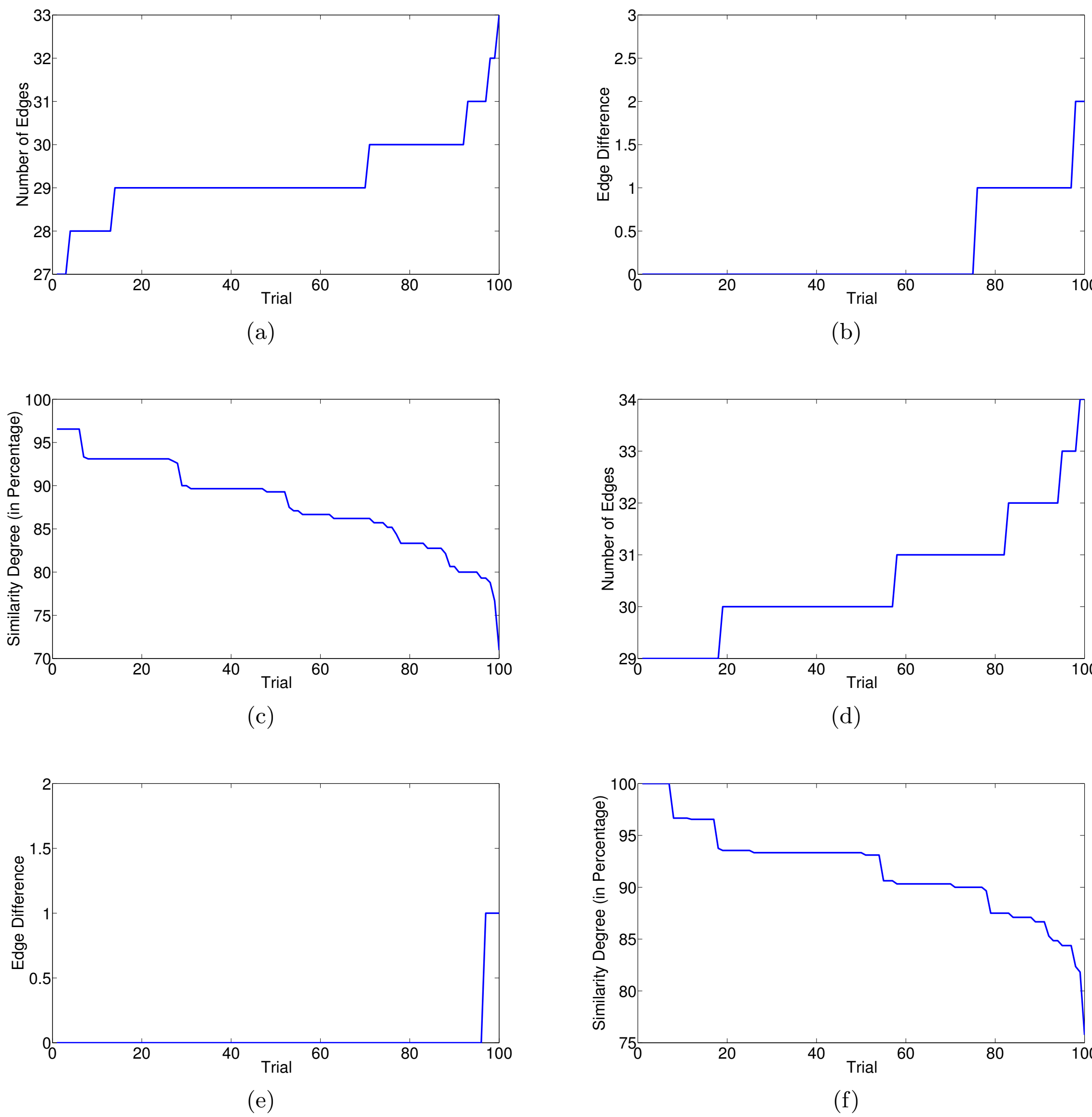


Figure 1: Figures (a), (b) and (c) show the number of edges of $\mathcal{G}(\mathcal{S}^{\text{opt}})$, the number of edges of $\mathcal{G}(\mathcal{S}^{\text{opt}}) - \mathcal{G}(\Sigma_k)$, and the similarity degree of thresholding and graphical lasso detected via Theorem 9 for Experiment I. Figures (d), (e) and (f) show the number of edges of $\mathcal{G}(\mathcal{S}^{\text{opt}})$, the number of edges of $\mathcal{G}(\mathcal{S}^{\text{opt}}) - \mathcal{G}(\Sigma_k)$, and the similarity degree of thresholding and graphical lasso detected via Theorem 9 for Experiment II.

