Global Testing and Large-Scale Multiple Testing for High-Dimensional Covariance Structures

T. Tony Cai[1]

[1]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, USA, 19104; email: tcai@wharton.upenn.edu

Xxxx. Xxx. Xxx. Xxx. YYYY. AA:1–27

This article's doi:
10.1146/((please add article doi))

Keywords

Correlation matrix, covariance matrix, differential network, false discovery proportion, false discovery rate, Gaussian graphical model, global testing, multiple testing, null distribution, precision matrix, thresholding.

Abstract

Driven by a wide range of contemporary applications, statistical inference for covariance structures has been an active area of current research in high-dimensional statistics. This paper provides a selective survey of some recent developments in hypothesis testing for high-dimensional covariance structures, including global testing for the overall pattern of the covariance structures and simultaneous testing of a large collection of hypotheses on the local covariance structures with false discovery proportion (FDP) and false discovery rate (FDR) control. Both one-sample and two-sample settings are considered. The specific testing problems discussed include global testing for the covariance, correlation, and precision matrices, and multiple testing for the correlations, Gaussian graphical models, and differential networks.

Contents

1. Introduction

Analysis of high dimensional data, whose dimension p can be much larger than the sample size n, has emerged as one of the most important and active areas of current research in statistics. It has a wide range of applications in many fields, including genomics, medical imaging, signal processing, social science, and financial economics. Covariance structure plays a central role in many of these analyses. In addition to being of significant interest in its own right for many important applications, knowledge of the covariance structures is critical for a large collection of fundamental statistical methods, including the principal component analysis, discriminant analysis, clustering analysis, and regression analysis.

There have been significant recent advances on statistical inference for high-dimensional covariance structures. These include estimation of and testing for large covariance matrices, volatility matrices, correlation matrices, precision matrices, Gaussian graphical models, differential correlations, and differential networks. The difficulties of these high-dimensional statistical problems can be traced back to the fact that the usual sample covariance matrix, which performs well in the classical low-dimensional cases, fails to produce good results in the high-dimensional settings. For example, when $p > n$, the sample covariance matrix is not invertible, and consequently does not lead directly to an estimate of the precision matrix (the inverse of the covariance matrix) which is required in many applications. It is also well known that the empirical principal components of the sample covariance matrix can be nearly orthogonal to the corresponding principal components of the population covariance matrix in the high-dimensional settings. See, e.g., Johnstone (2001), Paul (2007), and Johnstone and Lu (2009).

Structural assumptions are needed for accurate estimation of high-dimensional covariance and precision matrices. A collection of smoothing and regularization methods have been developed for estimation of covariance, correlation, and precision matrices under the various structural assumptions. These include the banding methods in Wu and Pourahmadi (2009) and Bickel and Levina (2008a), tapering in Furrer and Bengtsson (2007) and Cai et al. (2010), thresholding rules in Bickel and Levina (2008b), El Karoui (2008), Rothman et al. (2009), Cai and Liu (2011), and Fan et al. (2011), penalized log-likelihood estimation in Huang et al. (2006), Yuan and Lin (2007), d'Aspremont et al. (2008), Banerjee et al. (2008), Rothman et al. (2008), Lam and Fan (2009), Ravikumar et al. (2011), and Chandrasekaran et al. (2012), regularizing principal components in Johnstone and Lu (2009), Zou et al. (2006), Cai et al. (2013, 2015), Fan et al. (2013), and Vu and Lei (2013), and penalized regression for precision matrix estimation, Gaussian graphical models, and differential networks in Meinshausen and Bühlmann (2006), Yuan (2010), Cai et al. (2011), Sun and Zhang (2013), Zhao et al. (2014), and Ren et al. (2015).

Optimal rates of convergence and adaptive estimation for high-dimensional covariance and precision matrices have been studied. See, for example, Cai et al. (2010); Cai and Yuan (2012) for bandable covariance matrices, Cai and Zhou (2012); Cai and Liu (2011) for sparse covariance matrices, Cai, Ren, and Zhou (2013) for Toeplitz covariance matrices, Cai et al. (2015) for sparse spiked covariance matrices, and Cai et al. (2016) for sparse precision matrices. It should be noted that many of these matrix estimation problems exhibit new features that are significantly different from those that occur in the conventional nonparametric function estimation problems and require the development of new technical tools which can be useful for other estimation and inference problems as well. See Cai et al. (2016) for a survey of the recent results on optimal and adaptive estimation, and for a discussion of some key technical tools used in the theoretical analyses.

Parallel to estimation, there have been significant recent developments on the methods and theory for hypothesis testing on high-dimensional covariance structures. There are a wide array of applications, including brain connectivity analysis (Shaw et al. 2006), gene co-expression network analysis (Carter et al. 2004; Lee et al. 2004; Zhang et al. 2008; Dubois et al. 2010; Fuller et al. 2007), differential gene expression analysis (Shedden and Taylor (2005); de la Fuente (2010); Ideker and Krogan (2012); Fukushima (2013)), and correlation analysis of factors that interact to shape children's language development and reading ability (Dubois et al. 2010; Hirai et al. 2007; Raizada et al. 2008; Zhu et al. 2005).

