BIOMETRICS 64, 1–26 DOI: 10.1111/j.1541-0420.2005.00454.x
December 2008

Common Reducing Subspace Model and Network Alternation Analysis

Wenjing Wang*, Xin Zhang**

Department of Statistics, Florida State University, Tallahassee, Florida, 32306 U.S.A.

**email:* wenjing.wang@stat.fsu.edu

***email:* henry@stat.fsu.edu

and

Lexin Li

Division of Biostatistics, University of California, Berkeley, California, 94720 U.S.A.

email: lexinli@berkeley.edu

SUMMARY: Motivated by brain connectivity analysis and many other network data applications, we study the problem of estimating covariance and precision matrices, and their differences, across multiple populations. We propose a common reducing subspace model that leads to substantial dimension reduction and efficient parameter estimation. We explicitly quantify the efficiency gain through an asymptotic analysis. Our method is built upon and further extends a nascent technique, the envelope model, which adopts a generalized sparsity principle. This distinguishes our proposal from most of existing covariance and precision estimation methods that assume element-wise sparsity. Moreover, unlike most existing solutions, our method can naturally handle both covariance and precision matrices in a unified way, and work with matrix-valued data. We demonstrate the efficacy of our method through intensive simulations, and illustrate the method with an autism spectrum disorder data analysis.

KEY WORDS: Central subspace; Dimension reduction; Envelope models; Network analysis; Neuroimaging analysis; Reducing subspace.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

In this article, we consider the problem of estimating covariance and precision matrices, and in particular their differences, across multiple populations. Our motivation is brain connectivity analysis, whereas our proposal is potentially applicable to a wide range of network data analyses in e-commerce, genomics and social sciences. Accumulated evidence has suggested that, compared to healthy brains, the connectivity network changes in brains with neurological disorders, and such alternations in brain network hold crucial insights of disease pathologies (Fox and Greicius, 2010). Brain connectivity analysis is now in the foreground of neuroscience research (Bullmore and Sporns, 2009), and is also drawing increasing attention in the statistics field (Kim et al., 2014; Ahn et al., 2015; Chen et al., 2015a,b; Qiu et al., 2016; Wang et al., 2016, among others). A brain network is often encoded as a covariance or precision matrix, or their standardized counterpart, a correlation or partial correlation matrix. Nodes of the network represent brain regions, and links represent interactions among those regions (Fornito et al., 2013).

We propose a common reducing subspace model to efficiently estimate the differences of multiple covariance and precision matrices under different conditions, e.g., disease status. Our key idea is to assume that there is a common, lower-dimensional subspace that sufficiently captures all the heterogeneity of the individual matrices across multiple populations. Our method then seeks to estimate this subspace, from which we construct the efficient estimates of covariances, precisions and their differences. Our proposal enjoys several advantages. First, it effectively reduces the number of unknown parameters, and the resulting estimators are statistically efficient. The efficiency gain is explicitly quantified by our asymptotic analysis, and it translates to a competitive empirical performance under a limited sample size. Second, our method can naturally handle both covariance and precision matrices in a unified fashion. By contrast, most existing network estimation methods tackle only covariance, or precision, but rarely both. Third, our model focuses on estimating the differences of covariance and precision matrices across different conditions, since such alternations

are of key scientific interest (Zhao et al., 2014). Meanwhile our method can estimate the covariance and precision under each individual condition as a by-product of our model. Finally, unlike a pure dimension reduction solution, our method simultaneously achieves dimension reduction and obtains actual estimates of the matrices of interest.

There have been a large number of proposals of covariance or precision estimation for a single population, e.g., Yuan and Lin (2007); Friedman et al. (2008); Peng et al. (2009); Ravikumar et al. (2011); Cai and Liu (2011); Cai et al. (2011), or for multiple populations, e.g., Guo et al. (2011); Chiquet et al. (2011); Danaher et al. (2014); Zhu et al. (2014); Lee and Liu (2015); Cai et al. (2016). Compared to those solutions, our method notably differs in several ways. First, nearly all those estimation methods are built upon the sparsity principle, in that many individual entries of the covariance or precision matrix are exactly zero or very small. Our method is built on a principle that shares a similar spirit as, but is also clearly different from the sparsity principle. It assumes that some aspects of covariance or precision matrices share a common structure that is invariant as the underlying group status varies. However, the common structure does not have to take the form of individual entries of the matrices. In a sense, it can be viewed as a generalized version of the sparsity principle (Li and Zhang, 2017). As such, our solution offers a useful complement to the existing literature on covariance and precision estimation. Second, in applications such as gene network or protein network, each individual's data observation is a vector of genetic or proteomic measurements, and a covariance or precision matrix is estimated across multiple subjects. By contrast, in brain connectivity analysis, each individual's imaging data often takes the form of a location by time matrix, and a covariance or precision matrix is calculated for each individual subject. Covariance or precision estimation for matrix-valued data is only recently emerging, and is usually based on the sparsity principle as well (Yin and Li, 2012; Leng and Tang, 2012). We also numerically compare with two representative joint sparse precision estimation methods in Section 5. Finally, we comment that we address network estimation in this article, rather than

network inference. Estimation and inference are two different problems; see Xia and Li (2017). Our asymptotic analysis in Section 4 reveals useful insight of the properties of the proposed method, particularly, an explicit quantification of the efficiency gain. On the other hand, in applications such as brain connectivity analysis, the sample size is quite limited, whereas the network dimensionality is high. A high-dimensional asymptotic test under a limited sample size is a very challenging question and is left as future research.

There is another family of covariance estimation methods, the covariance spectral models, including common principal components analysis (Flury, 1984) and its extensions (Flury, 1987; Boik, 2002; Schott, 1999), the shared subspace model (Franks and Hoff, 2016), and the sufficient covariance reducing model (Cook and Forzani, 2008). Our method is similar to this family, in that all assume that there exists a common and lower-dimensional subspace that sufficiently captures the individual covariance matrices. We analytically compare them in Section 2.4, and show some of those models are the special cases of ours. We also numerically compare them in Section 5, and demonstrate the superior performance of our method.

2. Model

2.1 *Preparations*

The following notation is used throughout our exposition. For a matrix $\boldsymbol{B} \in \mathbb{R}^{p\times d}$ with full column rank d, let $\mathcal{B} = \mathrm{span}(\boldsymbol{B}) \subseteq \mathbb{R}^p$ denote the subspace spanned by the columns of $\boldsymbol{B}$. Let $\boldsymbol{P}_{\mathcal{B}} = \boldsymbol{P}_{\boldsymbol{B}} = \boldsymbol{B}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}$ denote the projection onto $\mathcal{B}$, and $\boldsymbol{Q}_{\mathcal{B}} = \boldsymbol{Q}_{\boldsymbol{B}} = \boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{B}}$ denote the projection onto the orthogonal complement of $\mathcal{B}$, where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. Let $\boldsymbol{P}_{\boldsymbol{B}(\boldsymbol{M})} = \boldsymbol{B}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{M}$ denote the projection onto $\mathcal{B}$ with respect to the $\boldsymbol{M}$-inner product, and $\boldsymbol{Q}_{\boldsymbol{B}(\boldsymbol{M})} = \boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{B}(\boldsymbol{M})}$, where $\boldsymbol{M}$ is a symmetric positive-definite matrix.

The next two concepts are essential for our method, and were defined in Cook et al. (2010). A reducing subspace of a symmetric positive-definite matrix $\boldsymbol{M} \in \mathbb{R}^{p\times p}$ is defined as the subspace

$\mathcal{R} \subseteq \mathbb{R}^p$ such that $\boldsymbol{M} = \boldsymbol{P}_{\mathcal{R}}\boldsymbol{M}\boldsymbol{P}_{\mathcal{R}} + \boldsymbol{Q}_{\mathcal{R}}\boldsymbol{M}\boldsymbol{Q}_{\mathcal{R}}$. It is easy to see that the subspace spanned by any set of eigenvectors of $\boldsymbol{M}$ is a reducing subspace of $\boldsymbol{M}$, though the reverse is not true. An $\boldsymbol{M}$-envelope of a subspace $\mathcal{B} \subseteq \mathbb{R}^p$ is defined as the intersection of all reducing subspaces of $\boldsymbol{M}$ that contain $\mathcal{B}$, and is denoted as $\mathcal{E}_{\boldsymbol{M}}(\mathcal{B})$.