sets are borrowed form Vértes et al. (2012). Each data set includes 134 samples of the low frequency oscillations, taken at 140 cortical brain regions in the right hemisphere. The goal is to find the brain functional connectivity network of each subject that specifies the direct interactions (or conditional dependence/independence) between the activities of these cortical regions.

Using the aforementioned time series data, a 140×140 sample correlation matrix can be computed for each subject. Note that the number of samples is smaller than the number of variables and, therefore, the sample correlation matrix is not invertible. As a result, it is not possible to take the inverse of the sample correlation matrix for estimating a partial correlation matrix. The brain network being sought has $n = 140$ nodes (brain regions). In an effort to find a subgraph of this network as closely as possible to a spanning tree, we choose the regularization parameter λ in the graphical lasso algorithm and the level of thresholding in such a way that they both lead to graphs with $n - 1 = 139$ edges. As illustrated in Figure 2, the similarity degree between the outcomes of these two techniques is above 90% for all twenty subjects. The outcomes of graphical lasso and thresholding obtained for 4 of these subjects are given in Figure 3 for illustration. Note that these graphs are not spanning trees as indented, which imply that graphical lasso and thresholding are able to obtain graphs with $n - 1$ edges but they inevitably have cycles.

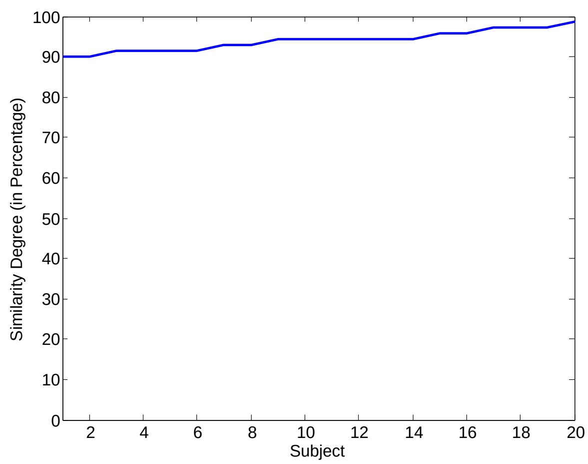


Figure 2: The similarity degree of thresholding and graphical lasso for obtaining brain networks of 20 subjects.

5. Case Study on Electrical Circuits

Consider an arbitrary resistor-capacitor (RC) circuit with n nodes, where certain nodes are connected to each other or the ground via resistor-capacitor elements. Figure 4(a) illustrates an RC circuit. The connectivity of each circuit can be represented by a graph, as demonstrated in Figure 4(b). Assume that the physical structure of the circuit is unknown and only the nodal voltages are available for measurement. It is desirable to find the structural connectivity of the circuit from the measured signals. Given a time instance t, let $V(t)$ denote the vector of the voltages for nodes $1, ..., n$ at time t. Assume that the circuit elements are subject to white thermal noise, namely Johnson-Nyquist noise, and that the conductance and capacitance in each resistor-capacitor element are identical. Let