High dimensionality and dependency impose significant methodological and technical challenges. This is particularly true for testing on the structure of precision matrices and differential networks as there is no "sample precision matrix" as the natural starting point. The problems of hypothesis testing for high-dimensional covariance structures can be divided into two types: global testing for the overall pattern of the covariance structures and simultaneous testing of a large collection of hypotheses for the local covariance structures such as pairwise correlations or changes on conditional dependence. These two types of testing problems are related but also very different in nature. We discuss them separately in the present paper.

In the classical setting of low-dimension and large sample size, several methods have been developed for testing specific global patterns of covariance matrices. In the high-dimensional settings, these methods either do not perform well or are no longer applicable. A collection of new global testing procedures have been developed in the last few years both in the one-sample and two-sample cases. In addition to global testing, large-scale multiple testing for high-dimensional covariance structures, where one tests simultaneously thousands or millions of hypotheses on the local covariance structures or the changes of the local covariance structures, is an important and challenging problem. For a range of applications, a common feature in the covariance structures is sparsity under the alternatives. We will discuss a few recently proposed methods for both global testing and large-scale multiple testing for high-dimensional covariance structures that leverage the sparsity information. We will conclude with a discussion on some future issues and open problems.

Throughout the paper, $\mathbf{\Sigma} = (\sigma_{i,j})$, $\mathbf{R} = (r_{i,j})$, $\mathbf{\Omega} = (\omega_{i,j})$, and $\mathbf{I}$ denote the covariance, correlation, precision, and identity matrices, respectively. in the one-sample case, we assume that we observe a random sample $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ which consists of n independent copies of a p-dimensional random vector $\mathbf{X} = (X_1, \dots, X_p)^\intercal$ following some distribution with mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)$ and covariance matrix $\mathbf{\Sigma} = (\sigma_{i,j})_{p\times p}$. The sample mean is $\bar{\mathbf{X}} = \frac{1}{n}\sum_{k=1}^n \mathbf{X}_k$ and sample covariance matrix $\hat{\mathbf{\Sigma}} = (\hat{\sigma}_{i,j})_{p\times p}$ is defined as

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^\intercal. \tag{1}$$

In the two-sample case, we observe two independent random samples, $\{\mathbf{X}_1, \dots, \mathbf{X}_{n_1}\}$ are i.i.d. from a p-variate distribution with $\boldsymbol{\mu}_1 = (\mu_{1,1}, ..., \mu_{p,1})^\intercal$ and covariance matrix $\mathbf{\Sigma}_1 = (\sigma_{i,j,1})_{p\times p}$ and $\{\mathbf{Y}_1, \dots, \mathbf{Y}_{n_2}\}$ are i.i.d. from a distribution with $\boldsymbol{\mu}_2 = (\mu_{1,2}, ..., \mu_{p,2})^\intercal$ and covariance matrix $\mathbf{\Sigma}_2 = (\sigma_{i,j,2})_{p\times p}$. Given the two independent random samples, define the sample means by $\bar{\mathbf{X}} = \frac{1}{n_1}\sum_{k=1}^{n_1} \mathbf{X}_k$ and $\bar{\mathbf{Y}} = \frac{1}{n_2}\sum_{k=1}^{n_2} \mathbf{Y}_k$ and the sample covariance matrices $\hat{\mathbf{\Sigma}}_1 = (\hat{\sigma}_{i,j,1})_{p\times p}$ and $\hat{\mathbf{\Sigma}}_2 = (\hat{\sigma}_{i,j,2})_{p\times p}$ by

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{n_1}\sum_{k=1}^{n_1}(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^\intercal \quad \text{and} \quad \hat{\mathbf{\Sigma}}_2 = \frac{1}{n_2}\sum_{k=1}^{n_2}(\mathbf{Y}_k - \bar{\mathbf{Y}})(\mathbf{Y}_k - \bar{\mathbf{Y}})^\intercal. \tag{2}$$

1.1. Global Testing for High-Dimensional Covariance Structures

Global testing has been relatively well studied in the one-sample case. Several interesting two-sample global testing problems have also been investigated recently. We give a brief discussion here and consider these problems in detail in Section 2.

1.1.0.1. Testing the Simple Global Null Hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{I}$. Testing a simple global null hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{I}$ occupies a special position in the methods and theory for testing high-dimensional covariance structures. Much of the random matrix theory has been developed in the null case $\mathbf{\Sigma} = \mathbf{I}$ and testing for $H_0 : \mathbf{\Sigma} = \mathbf{I}$ is often used as an immediate application for the results from the random matrix theory. A collection of testing methods has been proposed in the literature, including tests based on the spectral norm (largest eigenvalue), tests based on the Frobenius norm, and tests using the maximum entrywise deviation. A concise summary of the methods for testing the global null $H_0 : \mathbf{\Sigma} = \mathbf{I}$ is given in Section 2.1.