2.2 *Common reducing subspace model*

We begin with covariance matrices. Let $\boldsymbol{D}_{i,k} \in \mathbb{R}^{p\times p}$ denote the covariance matrix for subject i in group k, $i = 1, \ldots, n_k, k = 1, \ldots, K$. If the parameter of interest is the correlation matrix, we first obtain an estimate of the covariance matrix then standardize it. Assume $\boldsymbol{D}_{i,k}$ follows a Wishart distribution with mean $\mathrm{E}(\boldsymbol{D}_{i,k}) = \boldsymbol{\Delta}_k$, $i = 1, \ldots, n_k$. Wishart distribution is commonly used to characterize the probability distribution of symmetric, nonnegative-definite random matrices. We then propose the following common reducing subspace model

$$\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}, \quad k = 1, \ldots, K, \tag{1}$$

where $\boldsymbol{\Omega}_k \in \mathbb{R}^{u\times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u)\times(p-u)}$ are symmetric positive definite matrices, $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times u}$ and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{p\times(p-u)}$ together form an orthonormal basis of $\mathbb{R}^p$, and $u \leqslant p$ is the smallest dimension such that (1) holds. This model is closely related to several existing covariance models, and we compare them in Section 2.4. It says that, there exists a common subspace $\mathcal{S} = \mathrm{span}(\boldsymbol{\Gamma}) \subseteq \mathbb{R}^p$, such that $\mathcal{S}$ is a reducing subspace for all $\{\boldsymbol{\Delta}_k\}_{k=1}^K$. Moreover, the projection of $\boldsymbol{\Delta}_k$ onto $\mathcal{S}$ has a different representation $\boldsymbol{\Omega}_k$, while its projection onto the orthogonal complement of $\mathcal{S}$ has the same representation $\boldsymbol{\Omega}_0$ regardless of the group k. In other words, $\mathcal{S}$ captures all the heterogeneous variations in the covariance matrices across multiple groups. It is easy to see that model (1) is equivalent to the following coordinate-free representation,

$$\boldsymbol{\Delta}_k = \boldsymbol{P}_{\mathcal{S}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\mathcal{S}} + \boldsymbol{Q}_{\mathcal{S}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\mathcal{S}}, \quad \boldsymbol{Q}_{\mathcal{S}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta}_j)\boldsymbol{Q}_{\mathcal{S}} = 0, \quad j, k = 1, \ldots, K. \tag{2}$$

This equivalence is formally established in Proposition 1. Our goal is to estimate this common subspace $\mathcal{S}$, or its basis $\boldsymbol{\Gamma}$, plus $\{\boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_K\}$. From those, we in turn obtain the estimates of the between-group covariance difference $\boldsymbol{\Delta}_k - \boldsymbol{\Delta}_j = \boldsymbol{\Gamma}(\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_j)\boldsymbol{\Gamma}^{\mathrm{T}}$, as well as the group-specific

covariance $\boldsymbol{\Delta}_k$. Model (1) leads to substantial dimension reduction in terms of the number of free parameters, especially when $u << p$. The number of unknowns is originally $Kp(p+1)/2$, and is now reduced to $p(p+1)/2+(K-1)u(u+1)/2$. The reduction is $(K-1)\{p(p+1)-u(u+1)\}/2$, which is the difference between the full covariance matrix $\boldsymbol{\Delta}_k$ and the heterogeneity component $\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Delta}_k\boldsymbol{\Gamma} = \boldsymbol{\Omega}_k$ multiplied by a factor of $(K-1)$.

We next consider precision matrices. Precision, or its standardized version, partial correlation, is frequently used in brain connectivity analysis, thanks to its conditional independence interpretation under the normal distribution (Wang et al., 2016). Without normality, it can still be a parameter of interest. Under the Wishart distribution assumption of $\boldsymbol{D}_{i,k}$, the precision matrix $\boldsymbol{D}_{i,k}^{-1}$ follows an inverse Wishart distribution with mean $\boldsymbol{\Delta}_k^{-1}$ up to a constant. We note that the common reducing subspace structure in (1) remains invariant when switching from $\boldsymbol{\Delta}_k$ to $\boldsymbol{\Delta}_k^{-1}$; that is

$$\boldsymbol{\Delta}_k^{-1} = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k^{-1}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^{\mathrm{T}}, \quad k = 1, \ldots, K. \tag{3}$$

In other words, the same reducing subspace $\mathcal{S}$ captures all the heterogeneous variations in the precision matrices across multiple groups as well. An analogy of this type of invariance can be found in matrix inverse calculation, where the inverse is applied to the eigenvalues of the matrix whereas the eigenvectors remain the same. Equivalently, (3) can be written as

$$\boldsymbol{\Delta}_k^{-1} = \boldsymbol{P}_{\mathcal{S}}\boldsymbol{\Delta}_k^{-1}\boldsymbol{P}_{\mathcal{S}} + \boldsymbol{Q}_{\mathcal{S}}\boldsymbol{\Delta}_k^{-1}\boldsymbol{Q}_{\mathcal{S}}, \quad \boldsymbol{Q}_{\mathcal{S}}(\boldsymbol{\Delta}_k^{-1} - \boldsymbol{\Delta}_j^{-1})\boldsymbol{Q}_{\mathcal{S}} = 0, \quad j, k = 1, \ldots, K. \tag{4}$$

The next proposition establishes the equivalence of the above statements, and we have an unified way of modeling covariance and precision matrices.

PROPOSITION 1: The four statements (1), (2), (3), and (4) are equivalent.

2.3 *Existence and uniqueness*

The common reducing subspace structure in (1) always exists, since one can trivially take $\mathcal{S} = \mathbb{R}^p$. However $\mathcal{S}$ is not unique. The idea is then to seek the intersection of all subspaces that satisfy (1), and this intersection is the smallest and unique subspace that satisfies (1). This leads to the notion

of covariance and precision envelope. Define $\mathcal{C} = \text{span}\{\boldsymbol{\Delta}_j - \boldsymbol{\Delta}_k \mid j, k = 1, \ldots, K\} \subseteq \mathbb{R}^p$ as the subspace that contains all the covariance variations across groups, and similarly, $\mathcal{P} = \text{span}\{\boldsymbol{\Delta}_j^{-1} - \boldsymbol{\Delta}_k^{-1} \mid j, k = 1, \ldots, K\} \subseteq \mathbb{R}^p$. According to (2), $\mathcal{S}$ is a reducing subspace of $\boldsymbol{\Delta}_k$ that contains $\mathcal{C}$. We define the intersection of all such reducing subspaces as the covariance envelope and denote it by $\mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{C})$. Similarly, we define the precision envelope $\mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{P})$, and the envelopes $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C})$ and $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{P})$ with respect to $\boldsymbol{\Delta}$. Here $\boldsymbol{\Delta} = \lim_{n\to\infty} \sum_{k=1}^K (n_k/n)\boldsymbol{\Delta}_k$, and aggregates $\boldsymbol{\Delta}_k$ across groups, and $n = \sum_{k=1}^K n_k$. The next proposition shows that those envelopes are all equal, so we only need to focus on one of them in subsequent estimation and inference.

PROPOSITION 2: $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) = \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{P}) = \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{C}) = \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{P})$ for all $k = 1, \ldots, p$.

The envelope $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C})$ uniquely exists, and is a key object of interest in our envelope-based covariance and precision estimation. Envelope is a nascent dimension reduction and efficient estimation technique first developed for multivariate-response regression (Cook et al., 2010), and later extended to many other high-dimensional regression models (Cook et al., 2013; Cook and Zhang, 2015; Su et al., 2016; Li and Zhang, 2017). However, this article is the first to develop the concept of covariance and precision envelopes.

2.4 *Connections with covariance spectral models*

Flury (1984) proposed common principal components analysis, where the goal is to find a set of common eigenvectors $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$ such that $\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Delta}_k\boldsymbol{\Psi}$ are simultaneously diagonalized. This can be restrictive, since it requires all $\boldsymbol{\Delta}_k$'s to have the same set of eigenvectors. Extensions have been made to allow to share only a subset but not all of the eigenvectors (Flury, 1987), or to allow no common eigenvectors but they span the same subspace (Schott, 1999; Boik, 2002). Nevertheless, this family of solutions required restrictive conditions on the eigenvectors, whereas our solution places no such restrictions.

Franks and Hoff (2016) proposed the shared subspace model, which assumes there exists a subspace $\mathrm{span}(\boldsymbol{\Gamma})$, $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times u}$, such that $\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Lambda}_k\boldsymbol{\Gamma}^{\mathrm{T}} + \sigma^2\boldsymbol{I}_p$, $\boldsymbol{\Lambda}_k > 0$. It is equivalent to

$$\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}(\boldsymbol{\Lambda}_k + \sigma^2\boldsymbol{I}_u)\boldsymbol{\Gamma}^{\mathrm{T}} + \sigma^2\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}, \quad \boldsymbol{\Lambda}_k > 0.$$

Therefore this model is a special case of our model (1). It assumes that the common covariance is isotropic; that is, $\boldsymbol{\Omega}_0 = \sigma^2\boldsymbol{I}_{p-u}$. It also restricts the shared subspace $\mathrm{span}(\boldsymbol{\Gamma})$ to contain the leading eigenvalues, since $\boldsymbol{\Omega}_k = \boldsymbol{\Lambda}_k + \sigma^2\boldsymbol{I}_u$ now has larger eigenvalues than $\boldsymbol{\Omega}_0$.