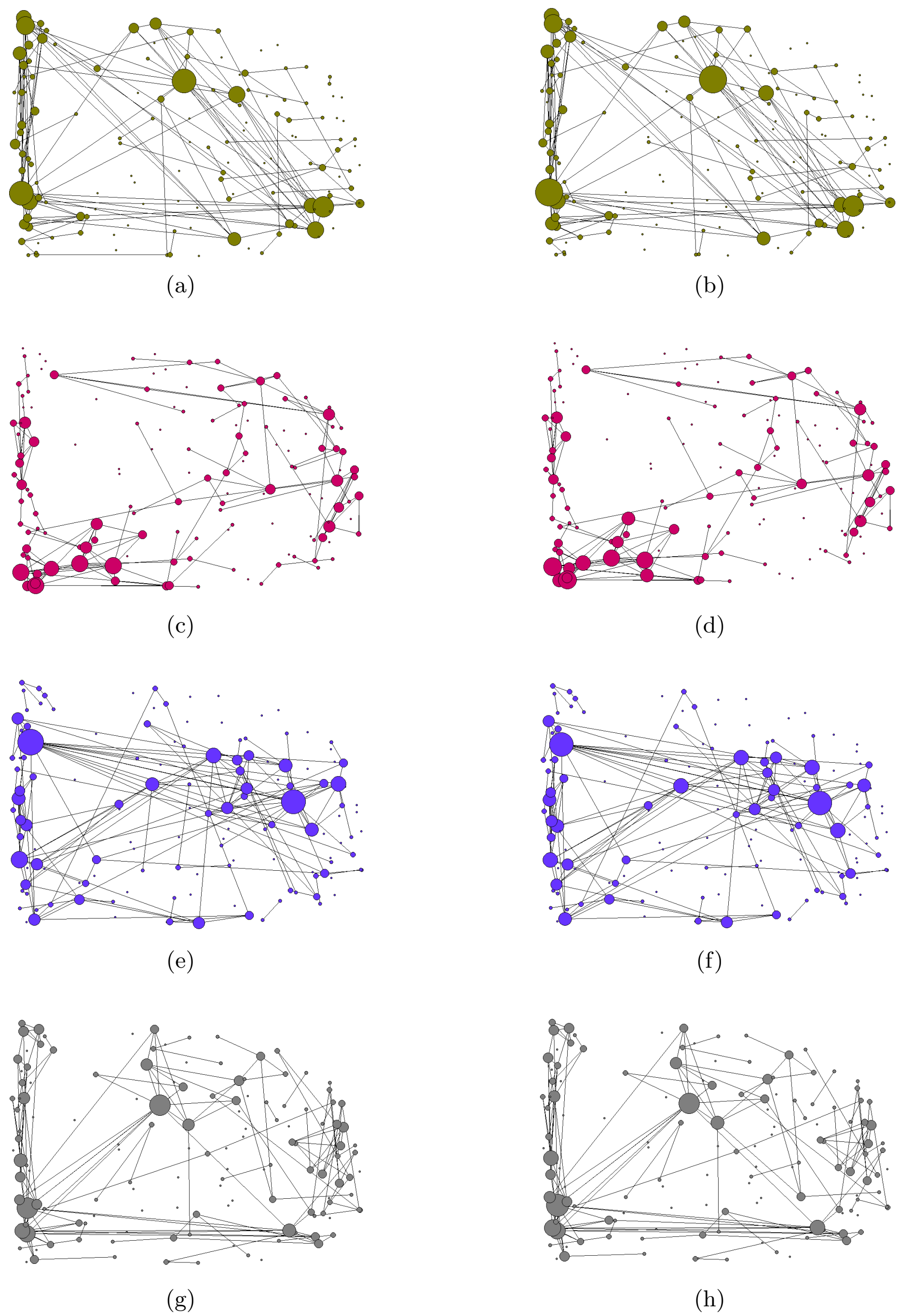
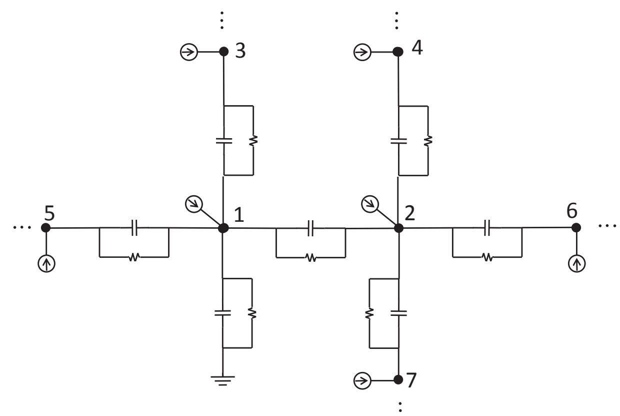