1.1.0.2. Testing a Composite Global Null Hypothesis in the One-Sample Case. In addition to testing the simple null $H_0 : \mathbf{\Sigma} = \mathbf{I}$, it is also of interest to test a range of composite global null hypotheses, including testing for sphericity with $H_0 : \mathbf{\Sigma} = \sigma^2\mathbf{I}$, testing for $H_0 : \mathbf{\Sigma}$ is diagonal, which is equivalent to $H_0 : \mathbf{R} = \mathbf{I}$, and testing for short-range dependence $H_0 : r_{i,j} = 0$ for all $|i - j| \geq k$. For testing against a composite null, the standard random theory results are often not directly applicable. Methods for testing these one-sample composite global nulls will be discussed in Section 2.2.

1.1.0.3. Testing for Equality of Two Covariance Matrices $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. Testing the equality of two covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ is an important problem. Many statistical procedures including the classical Fisher's linear discriminant analysis rely on the fundamental assumption of equal covariance matrices. The likelihood ratio test (LRT) is commonly used in the conventional low-dimensional case and enjoys certain optimality properties. In the high-dimensional setting where $p > n$, the LRT is not applicable. There are two types of tests in this setting: one is based on the maximum entrywise deviation and another is based on the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$. The former is particularly powerful when $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is "sparse" under the alternative and the latter performs well when $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is "dense". Section 2.3 discusses a recent proposal in Cai et al. (2013) that is powerful against sparse alternatives and robust with respect to the population distributions. A summary of a few Frobenius norm based methods is also given.

1.1.0.4. Global Testing for Differential Correlations $H_0 : \mathbf{R}_1 - \mathbf{R}_2 = 0$. Differential gene expression analysis is widely used in genomics for identifying disease-associated genes for complex diseases (Shedden and Taylor (2005); Bandyopadhyay et al. (2010); de la Fuente (2010); Ideker and Krogan (2012); Fukushima (2013)). For example, de la Fuente (2010) demonstrated that the gene expression networks can vary in different disease states and the differential correlations in gene expression are useful in disease studies. Global testing for the differential correlations $H_0 : \mathbf{R}_1 - \mathbf{R}_2 = 0$ based on two independent random samples is complicated as the entries of the sample differential correlation matrix are correlated under the null H_0. Section 2.4 considers a test recently introduced in Cai and Zhang (2016) that is based on the maximum entrywise deviation. The test is powerful when the differential correlation matrix is sparse under the alternative.

1.1.0.5. Global Testing for Differential Network $H_0 : \mathbf{\Omega}_1 - \mathbf{\Omega}_2 = 0$. Gaussian graphical models provide a powerful tool for modeling the conditional dependence relationships among a large number of random variables in a complex system. Applications include portfolio optimization, speech recognition, and genomics. It is well known that the structure of an undirected Gaussian graph is characterized by the precision matrix of the corresponding distribution (Lauritzen 1996). In addition to inference for a given Gaussian graph, it is of significant interest in many applications to understand whether and how the network changes between disease states (de la Fuente 2010; Ideker and Krogan 2012; Hudson et al. 2009; Li et al. 2007; Danaher et al. 2014; Zhao et al. 2014; Xia et al. 2015).

The problem of detecting whether the network changes between disease states under the Gaussian graphical model framework can be formulated as testing

$$H_0 : \ \mathbf{\Delta} = 0 \quad \text{versus} \quad H_1 : \ \mathbf{\Delta} \neq 0$$

where $\mathbf{\Delta} = (\delta_{i,j}) = \mathbf{\Omega}_1 - \mathbf{\Omega}_2$ is called the differential network. Equality of two precision matrices is mathematically equivalent to equality of two covariance matrices. However, it is often the case in many applications $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are sparse and it is important to test the precision matrices directly in order to incorporate the sparsity information in the testing procedure. Section 2.5 discusses a testing method recently proposed in Xia et al. (2015) that is based on a penalized regression approach.

1.2. Multiple Testing for Local Covariance Structures

Simultaneously testing a large number of hypotheses on local covariance structures arises in many important applications including genome-wide association studies (GWAS), phenome-wide association studies (PheWAS), and brain imaging. A common goal in multiple testing is often to control the false discovery proportion (FDP), which is defined to be the proportion of false positives among all rejections, and its expectation, the false discovery rate (FDR). Simultaneous testing for the covariance structures with FDP and FDR control is technically challenging, both in constructing a suitable test statistic for testing a given hypothesis and establishing its null distribution and in developing a multiple testing procedure to account for the multiplicity and dependency so that the overall FDP and FDR are controlled.

Multiple testing has been well studied in the literature, especially in the case where the test statistics are independent. See, for example, Benjamini and Hochberg (1995); Storey (2002); Efron (2004); Genovese and Wasserman (2004); Sun and Cai (2007). Standard FDR control procedures, such as the Benjamini-Hochberg (B-H) procedure (Benjamini and Hochberg 1995), typically built under the independence assumption, would fail to provide desired error controls in the presence of strong correlation, particularly when the signals are sparse. The effects of dependency on multiple testing procedures have been considered, for example, in Benjamini and Yekutieli (2001); Storey et al. (2004); Qiu et al. (2005); Farcomeni (2007); Wu (2008); Efron (2007); Sun and Cai (2009); Sun et al. (2015). In particular, Qiu et al. (2005) demonstrated that the dependency effects can significantly deteriorate the performance of many multiple testing procedures. Farcomeni (2007) and Wu (2008) showed that the FDR is controlled at the nominal level by the B-H procedure under some stringent dependency assumptions. The procedure in Benjamini and Yekutieli (2001) allows the general dependency by paying a logarithmic term loss on the FDR which makes the method very conservative.