Cook and Forzani (2008) proposed the covariance reducing model for characterizing the behavior of K covariance matrices $\{\boldsymbol{D}_k\}_{k=1}^K$. It seeks to find the minimum subspace $\mathrm{span}(\boldsymbol{\Upsilon})$, with $\boldsymbol{\Upsilon} \in \mathbb{R}^{p\times u}$, which is sufficient to capture the variance heterogeneity among all groups, in the sense that

$$\boldsymbol{D}_j \mid (\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{D}_j\boldsymbol{\Upsilon} = \boldsymbol{B}) \sim \boldsymbol{D}_k \mid (\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{D}_k\boldsymbol{\Upsilon} = \boldsymbol{B}), \quad j,k = 1,\ldots,K.$$

That is, the conditional distribution of $\boldsymbol{D}_k \mid \boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{D}_k\boldsymbol{\Upsilon}$ does not depend on k, and thus $\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{D}_k\boldsymbol{\Upsilon}$ is a sufficient dimension reduction of $\boldsymbol{D}_k$. Under the assumption that $\boldsymbol{D}_k$ follows a Wishart distribution with mean $\boldsymbol{\Delta}_k$, Cook and Forzani (2008) showed that $\boldsymbol{\Upsilon}$ satisfies

$$\boldsymbol{\Delta}_k = \boldsymbol{\Delta} + \boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})}^{\mathrm{T}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})}, \tag{5}$$

where $\boldsymbol{\Delta}$ is as defined before. In general, $\mathrm{span}(\boldsymbol{\Upsilon})$ from model (5) is contained in $\mathrm{span}(\boldsymbol{\Gamma})$ from our model (1). Hence (5) is targeting at a potentially smaller subspace than our common reducing subspace. However, this is not surprising, because the goal of (5) is to find the minimal sufficient reduction only, while our model aims to both achieve dimension reduction and also to provide a direct and more efficient estimator for $\boldsymbol{\Delta}_k$. For the special case when $\mathrm{span}(\boldsymbol{\Upsilon})$ itself is a reducing subspace of $\boldsymbol{\Delta}$, then $\boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})} = \boldsymbol{P}_{\boldsymbol{\Upsilon}}$, and thus (5) is simplified as $\boldsymbol{\Delta}_k = \boldsymbol{\Delta} + \boldsymbol{P}_{\boldsymbol{\Upsilon}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{P}_{\boldsymbol{\Upsilon}}$. Consequently, $\boldsymbol{Q}_{\boldsymbol{\Upsilon}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{Q}_{\boldsymbol{\Upsilon}} = 0$. By (2), models (1) and (5) become equivalent. In the Supplementary Materials, Section A2.1, we present additional theoretical results that connect our model with that of Cook and Forzani (2008).

3. Estimation

3.1 *Maximum likelihood estimation*

Consider i.i.d. samples $\boldsymbol{D}_{i,k}$, $i = 1, \ldots, n_k$, $k = 1, \ldots, K$, that follow a Wishart distribution $W_p(\boldsymbol{\Delta}_k/d, d)$, where d is the degrees of freedom. The standard maximum likelihood estimators of $\boldsymbol{\Delta}_k$ and $\boldsymbol{\Delta}$ are simply the sample means

$$\overline{\boldsymbol{D}}_k = n_k^{-1} \sum_{i=1}^{n_k} \boldsymbol{D}_{i,k} \quad \text{and} \quad \overline{\boldsymbol{D}} = \sum_{k=1}^{K} (n_k/n) \overline{\boldsymbol{D}}_k.$$

Next we pursue the maximum likelihood estimator of the individual covariance matrix $\boldsymbol{\Delta}_k$ under the common reducing subspace model (1). The next proposition summarizes the result.

PROPOSITION 3: Under the Wishart distribution assumption, the maximum likelihood estimator of $\boldsymbol{\Delta}_k$ is $\widehat{\boldsymbol{\Delta}}_k = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\overline{\boldsymbol{D}}_k\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}} + \widehat{\boldsymbol{\Gamma}}_0\widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}}\overline{\boldsymbol{D}}\widehat{\boldsymbol{\Gamma}}_0\widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}}$, where $\widehat{\boldsymbol{\Gamma}} \in \mathbb{R}^{p\times u}$ is the minimizer of the following objective function, subject to the constraint that $\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma} = \boldsymbol{I}_u$,

$$\ell(\boldsymbol{\Gamma}) = nd \log |\boldsymbol{\Gamma}^{\mathrm{T}}(\overline{\boldsymbol{D}})^{-1}\boldsymbol{\Gamma}| + \sum_{k=1}^{K} n_k d \log |\boldsymbol{\Gamma}^{\mathrm{T}}\overline{\boldsymbol{D}}_k\boldsymbol{\Gamma}|, \tag{6}$$

$\widehat{\boldsymbol{\Gamma}}_0$ is the orthonormal basis that is complement to $\widehat{\boldsymbol{\Gamma}}$, $\widehat{\boldsymbol{\Omega}}_k = \widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\overline{\boldsymbol{D}}_k\widehat{\boldsymbol{\Gamma}}$, and $\widehat{\boldsymbol{\Omega}}_0 = \widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}}\overline{\boldsymbol{D}}\widehat{\boldsymbol{\Gamma}}_0$.

Based on Proposition 3, once we obtain the minimizer $\widehat{\boldsymbol{\Gamma}}$ of (6) under the orthogonal constraint, we can obtain the estimators $\widehat{\boldsymbol{\Gamma}}_0$, $\widehat{\boldsymbol{\Omega}}_k$, $\widehat{\boldsymbol{\Omega}}_0$, and $\widehat{\boldsymbol{\Delta}}_k$. Subsequently, we estimate the between-group covariance matrix difference as $\widehat{\boldsymbol{\Delta}}_k - \widehat{\boldsymbol{\Delta}}_j = \widehat{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\Omega}}_k - \widehat{\boldsymbol{\Omega}}_j)\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}$.

Moreover, denoting $\boldsymbol{\Phi}_k = \boldsymbol{\Delta}_k^{-1}$, we can directly obtain the maximum likelihood estimator of the individual precision matrix under model (1) as $\widehat{\boldsymbol{\Phi}}_k = \widehat{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\overline{\boldsymbol{D}}_k\widehat{\boldsymbol{\Gamma}})^{-1}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}} + \widehat{\boldsymbol{\Gamma}}_0(\widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}}\overline{\boldsymbol{D}}\widehat{\boldsymbol{\Gamma}}_0)^{-1}\widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}}$. Accordingly, we estimate the between-group precision matrix difference as $\widehat{\boldsymbol{\Phi}}_k - \widehat{\boldsymbol{\Phi}}_j = \widehat{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\Omega}}_k^{-1} - \widehat{\boldsymbol{\Omega}}_j^{-1})\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}} = \widehat{\boldsymbol{\Gamma}}\left\{(\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\overline{\boldsymbol{D}}_k\widehat{\boldsymbol{\Gamma}})^{-1} - (\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\overline{\boldsymbol{D}}_j\widehat{\boldsymbol{\Gamma}})^{-1}\right\}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}$.

The constrained minimization of (6) is done through gradient descent on a Grassmann manifold. We employ the manifold optimization method of Huang et al. (2015) and its R implementation `ManifoldOptim` (Martin et al., 2016). In particular, we obtain the following closed-form ex-

pression for the gradient of $\ell(\mathbf{\Gamma})$, which is to greatly facilitate the computation,

$$\frac{\mathrm{d}\ell(\mathbf{\Gamma})}{\mathrm{d}\mathbf{\Gamma}} = 2dn\,\overline{\boldsymbol{D}}^{-1}\mathbf{\Gamma}\left\{\mathbf{\Gamma}^{\mathrm{T}}(\overline{\boldsymbol{D}})^{-1}\mathbf{\Gamma}\right\}^{-1} + \sum_{k=1}^{K} 2dn_k\,\overline{\boldsymbol{D}}_k\mathbf{\Gamma}\left(\mathbf{\Gamma}^{\mathrm{T}}\overline{\boldsymbol{D}}_k\mathbf{\Gamma}\right)^{-1}.$$

With a good initial value, the manifold optimization procedure converges fast. We next examine initialization and dimension selection.