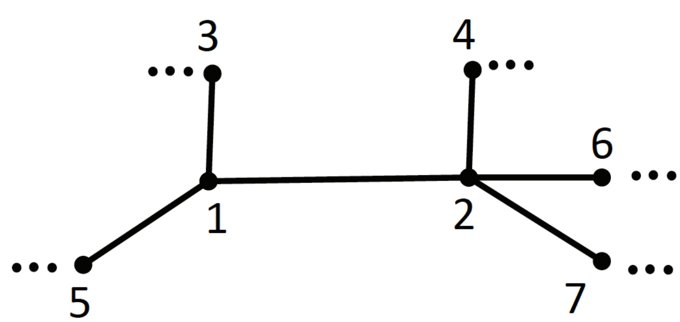
Figure 3: Figures (a) and (b) show the brain networks of one subject obtained from graphical lasso and thresholding, respectively. Similarly, figures (c)-(d), (e)-(f) and (g)-(h) show the networks obtained from graphical lasso and thresholding for three other subjects. The size of each node in these networks reflects its degree.

Σ_* denote the "steady-state" covariance of the voltage measurements. It can be shown that $\Sigma_* = C^{-1}$, where C is the $n \times n$ capacitance matrix of the circuit (Sojoudi and Doyle, 2014). Note that the sparsity pattern of C is consistent with the topology of the circuit. Therefore, the inverse covariance matrix Σ_*^{-1} and the partial correlation matrix both have the same sparsity structure as the circuit. In other words, the partial correlation matrix unveils the physical connectivity of the circuit. Assuming that the circuit under study has a sparse structure, it can be concluded that

- Σ_* is generically a dense matrix, due to being the inverse of the sparse matrix C.
- Σ_*^{-1} is sparse and its sparsity pattern conforms with the circuit topology.



(a)



(b)

Figure 4: a) An RC network with node 1 connected to the ground via a resistor and a capacitor, b) a graph representation of the RC network connectivity.

The above physical model illustrates the fact that the topology of a system may have been encoded in the partial correlation matrix. To recover the circuit topology from the voltage vector $V(t)$, one can sample the vector $V(t)$ at different times $t_1, t_2, ..., t_r$ and construct a sample covariance matrix as

$$\Sigma = \frac{1}{r}\sum_{i=1}^{r} V_i(t_i)V_i(t_i)^T \tag{39}$$

where r denotes the number of samples. Note that Σ converges to the population covariance Σ_* as $r \to \infty$. When r is finite, two possible scenarios arise: i) Σ is invertible but the inverse matrix needs to be thresholded to some level due to the error $\Sigma_* - \Sigma$, ii) Σ is not invertible and therefore alternative methods are needed to estimate the inverse matrix.

The above circuit model can be used to study the relationship between the thresholding and graphical lasso techniques. As an example, consider a mesh circuit with $n = 100$ nodes that are connected to one another through 180 links. The graphical model of this grid circuit is depicted in Figure 5(a). With no loss of generality, assume that $C_{ij} = -1$ for all $(i, j) \in \mathcal{E}$. Furthermore, suppose that nodes 2, 9, 22, 61 and 70 of the circuit are connected to the ground through parallel RC circuits with values equal to 0.1. For $r = 99$ and 10 different trials, we have calculated the sample covariance matrices and applied the thresholding technique and graphical lasso to these matrices in order to recover networks with 180 edges. Note that since the number of samples is less than the number of variables, the sample partial correlations cannot be obtained through matrix inversion. The degree of similarity between graphical lasso and thresholding for these 10 trials are given in Figure 5(b). Figures 5(c) and 5(d) show the networks obtained from thresholding and graphical lasso for one of the trials. False positives are marked in red and false negatives are colored in blue. These two graphs have 178 edges in common (out of 180 edges), which indicates a high degree of similarity between the outcomes of the two techniques under study (to observe some of the few differences, one can inspect the existence of the edges $(64, 75)$ and $(55, 56)$ in these two graphs). We have repeated the above experiment on many circuit models beyond mesh networks and observed a very similar result.