A natural starting point for large-scale multiple testing for local covariance structures is the sample covariance or correlation matrix, whose entries are intrinsically heteroscedastic and dependent even if the original observations are independent. The sample covariances/correlations may have a wide range of variability and their dependence structure is rather complicated. The analysis of these large-scale multiple testing problems poses many statistical challenges not present in smaller scale studies.

1.2.0.1. Multiple Testing for Correlations. The problem of correlation detection arises in both one-sample and two-sample settings. In the one-sample case, one wishes to simultaneously test $(p^2 - p)/2$ hypotheses

$$H_{0,i,j}: \ r_{i,j} = 0 \quad \text{versus} \quad H_{1,i,j}: \ r_{i,j} \neq 0, \quad \text{for} \quad 1 \leq i < j \leq p.$$

In the two-sample case, one is interested in the simultaneous testing of correlation changes,

$$H_{0,i,j}: \quad r_{i,j,1} = r_{i,j,2} \quad \text{versus} \quad H_{1,i,j}: \quad r_{i,j,1} \neq r_{i,j,2}, \quad \text{for} \quad 1 \leq i < j \leq p.$$

Section 3.2 discusses multiple testing procedures for correlations recently introduced in Cai and Liu (2015) in both the one-sample and two-sample settings.

1.2.0.2. Multiple Testing for Gaussian Graphical Models. As mentioned earlier, Gaussian graphical models have a wide range of applications and the conditional dependence structure is completely characterized by the corresponding precision matrix $\mathbf{\Omega} = (\omega_{i,j})_{p\times p}$ of the distributions. Given an i.i.d. random sample, the problem of multiple testing for a Gaussian graphical model can thus be formulated as simultaneously testing $(p^2 - p)/2$ hypotheses on the off-diagonal entries of the precision matrix $\mathbf{\Omega}$,

$$H_{0,i,j}: \omega_{i,j} = 0 \quad \text{versus} \quad H_{1,i,j}: \omega_{i,j} \neq 0, \quad 1 \leq i < j \leq p.$$

Section 3.3 discusses a multiple testing procedure for Gaussian graphical models proposed in Liu (2013), which was shown to control both the FDP and FDR asymptotically.

1.2.0.3. Multiple Testing for Differential Networks. Differential network analysis has many important applications in genomics and other fields. If the global null hypothesis $H_0: \mathbf{\Delta} \equiv \mathbf{\Omega}_1 - \mathbf{\Omega}_2 = 0$ is rejected, it is often of significant interest to investigate the local structures of the differential network $\mathbf{\Delta}$. A natural approach is to carry out simultaneous testing for the $(p^2 - p)/2$ hypotheses on the off-diagonal entries of $\mathbf{\Delta} = (\delta_{i,j})$,

$$H_{0,i,j}: \delta_{i,j} = 0 \quad \text{versus} \quad H_{1,i,j}: \delta_{i,j} \neq 0, \quad 1 \leq i < j \leq p. \tag{3}$$

This problem was recently studied in Xia et al. (2015). A procedure for simultaneously testing the hypotheses in (3) was developed and shown to control the FDP and FDR asymptotically. Section 3.4 will discuss this procedure in detail.

1.3. Notation and Definitions

We introduce here notation and definitions that will be used in the rest of the paper. Vectors and matrices are denoted by boldfaced letters.

For a vector $\mathbf{a} = (a_1, \ldots, a_p)^\intercal \in \mathbb{R}^p$, define the ℓ_q norm by $\|\mathbf{a}\|_q = (\sum_{i=1}^p |a_i|^q)^{1/q}$ for $1 \leq q \leq \infty$. A vector $\mathbf{a} \in \mathbb{R}^p$ is called k-sparse if it has at most k nonzero entries. We say a matrix $\mathbf{A}$ is k-sparse if each row/column has at most k nonzero entries. For any vector $\mathbf{a} \in \mathbb{R}^p$, let $\mathbf{a}_{-i}$ denote the vector in $\mathbb{R}^{p-1}$ by removing the i^{th} entry from $\mathbf{a}$. For a matrix $\mathbf{A} \in \mathbb{R}^{p\times q}$, $\mathbf{A}_{i,-j}$ denotes the i^{th} row of $\mathbf{A}$ with its j^{th} entry removed and $\mathbf{A}_{-i,j}$ denotes the j^{th} column of $\mathbf{A}$ with its i^{th} entry removed. The matrix $\mathbf{A}_{-i,-j}$ denotes a $(p-1)\times(q-1)$ matrix obtained by removing the i^{th} row and j^{th} column of $\mathbf{A}$. For $1 \le w \le \infty$ and a matrix $\mathbf{A} = (a_{i,j})_{p\times p}$, the matrix ℓ_w norm of a matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_w = \max_{\|x\|_w=1} \|\mathbf{A}x\|_w$. The spectral norm is the ℓ_2 norm. The Frobenius norm of $\mathbf{A}$ is $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{i,j}^2}$, and the trace of $\mathbf{A}$ is $\mathrm{tr}(\mathbf{A}) = \sum_i a_{i,i}$.