3.2 *Initialization and dimension selection*

Initialization is important for the constrained minimization of $\ell(\mathbf{\Gamma})$. To obtain a good initial estimator of $\mathbf{\Gamma}$, we adopt the sequential one-direction-at-a-time approach (Cook and Zhang, 2016). It is fast, stable, and widely used in the envelope literature (e.g. Cook and Zhang, 2015; Li and Zhang, 2017). Specifically, for $m = 0, \ldots, p-1$, let $\boldsymbol{g}_m \in \mathbb{R}^{p\times 1}$ denote the mth sequential direction to be obtained. Let $\boldsymbol{G}_m = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m)$, and let $(\boldsymbol{G}_m, \boldsymbol{G}_{0m})$ be an orthogonal basis for $\mathbb{R}^p$. We set $\boldsymbol{g}_0 = \boldsymbol{G}_{00} = 0$. Define $\overline{\boldsymbol{D}}_m = \boldsymbol{G}_{0m}^{\mathrm{T}}\overline{\boldsymbol{D}}\boldsymbol{G}_{0m}$ and $\overline{\boldsymbol{D}}_{k,m} = \boldsymbol{G}_{0m}^{\mathrm{T}}\overline{\boldsymbol{D}}_k\boldsymbol{G}_{0m}$, $k = 1, \ldots, K$. The objective function after the first m sequential steps is

$$\phi_{m+1}(\boldsymbol{w}) = nd\log\{\boldsymbol{w}^{\mathrm{T}}(\overline{\boldsymbol{D}}_m)^{-1}\boldsymbol{w}\} + \sum_{k=1}^{K} n_k d\log(\boldsymbol{w}^{\mathrm{T}}\overline{\boldsymbol{D}}_{k,m}\boldsymbol{w}).$$

We minimize $\phi_{m+1}(\boldsymbol{w})$ over $\boldsymbol{w} \in \mathbb{R}^{p-m}$ subject to $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} = 1$, and denote the minimizer as $\widehat{\boldsymbol{w}}_{m+1}$. Then the $(m+1)$th direction is $\boldsymbol{g}_{m+1} = \boldsymbol{G}_{0m}\widehat{\boldsymbol{w}}_{m+1}$. Given the dimension u of the common reducing subspace, we take $\boldsymbol{G}_u \in \mathbb{R}^{p\times u}$ as the initial estimator for $\mathbf{\Gamma}$.

The computational complexity of the initialization is $O\{(K+1)(p^3u + u^3p/6) + pu^2 + pu\}$ per iteration, where $(K+1)(p^3u+u^3p/6)+pu^2$ comes from the matrix multiplication and inversion within the iteration, and pu comes from the optimization over Grassmann manifold of dimension one on $\mathbb{R}^p$. The computational complexity of the full Grassmann manifold optimization after the initialization is $O(pu^2)$ per iteration. So the per iteration complexity is at the order of $O(p^3)$, which is in the same order as matrix multiplication of two $p \times p$ matrices, or matrix inversion of a $p \times p$ matrix. Practically, we find our method runs reasonably fast. We report the computation time, as well as the effect of initialization, in the Supplementary Materials, Sections A1.1 and A1.2, where

we found that the manifold optimization converges quickly after the 1D initialization but a random initialization usually results in a much larger estimation error.

To estimate the common reducing subspace dimension u, we adopt a BIC type criterion of Zhang et al. (2018) that is built on the above sequential estimation. Specifically, we select the dimension u by minimizing

$$\mathcal{I}_{1\mathrm{D}}(m) = \sum_{j=1}^{m} \phi_j(\widehat{\boldsymbol{w}}_j) + m\frac{(K-1)p}{2}\log(nd), \quad m = 1, ..., p, \tag{7}$$

and $\mathcal{I}_{1\mathrm{D}}(0)$ is defined as 0. We seek $m \in \{0, \ldots, p\}$ that minimizes $\mathcal{I}_{1\mathrm{D}}(m)$. This criterion is shown to be computationally efficient and accurate (Zhang et al., 2018). Once obtaining an estimator of u, then an initial estimator of $\boldsymbol{\Gamma}$, we resort to the gradient descent procedure in Section 3.1 to obtain the maximum likelihood estimator of $\boldsymbol{\Gamma}$.

4. Theory

We study the asymptotic properties of the proposed estimator, and compare with the standard sample estimator. For notational simplicity, we present our results for $K = 2$ only, but the conclusion holds for $K > 2$. We focus on the properties of the covariance estimators, while the properties of the precision estimators can be derived similarly. We first consider the scenario when the data follow a Wishart distribution, and later relax this distributional assumption.

Define two vector operators, $\mathrm{vec}(\cdot) : \mathbb{R}^{p\times q} \mapsto \mathbb{R}^{pq\times 1}$ that vectorizes a matrix by stacking all its columns, and $\mathrm{vech}(\cdot) : \mathbb{R}^{p\times p} \mapsto \mathbb{R}^{p(p+1)/2\times 1}$ that vectorizes a symmetric matrix by extracting its columns of elements below or on the diagonal. Define $\boldsymbol{\theta}$ that collects all the parameters in (1), and the estimable function $\boldsymbol{h} = \boldsymbol{h}(\boldsymbol{\theta})$ that involves the targeting parameters of interest $\boldsymbol{\Delta}_k$; that is,

$$\boldsymbol{\theta} = \begin{pmatrix} \mathrm{vec}\ \boldsymbol{\Gamma} \\ \mathrm{vech}\ \boldsymbol{\Omega}_1 \\ \mathrm{vech}\ \boldsymbol{\Omega}_2 \\ \mathrm{vech}\ \boldsymbol{\Omega}_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \boldsymbol{\theta}_3 \\ \boldsymbol{\theta}_4 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{h} = \begin{pmatrix} \mathrm{vech}\ \boldsymbol{\Delta}_1 \\ \mathrm{vech}\ \boldsymbol{\Delta}_2 \end{pmatrix}.$$

We first establish and compare the asymptotic properties of the standard sample mean estimator,

$\widehat{\boldsymbol{h}}_{\text{std}} = \{\text{vech}^{\text{T}}(\overline{\boldsymbol{D}}_1), \text{vech}^{\text{T}}(\overline{\boldsymbol{D}}_2)\}^{\text{T}}$, and our estimator under the common reducing subspace model (1), $\widehat{\boldsymbol{h}}_{\text{crs}} = \{\text{vech}^{\text{T}}(\widehat{\boldsymbol{\Delta}}_1), \text{vech}^{\text{T}}(\widehat{\boldsymbol{\Delta}}_2)\}^{\text{T}}$. In the following, for any $\boldsymbol{M} \in \mathcal{S}^{p\times p}$, $\boldsymbol{M}^{\dagger}$ denotes the Moore-Penrose inverse; $\boldsymbol{C}_p$ and $\boldsymbol{E}_p$ are defined such that $\text{vech}(\boldsymbol{M}) = \boldsymbol{C}_p\text{vec}(\boldsymbol{M})$ and $\text{vec}(\boldsymbol{M}) = \boldsymbol{E}_p\text{vech}(\boldsymbol{M})$.

THEOREM 1: *Assuming the observed data matrices $\boldsymbol{D}_{i,k}$, $i = 1, \ldots, n_k$, $k = 1, \ldots, K$, are i.i.d. samples from a Wishart distribution $W_p(\boldsymbol{\Delta}_k/d, d)$ with d degrees of freedom, we have*

$$\sqrt{n}(\widehat{\boldsymbol{h}}_{\text{std}} - \boldsymbol{h}) \xrightarrow{D} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{V}), \quad \sqrt{n}(\widehat{\boldsymbol{h}}_{\text{crs}} - \boldsymbol{h}) \xrightarrow{D} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{U}),$$

where $\boldsymbol{V} = \boldsymbol{J}_h^{-1}$ is the inverse of the Fisher information matrix,

$$\boldsymbol{J}_h = \begin{pmatrix} \frac{d\pi_1}{2}\boldsymbol{E}_p^{\text{T}}(\boldsymbol{\Delta}_1^{-1} \otimes \boldsymbol{\Delta}_1^{-1})\boldsymbol{E}_p & \boldsymbol{0} \\ \boldsymbol{0} & \frac{d\pi_2}{2}\boldsymbol{E}_p^{\text{T}}(\boldsymbol{\Delta}_2^{-1} \otimes \boldsymbol{\Delta}_2^{-1})\boldsymbol{E}_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{J}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_2 \end{pmatrix},$$

$\boldsymbol{U} = \boldsymbol{H}(\boldsymbol{H}^{\text{T}}\boldsymbol{J}_h\boldsymbol{H})^{\dagger}\boldsymbol{H}^{\text{T}}$, and $\boldsymbol{H}$ is the gradient matrix $\partial\boldsymbol{h}/\partial\boldsymbol{\theta}$ of the form

$$\boldsymbol{H} = \begin{pmatrix} 2\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1 \otimes \boldsymbol{I}_p - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\text{T}}) & \boldsymbol{C}_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\boldsymbol{E}_u & \boldsymbol{0} & \boldsymbol{C}_p(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\boldsymbol{E}_{p-u} \\ 2\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_2 \otimes \boldsymbol{I}_p - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\text{T}}) & \boldsymbol{0} & \boldsymbol{C}_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\boldsymbol{E}_u & \boldsymbol{C}_p(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\boldsymbol{E}_{p-u} \end{pmatrix}.$$

Moreover, we have $\boldsymbol{U} \leqslant \boldsymbol{V}$, in that $\boldsymbol{V} - \boldsymbol{U}$ is a positive semi-definite matrix.