6. Conclusions

The objective of this paper is to study the problem of finding a sparse conditional dependence graph associated with a multivariate random vector, where the only available information is a sample correlation matrix. A commonly used technique for this problem is a convex program, named graphical lasso, where the objective function of this optimization problem has a sparsity-promoting penalty term. In this work, a simple condition is derived under which the graph obtained from graphical lasso coincides with the one obtained by simply thresholding the sample correlation matrix. This condition depends only on the solution of graphical lasso, and makes no use of the regularization coefficient or the sample correlation matrix. The focus of the paper is on the case where the regularization term is high enough to search for a sparse graph. A theoretical result is developed to support that graphical lasso and thresholding behave similarly in this regime, and this statement is verified in several random simulations as well as two case studies on brain connectivity networks and electrical circuits.

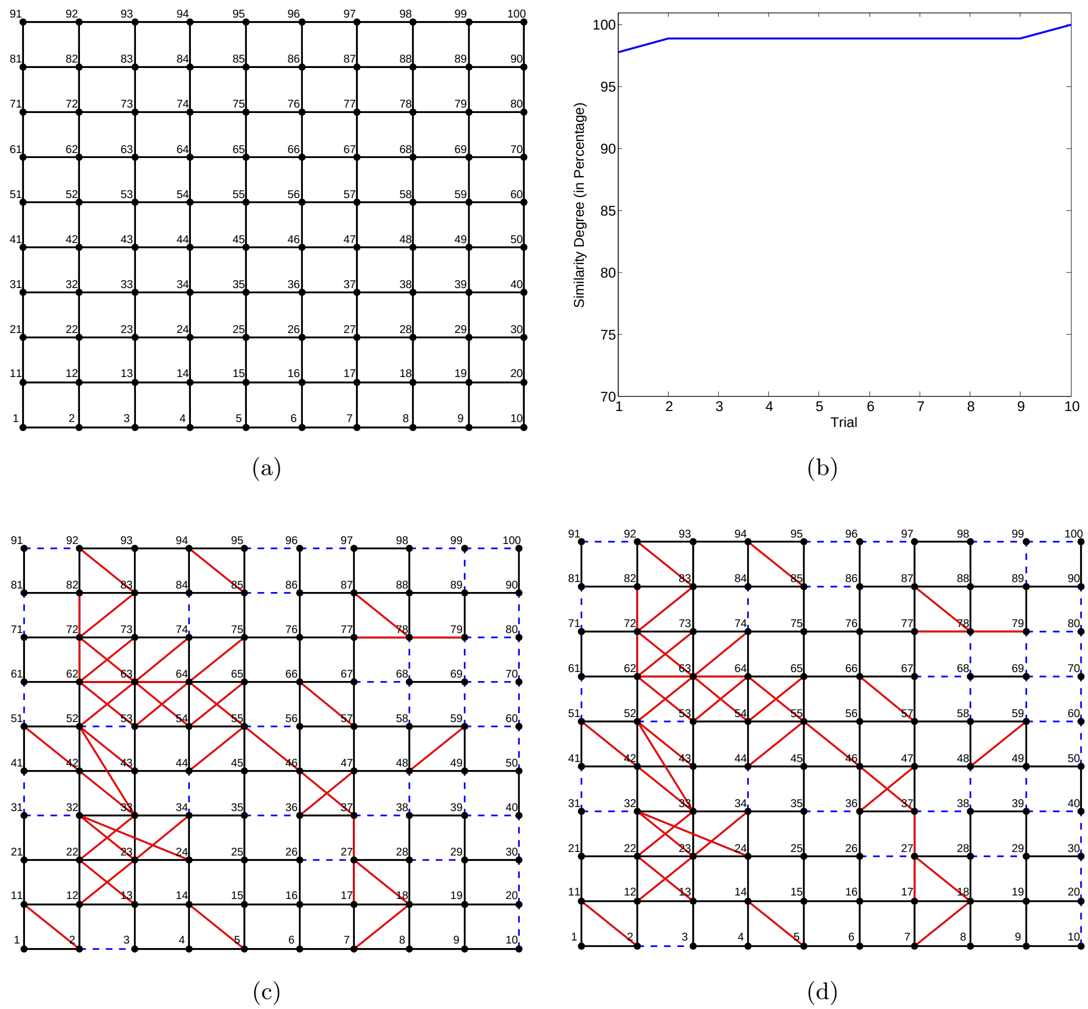


Figure 5: a) The mesh RC circuit studied in Section 5, b) the similarity degree of thresholding and graphical lasso for the mesh network over 10 trails, c) the network obtained from thresholding the sample correlation matrix for one trial, d) the network obtained from graphical lasso for the same trial used for Figure (c).

References

Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516, 2008.

Alfred M Bruckstein, David L Donoho, and Michael Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.

Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

Venkat Chandrasekaran, Pablo Parrilo, Alan S Willsky, et al. Latent variable graphical model selection via convex optimization. In *48th Annual Allerton Conference on Communication, Control, and Computing*, pages 1610–1613, 2010.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373–397, 2014.

Jianqing Fan and Jinchi Lv. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101, 2010.

Salar Fattahi and Javad Lavaei. On the convexity of optimal decentralized control problem and sparsity path. *http://www.ieor.berkeley.edu/~lavaei/SODC_2016.pdf*, 2016.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Tom Goldstein and Stanley Osher. The Split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

Dominique Guillot and Bala Rajaratnam. Retaining positive definiteness in thresholded matrices. *http://arxiv.org/abs/1108.3325*, 2011.