For a set $\mathcal{H}$, denote by $|\mathcal{H}|$ the cardinality of $\mathcal{H}$. For two sequences of positive real numbers $\{a_n\}$ and $\{b_n\}$, write $a_n = O(b_n)$ if there exists a constant C such that $|a_n| \le C|b_n|$ holds for all n, write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$, $a_n \lesssim b_n$ means $a_n \le Cb_n$ for all n, and $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. We use $I\{\cdot\}$ to denote the indicator function, use ϕ and Φ to denote respectively the density and cumulative distribution function of the standard normal distribution, and use $G(t) = 2 - 2\Phi(t)$ for the tail probability of the standard normal distribution.

2. Global Testing for Covariance Structures

We discuss in this section several problems on global testing for the overall patterns of the covariance structures. As mentioned in the introduction, in many applications, a common feature in the covariance structures is sparsity under the alternatives. A natural strategy to leverage the sparsity information is to use the maximum of entrywise deviations from the null as the test statistic. The null distribution of such a statistic is often the type I extreme value distribution. In this paper, we say a test statistic $M_{n,p}$ has asymptotically the extreme value distribution of type I if for any given $t \in \mathbb{R}$,

$$\mathbb{P}\left(M_{n,p} - 4\log p + \log\log p \le t\right) \to \exp\left(-\frac{1}{\sqrt{8\pi}} e^{-t/2}\right), \quad \text{as } n, p \to \infty. \tag{4}$$

The $1-\alpha$ quantile, denoted by q_α, of the type I extreme value distribution is given by

$$q_\alpha = -\log(8\pi) - 2\log\log(1-\alpha)^{-1}. \tag{5}$$

We begin with a brief summary of the methods for testing for a simple null $H_0 : \mathbf{\Sigma} = \mathbf{I}$. This problem has been well studied in the literature due to its close connection to the random matrix theory.

2.1. Testing a Simple Global Null Hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{I}$

In the classical setting where the dimension p is fixed and the distribution is Gaussian, the most commonly used test for testing the simple null $H_0 : \mathbf{\Sigma} = \mathbf{I}$ is the likelihood ratio test, whose test statistic is given by $LR = \mathrm{tr}(\hat{\mathbf{\Sigma}}) - \log\det(\hat{\mathbf{\Sigma}}) - p$. It is known that the asymptotic null distribution of the test statistic LR is $\chi^2_{p(p+1)/2}$. See Anderson (2003); Muirhead (1982). When the dimension p grows with the sample size n, the standard LRT is no longer applicable. Bai et al. (2009) proposed a corrected LRT for which the test statistic has a Gaussian limiting null distribution when $p/n \to c \in (0, 1)$. The result is further extended in Jiang et al. (2012); Zheng et al. (2015).

In the high-dimensional setting, a natural approach to testing the simple null $H_0 : \mathbf{\Sigma} = \mathbf{I}$ is to use test statistics that are based on some distance $\|\mathbf{\Sigma} - \mathbf{I}\|$ where $\|\cdot\|$ is a matrix norm such as the spectral norm or the Frobenius norm. When the dimension p and the sample size n are comparable, i.e., $p/n \to \gamma \in (0, \infty)$, testing of the hypotheses $H_0 : \mathbf{\Sigma} = \mathbf{I}$ against $H_1 : \mathbf{\Sigma} \neq \mathbf{I}$ has been considered by Johnstone (2001) in the Gaussian case, Soshnikov (2002) in the sub-Gaussian case, and by Peche (2009) in a more general setting with moment conditions and where the ratio p/n can converge to either a positive number γ, 0 or ∞. Johnstone (2001) showed that in the Gaussian case, when $p/n \to c \in (0, \infty)$, the limiting null distribution of the largest eigenvalue of the sample covariance matrix is the Tracy-Widom distribution. The result immediately yields a test for $H_0 : \mathbf{\Sigma} = \mathbf{I}$ in the Gaussian case. See Johnstone (2001); Soshnikov (2002); Peche (2009) for further details.

In addition to the spectral norm (largest eigenvalue) based tests, several testing procedures based on the squared Frobenius norm $\|\mathbf{\Sigma} - \mathbf{I}\|_F^2$ have been proposed. Tests using the Frobenius norm was first introduced in John (1971) and Nagao (1973) in the low-dimensional setting by simply plugging in the sample covariance matrix $\hat{\mathbf{\Sigma}}$ and using $p^{-1}\|\hat{\mathbf{\Sigma}} - \mathbf{I}\|_F^2$ as the test statistic. Ledoit and Wolf (2002) showed that using $p^{-1}\|\hat{\mathbf{\Sigma}} - \mathbf{I}\|_F^2$ leads to an inconsistent test in the high-dimensional setting and introduced a correction term $-\frac{1}{np}\mathrm{tr}^2(\hat{\mathbf{\Sigma}}) + \frac{p}{n}$. It is shown that the new test statistic has a Gaussian limiting null distribution. The results have been extended by Birke and Dette (2005); Srivastava (2005). Chen et al. (2010) proposed new test statistic by using U-statistics to estimate the traces $(\mathrm{tr}(\mathbf{\Sigma}), \mathrm{tr}(\mathbf{\Sigma}^2))$ and studied the asymptotic power of the test.