The last result implies that our common reducing subspace estimator under model (1) is asymptotically more efficient than the standard estimator. It also implies that any estimable functions of $\boldsymbol{h}$ can be estimated more efficiently via $\widehat{\boldsymbol{h}}_{\text{crs}}$ than via $\widehat{\boldsymbol{h}}_{\text{std}}$. As a direct result, $\text{avar}\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_k)\} \leqslant \text{avar}\{\sqrt{n}\text{vech}(\overline{\boldsymbol{D}}_k)\}$, where $\text{avar}(\cdot)$ stands for the asymptotic variance.

To better understand the asymptotic efficiency gain of our estimators, we further decompose the asymptotic variances of $\widehat{\boldsymbol{\Delta}}_k$ and $\widehat{\boldsymbol{\Delta}}_k - \widehat{\boldsymbol{\Delta}}_j$. Toward that end, we introduce an intermediate estimator $\widehat{\boldsymbol{\Delta}}_{k,\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\text{T}}\overline{\boldsymbol{D}}_k\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\text{T}}$, which is built on a known $\boldsymbol{\Gamma}$. Then we have the following result.

THEOREM 2: *Under the same assumption as in Theorem 1, we have*

$$\begin{aligned}\text{avar}\big\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_k)\big\} &= \text{avar}\big\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_{k,\boldsymbol{\Gamma}})\big\} + \text{avar}\big\{\sqrt{n}\text{vech}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{k,\boldsymbol{\Omega}_k,\boldsymbol{\Omega}_0}\boldsymbol{Q}_{\boldsymbol{\Gamma}})\big\} \\ &\quad + \text{avar}\big\{\sqrt{n}\text{vech}(\boldsymbol{P}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{k,\boldsymbol{\Omega}_k,\boldsymbol{\Omega}_0}\boldsymbol{P}_{\boldsymbol{\Gamma}})\big\},\ \textit{for } k=1,2, \qquad (8)\\ \text{avar}\big\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_1-\widehat{\boldsymbol{\Delta}}_2)\big\} &= \text{avar}\big\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Gamma}}-\widehat{\boldsymbol{\Delta}}_{2,\boldsymbol{\Gamma}})\big\} \\ &\quad + \text{avar}\big\{\sqrt{n}\text{vech}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}-\widehat{\boldsymbol{\Delta}}_{2,\boldsymbol{\Omega}_2,\boldsymbol{\Omega}_0})\boldsymbol{Q}_{\boldsymbol{\Gamma}})\big\}, \qquad (9)\end{aligned}$$

where $\widehat{\boldsymbol{\Delta}}_{k,\boldsymbol{\Omega}_k,\boldsymbol{\Omega}_0}$ *denotes the estimator of* $\boldsymbol{\Delta}_k$ *when* $\boldsymbol{\Omega}_k, \boldsymbol{\Omega}_0$ *are treated as known.*

According to (8), the asymptotic variance of the envelope-based covariance estimator $\widehat{\boldsymbol{\Delta}}_k$ is decomposed into three terms. The first term in (8) is the asymptotic variance of the intermediate estimator $\widehat{\boldsymbol{\Delta}}_{k,\boldsymbol{\Gamma}}$ when $\boldsymbol{\Gamma}$ is known, and the second and third terms together are the additional asymptotic cost of estimating $\boldsymbol{\Gamma}$. For the first term, we observe that

$$\text{avar}\big\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_{k,\boldsymbol{\Gamma}})\big\} = \boldsymbol{\Gamma}(\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{J}_k\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^{\mathrm{T}} < \boldsymbol{J}_k^{-1} = \text{avar}\big\{\sqrt{n}\text{vech}(\overline{\boldsymbol{D}}_k)\big\}.$$

Therefore, when $\boldsymbol{\Gamma}$ is known, our estimator achieves a clear efficiency gain compared to the standard estimator. This gain is more substantial when the dimension u of the envelope $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C})$ is small compared to p. On the other hand, when $\boldsymbol{\Gamma}$ is unknown and is estimated given the data, the asymptotic variance of our estimator is still smaller than that of the standard estimator, as indicated by Theorem 1.

The asymptotic cost of estimating $\boldsymbol{\Gamma}$ is further decomposed into the second and third terms in (8), where one is a projection into $\text{span}(\boldsymbol{\Gamma})$, and the other is a projection into the orthogonal complement of $\text{span}(\boldsymbol{\Gamma})$. The term within $\text{span}(\boldsymbol{\Gamma})$ vanishes when we estimate the difference of the two covariances. This leads to (9), which implies that the asymptotic efficiency gain is even more substantial when we focus on the difference $\boldsymbol{\Delta}_1 - \boldsymbol{\Delta}_2$ rather than the individual $\boldsymbol{\Delta}_k$.

Finally, we show that, when the data do not follow a Wishart distribution, but instead a general symmetric matrix distribution, our estimators from Proposition 3 are still $\sqrt{n}$-consistent and asymptotically normal, under some mild moment conditions.

THEOREM 3: *Assuming the observed data matrices* $\boldsymbol{D}_{i,k}$, $i = 1, \ldots, n_k$, $k = 1, \ldots, K$, *are i.i.d. samples from a symmetric matrix distribution, with mean* $\boldsymbol{\Delta}_k$, *and* $\mathrm{E}\{\mathrm{vech}(\boldsymbol{D}_{i,k}^2)\} < \infty$, *then* $\sqrt{n}(\widehat{\boldsymbol{h}}_{\text{std}} - \boldsymbol{h}) \xrightarrow{D} \boldsymbol{N}(0, \boldsymbol{W})$ *for some positive definite covariance matrix* $\boldsymbol{W}$, *and* $\sqrt{n}(\widehat{\boldsymbol{h}}_{\text{crs}} - \boldsymbol{h}) \xrightarrow{D} \boldsymbol{N}(0, \boldsymbol{K})$, *where* $\boldsymbol{K} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H})^{\dagger}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{W}\boldsymbol{J}_h\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H})^{\dagger}\boldsymbol{H}^{\mathrm{T}}$. *Moreover,* $\boldsymbol{K} \leqslant \boldsymbol{W}$ *if* $\mathrm{span}(\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{H})$ *is a reducing subspace of* $\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{J}_h^{\frac{1}{2}}$.

Even without the Wishart distribution assumption, our common reducing subspace-based estimator is still potentially more efficient than the standard estimator.

We also present some theoretical result when the sample size $n = \sum_{k=1}^{K} n_k$ is fixed in the Supplementary Materials, Section A2.2.

5. Simulations

5.1 *Covariance matrix estimation*

We study the empirical performance of our proposed common reducing subspace model, and compare with three alternative methods examined in Section 2.4. We investigate the covariance matrix estimation in this section, then the precision matrix estimation in the next. We first assume the true envelope dimension is known, and study its selection through (7) in Section 5.3.

We consider the following models to generate the covariance matrices.

$$\begin{aligned}
&\text{Model I:} && \boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T. \\
&\text{Model II:} && \boldsymbol{\Delta}_k = \sigma_k^2\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T, \qquad \text{where } \sigma_1 = 0.1, \sigma_2 = 0.2. \\
&\text{Model III:} && \boldsymbol{\Delta}_k = \sigma_k^2\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T + 0.1\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T, \quad \text{where } \sigma_1 = 1, \sigma_2 = 2.
\end{aligned}$$

Here we first generate the basis matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times u}$ with all its elements from a random uniform $(0, 1)$ distribution, then orthonormalize this matrix. We construct $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{p\times(p-u)}$ such that $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ forms an orthogonal basis of $\mathbb{R}^{p\times p}$. In Model I, $\boldsymbol{\Omega}_k$ is generated as $\boldsymbol{O}\boldsymbol{J}_k\boldsymbol{O}^T$, where $\boldsymbol{O}$ is an orthogonal matrix, and $\boldsymbol{J}_k$ is a diagonal matrix taking values $k \times (1, ..., u)$ as its diagonal elements, and $\boldsymbol{\Omega}_0$ is generated as $\boldsymbol{O}\boldsymbol{J}\boldsymbol{O}^T$, where $\boldsymbol{J}$ is a diagonal matrix taking equal-spaced values between e^{-2} and

e^2 as its diagonal elements. In Models II and III, $\mathbf{\Omega}_0$ is isotropic. Model II assumes $\mathbf{\Omega}_k$ contains the smallest eigenvalues of $\mathbf{\Delta}_k$, whereas Model III assumes $\mathbf{\Omega}_k$ contains the largest eigenvalues of $\mathbf{\Delta}_k$. Model III is constructed following the setup of the shared subspace model of Franks and Hoff (2016), whereas there is no specific ordering of the eigenvalues in $\mathbf{\Omega}_k$ and $\mathbf{\Omega}_0$ as in Model I. For all three models, we set the number of groups $K = 2$, the degree of freedom $d = 100$, the envelope dimension $u = 5$, the dimension of the covariance $p = 100, 200$, and the per-group sample size $n_k = 40, 100$.