Shuai Huang, Jing Li, Liang Sun, Jieping Ye, Adam Fleisher, Teresa Wu, Kewei Chen, Eric Reiman, Alzheimer's Disease NeuroImaging Initiative, et al. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010.

Ali Jalali, Pradeep D Ravikumar, Vishvas Vasuki, and Sujay Sanghavi. On learning discrete graphical models using group-sparse regularization. In *International Conference on Artificial Intelligence and Statistics*, pages 378–387, 2011.

Nicole Krämer, Juliane Schäfer, and Anne-Laure Boulesteix. Regularized estimation of large-scale gene association networks using graphical Gaussian models. *BMC bioinformatics*, 10(1):384, 2009.

Han Liu, Kathryn Roeder, and Larry Wasserman. Stability approach to regularization selection (stars) for high dimensional graphical models. In *Advances in Neural Information Processing Systems*, pages 1432–1440, 2010.

Rahul Mazumder and Trevor Hastie. Exact covariance thresholding into connected components for large-scale graphical lasso. *The Journal of Machine Learning Research*, 13(1): 781–794, 2012.

Nicolai Meinshausen and Bin Yu. Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics*, pages 246–270, 2009.

Peter Richtárik and Martin Takáč. Parallel coordinate descent methods for big data optimization. *Mathematical Programming*, pages 1–52, 2012.

Mark Schmidt, Alexandru Niculescu-Mizil, Kevin Murphy, et al. Learning graphical model structure using L1-regularization paths. In *AAAI*, volume 7, pages 1278–1283, 2007.

Somayeh Sojoudi. Graphical lasso and thresholding: Conditions for equivalence. *http://www.somayehsojoudi.com/GL_Versus_TH2.pdf*, 2016.

Somayeh Sojoudi and John Doyle. Study of the brain functional network using synthetic data. In *52nd Annual Allerton Conference on Communication, Control, and Computing*, pages 350–357, 2014.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

Petra E Vértes, Aaron F Alexander-Bloch, Nitin Gogtay, Jay N Giedd, Judith L Rapoport, and Edward T Bullmore. Simple models of human brain functional networks. *Proceedings of the National Academy of Sciences*, 109(15):5868–5873, 2012.

Daniela M Witten, Jerome H Friedman, and Noah Simon. New insights and faster computations for the graphical lasso. *Journal of Computational and Graphical Statistics*, 20(4): 892–900, 2011.

John Wright, Allen Yang, Arvind Ganesh, Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

Ming Yuan and Yi Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

Cun-Hui Zhang and Jian Huang. The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics*, pages 1567–1594, 2008.