A few optimality results on testing $H_0 : \mathbf{\Sigma} = \mathbf{I}$ have been established. Cai and Ma (2013) characterized the optimal testing boundary that separates the testable region from the non-testable region by the Frobenius norm in the setting where p/n is bounded. A U-statistics based test that is similar to the one proposed in Chen et al. (2010) is shown to be rate optimal over this asymptotic regime. Moreover, the power of this test uniformly dominates that of the corrected LRT over the entire asymptotic regime under which the corrected LRT is applicable (i.e. $p < n$ and $p/n \to c \in [0, 1]$). See also Baik et al. (2005); El Karoui (2007); Onatski et al. (2013) for other optimality results.

2.2. Testing a Composite Global Null Hypothesis in the One-Sample Case

In addition to testing the simple null hypothesis $H_0 : \mathbf{\Sigma} = \mathbf{I}$, there are several interesting global testing problems in the one-sample case with composite null hypotheses. Tests based on the largest eigenvalue of the sample covariance matrix cannot be easily modified for testing composite nulls. In particular, testing for sphericity with $H_0 : \mathbf{\Sigma} = \sigma^2\mathbf{I}$ has been studied. LRT is the standard choice in the low-dimensional setting. John (1971) introduced an invariant test based on the test statistic $\frac{1}{p}\mathrm{tr}\left\{(\frac{\hat{\mathbf{\Sigma}}}{p^{-1}\mathrm{tr}(\hat{\mathbf{\Sigma}})} - \mathbf{I})^2\right\}$ and proved that the test is locally most powerful invariant test for sphericity where the dimension p is fixed. Ledoit and Wolf (2002) showed this test is consistent even when p grows with n.

A more general null hypothesis is that $\mathbf{\Sigma}$ is diagonal. In the Gaussian case, this is equivalent to the independence of the components. The null is equivalent to $H_0 : \mathbf{R} = \mathbf{I}$, where $\mathbf{R}$ is the correlation matrix. A natural test is then to use the maximum of the absolute off-diagonal entries of the sample correlation matrix $\hat{\mathbf{R}} = (\hat{r}_{i,j})$. Let $L_{n,p} = \max_{1 \le i < j \le p} |\hat{r}_{i,j}|$. $L_{n,p}$ is called the coherence of the data matrix $(\mathbf{X}_1, ..., \mathbf{X}_n)$ (Cai and Jiang 2011). Jiang (2004) showed that $nL_{n,p}^2$ has asymptotically the extreme value distribution of type I as in (4), under the conditions $\mathbb{E}|X_i|^{30+\varepsilon} < \infty$ and $p/n \to \gamma \in (0, \infty)$. Cai and Jiang (2011, 2012); Shao and Zhou (2014) extended the results on the limiting null distribution to the ultra-high-dimensional settings. The asymptotic null distribution results lead immediately to a test for the hypothesis $H_0 : \mathbf{R} = \mathbf{I}$ by using $nL_{n,p}^2$ as the test statistic.

Testing $H_0 : \mathbf{R} = \mathbf{I}$ can be viewed as a special case of testing for short-range dependence. More precisely, for a given bandwidth $k = k(n,p) \geq 1$, one wishes to test H_0: $\sigma_{i,j} = 0$ for all $|i-j| \geq k$. Such a problem arises, for example, in econometrics and in time series analysis. See, for example, Andrews (1991); Ligeralde and Brown (1995). Cai and Jiang (2011) proposed the following test statistic, called k-coherence,

$$L_{n,p,k} = \max_{|i-j|\geq k} |\hat{\rho}_{i,j}|$$

and showed that $nL_{n,p,k}^2$ has the same limiting extreme value distribution as in (4) provided that most correlations are bounded away from 1 and $k = o(p^\epsilon)$ for some small ϵ. The results have since been extended by Shao and Zhou (2014); Xiao and Wu (2011). Qiu and Chen (2012) proposed a test based on a U-statistic which is an unbiased estimator of $\sum_{|i-j|\geq k} \sigma_{i,j}^2$ for testing bandedness of $\mathbf{\Sigma}$. Compared with the test proposed in Cai and Jiang (2011), which is powerful for sparse alternatives, the test introduced in Qiu and Chen (2012) is powerful for dense alternatives.

2.3. Testing for Equality of Two Covariance Matrices $H_0 : \mathbf{\Sigma_1} = \mathbf{\Sigma_2}$

In the classical low-dimensional setting, the problem of testing for the equality of two covariance matrices has been well studied. The likelihood ratio test (LRT) is most commonly used and it has certain optimality properties. See, for example, Sugiura and Nagao (1968); Gupta and Giri (1973); Perlman (1980); Gupta and Tang (1984); Anderson (2003).

In the high dimensional setting, where the dimension p can be much larger than the sample sizes, the LRT is not well defined. In such a setting, several new tests for testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ have been proposed. These are either based on the squared Frobenius norm $\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_F^2$ or the entrywise deviations. In many applications such as gene selection in genomics, the covariance matrices of the two populations can be either equal or quite similar in the sense that they only possibly differ in a small number of entries. In such a setting, under the alternative the differential covariance matrix $\mathbf{\Delta} \equiv \mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is sparse.