We evaluate the performance by the estimation error $\|\mathbf{\Delta}_1 - \mathbf{\Delta}_2 - (\widehat{\mathbf{\Delta}}_1 - \widehat{\mathbf{\Delta}}_2)\|_F$, where $\|\cdot\|_F$ is the Frobenius norm. Table 1 reports the average estimation error over 100 replications. It is seen that our proposed estimator clearly outperforms the alternative methods in Models I and II. The huge difference in performance is due to the fact that, for Model I, although $\|\mathbf{\Delta}_2 - \mathbf{\Delta}_1\|_F$ is small, each individual $\|\mathbf{\Delta}_k\|_F, k = 1, 2$ is fairly large. When one of estimates for $\mathbf{\Delta}_k$ is poor, the estimation error is large. The shared subspace model of Franks and Hoff (2016) achieves the best performance in Model III, which is not surprising though, since model III is designed exactly following the specification of Franks and Hoff (2016). In this setting, the performance of our method is very close to that of Franks and Hoff (2016).

[Table 1 about here.]

5.2 *Precision matrix estimation*

Next we investigate the precision matrix estimation. Recall the notation $\mathbf{\Phi}_k = \mathbf{\Delta}_k^{-1}$. We consider the following models to generate the precision matrices.

Model I: $\mathbf{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}$, where $\boldsymbol{B}_k = \mathbf{\Gamma}\mathbf{\Psi}_k\mathbf{\Gamma}^T$, $\delta = 0.5$.

Model II: $\mathbf{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}$, where $\boldsymbol{B}_k = \sigma_k^2\mathbf{\Gamma}\mathbf{\Gamma}^T$, $\sigma_1 = 1, \sigma_2 = 2, \delta = 0.5$.

Model III: $\mathbf{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}$, where $\boldsymbol{B}_k = \sigma_k\mathbf{\Gamma}\mathbf{\Gamma}^T$, $\sigma_1 = -3, \sigma_2 = -4, \delta = 5$.

Here in Model I, we generate $\mathbf{\Psi}_k$ as $\boldsymbol{O}\boldsymbol{J}_k\boldsymbol{O}^T$, where $\boldsymbol{O}$ is an orthogonal matrix, and $\boldsymbol{J}_k$ is a diagonal matrix with $0.1 \times k \times (1, ..., u)$ as the diagonal elements. Moreover, we generate a sparse $\mathbf{\Gamma} \in \mathbb{R}^{p\times u}$

so that the precision matrix $\mathbf{\Phi}_k$ is sparse. Specifically, we first generate a semi-orthogonol $\mathbf{\Gamma}_1 \in \mathbb{R}^{p_1 \times u}$, where p_1 is chosen to keep the sparsity proportion of $\mathbf{\Phi}_k$ to be about 90%. We then randomly choose p_1 indexes out of $\{1, \ldots, p\}$, fill those rows of $\mathbf{\Gamma}$ with the rows from $\mathbf{\Gamma}_1$, and keep other elements of $\mathbf{\Gamma}$ to be 0. The rest of the setup follows that in Section 5.1.

We evaluate the performance by the estimation error $||\mathbf{\Phi}_1 - \mathbf{\Phi}_2 - (\widehat{\mathbf{\Phi}}_1 - \widehat{\mathbf{\Phi}}_2)||_F$. Table 2 reports the average estimation error over 100 replications. It is seen that our proposed method performs the best when estimating the difference of the precision matrices.

[Table 2 about here.]

In many applications, it is often of interest to obtain a sparse estimator of the precision matrix. Toward that end, we propose to first obtain an estimator of the covariance matrix, then feed it into the sparse column-wise inverse operator method of Liu and Luo (2015) to obtain a sparse precision matrix estimator. We have chosen the sparse precision estimation method of Liu and Luo (2015) due to its ease of use and fast computation. We also compare with the fused graphical Lasso estimator of Danaher et al. (2014), and the joint sparse precision estimator of Lee and Liu (2015). For a fair comparison, we have tuned the sparsity of all the methods in the same way. That is, we generate an independent validation data set in the same way as the training set, then choose the tuning parameter that minimizes the Bregman loss: $\widehat{\lambda} = \arg\min_\lambda \sum_k \left\{ \mathrm{trace}(\overline{\boldsymbol{D}}_k, \widehat{\mathbf{\Phi}}_{k,\lambda}) - \log\det(\widehat{\mathbf{\Phi}}_{k,\lambda}) \right\}$. Table 3 reports both the average estimation error as well as the sensitivity and specificity of nonzero elements selection based on 100 data replications. Compared to the family of covariance spectral models, our method yields the best performance in Models I and II, and a comparable performance as Franks and Hoff (2016) in Model III. Compared to the fused graphical Lasso estimator and the joint sparse precision estimator, our method continues to outperform the two alternatives. This is because our method is designed to utilize the shared low-dimensional subspace, but the methods of Danaher et al. (2014); Lee and Liu (2015) were not designed this way.

[Table 3 about here.]

5.3 *Subspace dimension selection*

Next we investigate the performance of the common reducing subspace dimension selection criterion (7). We employ the models in Section 5.2 with fixed $u = 5, d = 100$ and varying (p, n_k). Table 4 reports the percentage of times that the true dimension is selected by minimizing (7) out of 100 data replications. It is seen that the dimension selection is accurate, especially with a reasonable sample size.

[Table 4 about here.]

6. Brain connectivity analysis

In this section, we illustrate our proposed method on a brain connectivity analysis of an autism spectrum disorder data. Autism spectrum disorder is an increasingly prevalent neurodevelopmental disorder, and its symptoms include social difficulties, communication deficits, stereotyped behaviors and cognitive delays (Rudie et al., 2013). It is of central scientific interest to understand the interactions among numerous brain regions, and to identify the regions that exhibit different connectivity patterns between the subjects with the disorder and the normal controls. The data we analyzed was the resting-state functional magnetic resonance imaging from the Autism Brain Imaging Data Exchange (Di Martino et al., 2014). It consists of 795 subjects in two groups, with one group of 362 subjects with autism spectrum disorder, and the other group of 433 normal controls. For each subject, the imaging data was preprocessed and summarized in the form of a 116×116 covariance matrix, corresponding to the connectivities of 116 brain regions-of-interest from the Anatomical Automatic Labeling atlas (Tzourio-Mazoyer et al., 2002). See Chen et al. (2015a,b) for some analyses of the same data.

We applied our proposed method to this data. Since partial correlation is frequently employed to portray brain connectivity network (Ryali et al., 2012; Chen et al., 2013), we focused our analysis on sparse precision matrix estimation. Following our approach in Section 5.2, we first fitted the

common reducing subspace model to obtain an envelope-based covariance matrix estimator, then fed it into the method of Liu and Luo (2015). The envelope dimension selection criterion (7) was minimized at $u = 18$, while the sparsity parameter was tuned based on five-fold cross-validation. Figure 1 shows the top 20 links found by our method and the associated brain regions visualized using the BrainNet Viewer (Xia et al., 2013). Table 5 further reports those links. We found a number of brain regions exhibiting different connectivity patterns between the autism group and the normal control, including cerebellum, middle temporal gyrus, inferior temporal gyrus, and fusiform gyrus. These findings agree with the literature on autism studies (Di Martino et al., 2014; Cheng et al., 2015; Long et al., 2016). For instance, cerebellum has long been known for its importance in motor learning, coordination, and more recently, cognitive functions and affective regulation. It has emerged as one of the key brain regions affected in autism (Becker and Stoodley, 2013).

[Figure 1 about here.]

[Table 5 about here.]

ACKNOWLEDGEMENTS

Research for this paper was supported in part by grants CCF-1617691, DMS-1613154 and DMS-1613137 from the U.S. National Science Foundation.

SUPPLEMENTARY MATERIALS

Supplementary materials, including additional simulations and proofs, are available with this paper at the Biometrics website on Wiley Online Library.

REFERENCES

Ahn, M., Shen, H., Lin, W., and Zhu, H. (2015). A sparse reduced rank framework for group analysis of functional neuroimaging data. *Statistica Sinica*, 25:295–312.

Becker, E. B. and Stoodley, C. J. (2013). Chapter one - autism spectrum disorder and the cerebellum. In Konopka, G., editor, *Neurobiology of Autism*, volume 113 of *International Review of Neurobiology*, pages 1 – 34. Academic Press.

Boik, R. J. (2002). Spectral models for covariance matrices. *Biometrika*, 89(1):159–182.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, 10:186–198.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684.

Cai, T. T., Li, H., Liu, W., and Xie, J. (2016). Joint estimation of multiple high-dimensional precision matrices. *Statistica Sinica*, 26:445–464.

Cai, T. T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.*, 106(494):594–607.

Chen, S., Kang, J., and Wang, G. (2015a). An empirical bayes normalization method for connectivity metrics in resting state fmri. *Frontiers in Neuroscience*, 9(316).

Chen, S., Kang, J., Xing, Y., and Wang, G. (2015b). A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. *Human Brain Mapping*, 36:5196–5206.

Chen, T., Ryali, S., Qin, S., and Menon, V. (2013). Estimation of resting-state functional connectivity using random subspace based partial correlation: A novel method for reducing global artifacts. *NeuroImage*, 82:87–100.

Cheng, W., Rolls, E. T., Gu, H., Zhang, J., and Feng, J. (2015). Autism: reduced connectivity between cortical areas involved in face expression, theory of mind, and the sense of self. *Brain*, 138(5):1382–1393.

Chiquet, J., Grandvalet, Y., and Ambroise, C. (2011). Inferring multiple graphical structures. *Statistics and Computing*, 21(4):537–553.

Cook, R. D. and Forzani, L. (2008). Covariance reducing models: An alternative to spectral modelling of covariance matrices. *Biometrika*, 95(4):799–812.

Cook, R. D., Helland, I. S., and Su, Z. (2013). Envelopes and partial least squares regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(5):851–877.

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica*, 20(3):927–960.

Cook, R. D. and Zhang, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association*, 110(510):599–611.

Cook, R. D. and Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics*, 25(1):284–300.

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

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., and et al (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6):659–667.

Flury, B. K. (1987). Two generalizations of the common principal component model. *Biometrika*, 74(1):59–69.

Flury, B. N. (1984). Common principal components in k groups. *Journal of the American Statistical Association*, 79(388):892–898.

Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage*, 80:426–444.

Fox, M. D. and Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience*, 4.

Franks, A. and Hoff, P. (2016). Shared subspace models for multi-group covariance estimation. *arXiv preprint arXiv:1607.03045*.

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

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*, 98(1):1–15.

Huang, W., Gallivan, K. A., and Absil, P.-A. (2015). A broyden class of quasi-newton methods for riemannian optimization. *SIAM Journal on Optimization*, 25(3):1660–1685.

Kim, J., Wozniak, J. R., Mueller, B. A., Shen, X., and Pan, W. (2014). Comparison of statistical tests for group differences in brain functional networks. *NeuroImage*, 101:681–694.

Lee, W. and Liu, Y. (2015). Joint estimation of multiple precision matrices with common structures. *The Journal of Machine Learning Research*, 16(1):1035–1062.

Leng, C. and Tang, C. Y. (2012). Sparse matrix graphical models. *J. Amer. Statist. Assoc.*, 107(499):1187–1200.

Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519):1131–1146.

Liu, W. and Luo, X. (2015). Fast and adaptive sparse precision matrix estimation in high dimensions. *Journal of Multivariate Analysis*, 135:153–162.

Long, Z., Duan, X., Mantini, D., and Chen, H. (2016). Alteration of functional connectivity in autism spectrum disorder: effect of age and anatomical distance. *Scientific reports*, 6:26527.

Martin, S., Raim, A. M., Huang, W., and Adragni, K. P. (2016). Manifoldoptim: An r interface to the roptlib library for riemannian manifold optimization. *arXiv preprint arXiv:1612.03930*.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486):735–746.

Qiu, H., Han, F., Liu, H., and Caffo, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of Royal Statistical Society, Series B.*, 78:487–504.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.*, 5:935–980.

Rudie, J., Brown, J., Beck-Pancer, D., Hernandez, L., Dennis, E., Thompson, P., Bookheimer, S., and Dapretto, M. (2013). Altered functional and structural brain network organization in autism. *NeuroImage: Clinical*, 2:79 – 94.

Ryali, S., Chen, T., Supekar, K., and Menon, V. (2012). Estimation of functional connectivity in fmri data using stability selection-based sparse partial correlation with elastic net penalty. *NeuroImage*, 59(4):3852–3861.

Schott, J. R. (1999). Partial common principal component subspaces. *Biometrika*, 86(4):899–908.

Su, Z., Zhu, G., Chen, X., and Yang, Y. (2016). Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika*, 103(3):579–593.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.

Wang, Y., Kang, J., Kemmer, P. B., and Guo, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Frontiers in Neuroscience*, 10:1–17.

Xia, M., Wang, J., and He, Y. (2013). Brainnet viewer: A network visualization tool for human brain connectomics. *PLOS ONE*, 8:1–15.

Xia, Y. and Li, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics*, 73:780–791.

Yin, J. and Li, H. (2012). Model selection and estimation in the matrix normal graphical model. *Journal of Multivariate Analysis*, 107:119 – 140.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

Zhang, X., Mai, Q., et al. (2018). Model-free envelope dimension selection. *Electronic Journal of Statistics*, 12(2):2193–2216.

Zhao, S. D., Cai, T. T., and Li, H. (2014). Direct estimation of differential networks. *Biometrika*, 101(2):253–268.

Zhu, Y., Shen, X., and Pan, W. (2014). Structural Pursuit Over Multiple Undirected Graphs. *J. Amer. Statist. Assoc.*, 109(508):1683–1696.

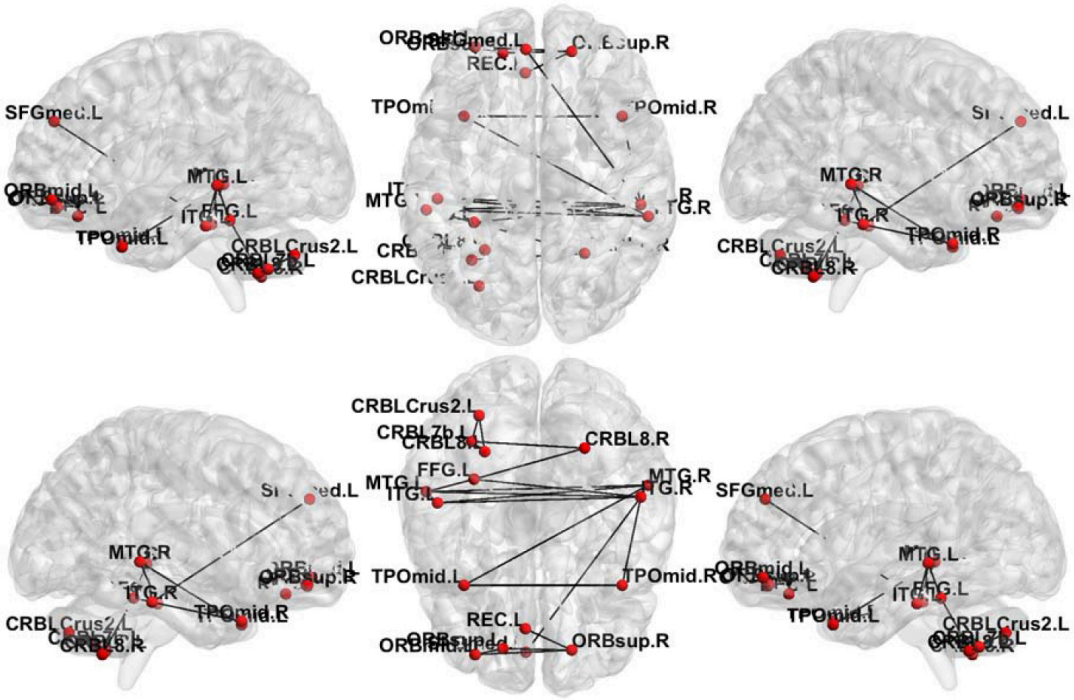

Figure 1. Visualization of the top 20 links selected by the envelope-based sparse precision matrix estimation and the associated brain regions.