In a recent paper, Cai et al. (2013) introduced a test that is powerful against sparse alternatives and robust with respect to the population distributions. The null hypothesis $H_0 : \mathbf{\Sigma}_1 - \mathbf{\Sigma}_2 = 0$ is equivalent to $H_0 : \max_{1\leq i\leq j\leq p} |\sigma_{i,j,1} - \sigma_{i,j,2}| = 0$. Let the sample covariance matrices $\hat{\mathbf{\Sigma}}_1 = (\hat{\sigma}_{i,j,1})$ and $\hat{\mathbf{\Sigma}}_2 = (\hat{\sigma}_{i,j,2})$ be given in (2). A natural approach is to compare $\hat{\sigma}_{i,j,1}$ and $\hat{\sigma}_{i,j,2}$ and to base the test on the maximum differences. However, due to the heteroscedasticity of the sample covariances $\hat{\sigma}_{i,j,1}$'s and $\hat{\sigma}_{i,j,2}$'s, it is necessary to first standardize $\hat{\sigma}_{i,j,1} - \hat{\sigma}_{i,j,2}$ before making a comparison among different entries.

To be more specific, define the variances $\theta_{i,j,1} = \mathrm{Var}((X_i - \mu_{i,1})(X_j - \mu_{j,1}))$ and $\theta_{i,j,2} = \mathrm{Var}((Y_i - \mu_{i,2})(Y_j - \mu_{j,2}))$. Note that $\theta_{i,j,1}$ and $\theta_{i,j,2}$ can be respectively estimated by

$$\hat{\theta}_{i,j,1} = \frac{1}{n_1}\sum_{k=1}^{n_1}\Big[(X_{k,i}-\bar{X}_i)(X_{k,j}-\bar{X}_j)-\hat{\sigma}_{i,j,1}\Big]^2,\ \hat{\theta}_{i,j,2} = \frac{1}{n_2}\sum_{k=1}^{n_2}\Big[(Y_{k,i}-\bar{Y}_i)(Y_{k,j}-\bar{Y}_j)-\hat{\sigma}_{i,j,2}\Big]^2,$$

where $\bar{X}_i = \frac{1}{n_1}\sum_{k=1}^{n_1} X_{k,i}$ and $\bar{Y}_i = \frac{1}{n_2}\sum_{k=1}^{n_2} Y_{k,i}$. Cai and Liu (2011) used such a variance estimator for adaptive estimation of a sparse covariance matrix. Given the variance estimators $\hat{\theta}_{i,j,1}$ and $\hat{\theta}_{i,j,2}$, one can standardize the squared difference of the sample covariances $(\hat{\sigma}_{i,j,1} - \hat{\sigma}_{i,j,2})^2$ as

$$T_{i,j} = \frac{(\hat{\sigma}_{i,j,1} - \hat{\sigma}_{i,j,2})^2}{\hat{\theta}_{i,j,1}/n_1 + \hat{\theta}_{i,j,2}/n_2}, \quad 1 \leq i \leq j \leq p. \tag{6}$$

Cai et al. (2013) proposed to use the maximum of the $T_{i,j}$ as the test statistic

$$M_{n,p} = \max_{1\le i\le j\le p} T_{i,j} = \max_{1\le i\le j\le p} \frac{(\hat{\sigma}_{i,j,1} - \hat{\sigma}_{i,j,2})^2}{\hat{\theta}_{i,j,1}/n_1 + \hat{\theta}_{i,j,2}/n_2} \tag{7}$$

for testing the null $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. It is shown that, under H_0 and regularity conditions, $M_{n,p}$ has asymptotically the extreme value distribution of type I as in (4). The technical difficulty in establishing the asymptotic null distribution lies in dealing with the dependence among the $T_{i,j}$. The limiting null distribution of $M_{n,p}$ leads naturally to the test

$$\Phi_\alpha = I\{M_{n,p} \ge q_\alpha + 4\log p - \log\log p\} \tag{8}$$

for a given significance level $0 < \alpha < 1$, where q_α is the $1-\alpha$ quantile of the type I extreme value distribution given in (5). The hypothesis $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ is rejected whenever $\Phi_\alpha = 1$.

The test Φ_α is particularly well suited for testing against sparse alternatives. The theoretical analysis shows that the proposed test enjoys certain optimality against a large class of sparse alternatives in terms of the power. It only requires one of the entries of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ having a magnitude more than $C\sqrt{\log p/n}$ in order for the test to correctly reject the null H_0. It is also shown that this lower bound $C\sqrt{\log p/n}$ is rate-optimal.

In addition to the test based on the maximum entrywise deviations proposed in Cai et al. (2013), there are several testing procedures based on the Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$. Schott (2007) introduced a test using an estimate of the squared Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$. Srivastava and Yanagihara (2010) constructed a test that relied on a measure of distance by $\mathrm{tr}(\mathbf{\Sigma}_1^2)/(\mathrm{tr}(\mathbf{\Sigma}_1))^2 - \mathrm{tr}(\mathbf{\Sigma}_2^2)/(\mathrm{tr}(\mathbf{\Sigma}_2))^2$. Both of these two tests are designed for the Gaussian setting. Li and Chen (2012) proposed a test using a combination of U-statistics which was also motivated by an unbiased estimator of the squared Frobenius norm of $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$. These Frobenius norm-based testing procedures perform well when $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is "dense", but they are not powerful against sparse alternatives.