Table 1

Covariance matrix estimation. Reported are the average estimation error $||\boldsymbol{\Delta}_1 - \boldsymbol{\Delta}_2 - (\widehat{\boldsymbol{\Delta}}_1 - \widehat{\boldsymbol{\Delta}}_2)||_F$ over 100 data replications. CRS: the proposed common reducing subspace model; FH16: the shared subspace model of Franks and Hoff (2016); CF08: the covariance reducing model of Cook and Forzani (2008); Sample: the sample mean estimator. The last column, max SE: the maximum standard error.

	$p = 100, n_k = 40$	$p = 100, n_k = 100$	$p = 200, n_k = 40$	$p = 200, n_k = 100$	max SE
			Model I		
CRS	1·469	0·913	2·388	1·387	0·061
FH16	109·1	86·07	475·6	271·1	33·15
CF08	92·77	80·73	444·9	266·9	34·58
Sample	131·9	98·77	609·4	359·1	31·74
			Model II		
CRS	0·004	0·002	0·004	0·003	0·000
FH16	0·142	0·107	0·207	0·155	0·004
CF08	0·155	0·106	0·198	0·117	0·016
Sample	2·141	1·352	4·380	2·765	0·002
			Model III		
CRS	0·413	0·263	0·471	0·293	0·006
FH16	0·413	0·215	0·446	0·255	0·008
CF08	0·823	0·450	1·251	0·618	0·081
Sample	0·539	0·341	0·751	0·471	0·004

Table 2

Precision matrix estimation. Reported are the average estimation error $||\mathbf{\Phi}_1 - \mathbf{\Phi}_2 - (\widehat{\mathbf{\Phi}}_1 - \widehat{\mathbf{\Phi}}_2)||_F$ over 100 data replications. CRS: the proposed common reducing subspace model; FH16: the shared subspace model of Franks and Hoff (2016); CF08: the covariance reducing model of Cook and Forzani (2008); Sample: the sample mean estimator. The last column, max SE: the maximum standard error.

	$p = 100, n_k = 40$	$p = 100, n_k = 100$	$p = 200, n_k = 40$	$p = 200, n_k = 100$	max SE
			Model I		
CRS	0·372	0·187	0·740	0·295	0·009
FH16	0·734	0·734	0·740	0·739	0·000
CF08	0·744	0·738	0·747	0·744	0·002
Sample	1·219	0·755	2·477	1·494	0·001
			Model II		
CRS	0·700	0·424	1·008	0·569	0·008
FH16	6·693	6·693	6·701	6·701	0·000
CF08	6·693	6·540	6·708	6·643	0·049
Sample	1·481	0·915	2·760	1·657	0·004
			Model III		
CRS	0·306	0·193	0·394	0·248	0·002
FH16	0·325	0·204	0·416	0·260	0·005
CF08	2·312	1·558	2·911	2·103	0·015
Sample	11·27	6·969	23·80	14·35	0·014

Table 3

Sparse precision matrix estimation. Reported are the average estimation error, the sensitivity (%) and specificity (%) of nonzero elements selection, over 100 data replications. CRS: the proposed common reducing subspace model; FH16: the shared subspace model of Franks and Hoff (2016); CF08: the covariance reducing model of Cook and Forzani (2008); DWW14: the fused graphical Lasso estimator of Danaher et al. (2014); LL15: the joint prcesion estimator of Lee and Liu (2015); Sample: the sample mean estimator. For CRS, FH16, CF08 and Sample estimators, we further combined with a fast sparse precision matrix estimation method (SCIO; Liu and Luo, 2015) to achieve variable selection and to encourage sparsity. The last column, max SE: the maximum standard error. The left panel reports the estimation error. The right panel reports the sensitivity (the first number) and specificity (the second number).

	Estimation error					Selection specificity and sensitivity			
	$p = 100$ $n_k = 40$	$p = 100$ $n_k = 100$	$p = 200$ $n_k = 40$	$p = 200$ $n_k = 100$	max SE	$p = 100$ $n_k = 40$	$p = 100$ $n_k = 100$	$p = 200$ $n_k = 40$	$p = 200$ $n_k = 100$
			Model I					Model I	
CRS	0·246	0·143	0·500	0·246	0·005	83·3, 74·3	90·7, 72·8	54·7, 83·2	77·9, 77·5
FH16	0·734	0·734	0·738	0·738	0·000	3·99, 98·3	4·57, 97·8	2·14, 99·1	1·93, 99·3
CF08	0·618	0·611	0·640	0·622	0·002	80·4, 70·3	89·8, 62·5	50·4, 83·7	73·8, 74·6
DWW14	0.701	0.509	0.671	0.566	0.009	65.6, 77.3	88.4, 58.2	39.4, 87.4	78.2, 60.9
LL15	0.654	0.677	0.830	0.733	0.002	43.0, 92.1	24.2, 99.5	35.6, 91.2	16.9, 99.3
Sample	0·444	0·315	0·651	0·501	0·002	77·4, 70·5	88·2, 62·9	36·9, 89·4	69·9, 74·4
			Model II					Model II	
CRS	0·654	0·435	1·342	0·808	0·011	96·2, 63·4	97·6, 62·2	93·0, 67·3	96·1, 64·2
FH16	6·695	6·694	6·702	6·702	0·000	6·71, 95·3	6·54, 94·9	2·73, 98·5	2·82, 98·4
CF08	5·827	5·678	5·730	5·753	0·083	96·0, 56.3	98·4, 35·9	87·8, 59·5	95·5, 50·0
DWW14	3.785	1.918	5.267	2.963	0.022	94.7,42.1	98.1, 24.5	96.1, 23.3	97.9, 21.4
LL15	3.134	3.123	5.393	5.582	0.004	86.7, 91.7	86.0, 99.5	72.9, 90.8	64.1, 99.3
Sample	1·045	0·670	2·216	1·382	0·010	95·0, 51.8	97·1, 48·8	89·7, 59·1	94·8, 53·1
			Model III					Model III	
CRS	0·599	0·376	1·324	0·724	0·019	85·3, 73·5	91·0, 70·3	71·7, 79.0	81·4, 75·5
FH16	0·290	0·176	0·349	0·212	0·008	97·4, 79·7	98·1, 82·5	95·5, 84·0	97·1, 84·6
CF08	2·103	1·483	3·154	2·139	0·048	88·0, 66·6	93·1, 63·8	73·7, 73·5	83·9, 69·8
DWW14	1.646	1.377	2.342	2.244	0.012	60.7, 99.2	80.8, 97.6	46.2, 98.6	66.7, 98.6
LL15	2.869	1.880	4.849	2.558	0.001	39.4, 91.2	26.3, 99.4	26.6, 90.4	12.0, 99.2
Sample	4·096	2·822	6·613	4·896	0·011	85·4, 63·1	91·4, 57·1	70·3, 74·1	81·3, 66·1

Table 4

Envelope dimension selection. Reported are the percentage of times that the true dimension is selected out of 100 data replications.

	$p = 100$			$p = 200$		
	$n_k = 100$	$n_k = 200$	$n_k = 400$	$n_k = 200$	$n_k = 400$	$n_k = 600$
Model I	86	99	100	79	99	100
Model II	100	100	100	100	100	100
Model III	100	100	100	100	100	100

Table 5

The top 20 links selected by the envelope-based sparse precision matrix estimation and the associated brain regions.

Connected Regions	Difference	Connected Regions	Difference
Cerebellum_7b_L ↔ Cerebellum_8_L	0·67	Cerebellum_Crus2_L ↔ Cerebellum_8_L	-0·17
Temporal_Inf_L ↔ Temporal_Inf_R	0·36	Temporal_Mid_L ↔ Temporal_Inf_R	-0·15
Temporal_Pole_Mid_L ↔ Temporal_Pole_Mid_R	0·29	Cerebellum_7b_L ↔ Cerebellum_8_R	-0·15
Frontal_Sup_Orb_L ↔ Frontal_Sup_Orb_R	0·27	Fusiform_L ↔ Temporal_Mid_L	-0·14
Frontal_Sup_Orb_L ↔ Frontal_Mid_Orb_L	0·21	Frontal_Sup_Orb_R ↔ Rectus_L	-0·13
Temporal_Mid_L ↔ Temporal_Inf_L	0·19	Temporal_Mid_R ↔ Temporal_Pole_Mid_L	-0·13
Temporal_Pole_Mid_R ↔ Temporal_Inf_R	0·18	Fusiform_L ↔ Cerebelum_8_R	-0·12
Temporal_Mid_L ↔ Temporal_Mid_R	0·16	Frontal_Sup_Medial_L ↔ Temporal_Inf_R	-0·11
Cerebellum_Crus2_L ↔ Cerebellum_7b_L	0·15	Frontal_Sup_Orb_R ↔ Frontal_Mid_Orb_L	-0·10
Fusiform_L ↔ Temporal_Inf_R	0·14	Temporal_Mid_R ↔ Temporal_Inf_L	-0·09