2.4. Global Testing for Differential Correlations $H_0 : \mathbf{R}_1 - \mathbf{R}_2 = 0$

In many applications, the correlation matrices are of more direct interest than the covariance matrices. In this case, it is naturally to test for the equality of two correlation matrices,

$$H_0 : \mathbf{R}_1 - \mathbf{R}_2 = 0 \quad \text{v.s.} \quad H_1 : \mathbf{R}_1 - \mathbf{R}_2 \ne 0. \tag{9}$$

This testing problem was considered in Cai and Zhang (2016).

Let the sample covariance matrices $\hat{\mathbf{\Sigma}}_1 = (\hat{\sigma}_{i,j,1})_{p\times p}$ and $\hat{\mathbf{\Sigma}}_2 = (\hat{\sigma}_{i,j,2})_{p\times p}$ be given as in (2). The sample correlation matrices are $\hat{\mathbf{R}}_d = (\hat{r}_{i,j,d})_{p\times p}$ with

$$\hat{r}_{i,j,d} = \frac{\hat{\sigma}_{i,j,d}}{(\hat{\sigma}_{i,i,d}\hat{\sigma}_{j,j,d})^{1/2}}, \quad 1 \le i \le j \le p,\ d = 1, 2.$$

Similar to testing for the equality of two covariance matrices, a natural test for testing the hypotheses in (9) is to use the maximum entrywise differences of the two sample correlation matrices. And for the same reason, it is also necessary to standardize $\hat{r}_{i,j,1} - \hat{r}_{i,j,2}$ before making the comparisons. Define

$$\eta_{i,j,1} = \mathrm{Var}\left[\frac{(X_i - \mu_{i,1})(X_j - \mu_{j,1})}{(\sigma_{i,i,1}\sigma_{j,j,1})^{1/2}} - \frac{r_{i,j,1}}{2}\left(\frac{(X_i - \mu_{i,1})^2}{\sigma_{i,i,1}} + \frac{(X_j - \mu_{j,1})^2}{\sigma_{j,j,1}}\right)\right]$$

and define $\eta_{i,j,2}$ similarly. Then, for $d = 1$ and 2, asymptotically as $n, p \to \infty$,

$$\sqrt{n}(\hat{r}_{i,j,d} - r_{i,j,d}) \approx \sqrt{\eta_{i,j,d}}\, z_{i,j,d}, \quad \text{where} \quad z_{i,j,d} \sim N(0,1).$$

The asymptotic variances are unknown but can be estimated by

$$\hat{\eta}_{i,j,1} = \frac{1}{n_1} \sum_{k=1}^{n_1} \left\{ \frac{(X_{i,k} - \bar{X}_i)(X_{j,k} - \bar{X}_j)}{(\hat{\sigma}_{i,i,1}\hat{\sigma}_{j,j,1})^{1/2}} - \frac{\hat{r}_{i,j,1}}{2} \left(\frac{(X_{i,k} - \bar{X}_i)^2}{\hat{\sigma}_{i,i,1}} + \frac{(X_{j,k} - \bar{X}_j)^2}{\hat{\sigma}_{j,j,1}} \right) \right\}^2$$

with $\hat{\eta}_{i,j,2}$ defined similarly. Given the variance estimates $\hat{\eta}_{i,j,1}$ and $\hat{\eta}_{i,j,2}$, we standardize the squared difference of the sample correlations $(\hat{r}_{i,j,1} - \hat{r}_{i,j,2})^2$ as

$$T_{i,j} = \frac{(\hat{r}_{i,j,1} - \hat{r}_{i,j,2})^2}{\hat{\eta}_{i,j,1}/n_1 + \hat{\eta}_{i,j,2}/n_2}, \quad 1 \le i, j \le p,$$

and define the test statistic for testing $H_0 : \mathbf{R}_1 - \mathbf{R}_2 = 0$ by

$$M_{n,p} = \max_{1 \le i \le j \le p} T_{i,j}.$$

Under regularity conditions, $M_{n,p}$ can be shown to have asymptotically the extreme value distribution of type I as in (4). The result yields immediately a test for testing the hypothesis $H_0 : \mathbf{R}_1 - \mathbf{R}_2 = 0$ at a given significance level $0 < \alpha < 1$,

$$\Psi_\alpha = I(M_{n,p} \ge 4 \log p - \log \log p + q_\alpha) \tag{10}$$

where q_α is given in (5). The null hypothesis $H_0 : \mathbf{R}_1 - \mathbf{R}_2 = 0$ is rejected whenever $\Psi_\alpha = 1$. The test Ψ_α shares similar properties as the test proposed in Cai et al. (2013) for testing the equality of two covariance matrices. In particular, it is also particularly well suited for testing against sparse alternatives.

2.5. Global Testing for Differential Network $H_0 : \Omega_1 - \Omega_2 = 0$

Precision matrix plays a fundamental role in many high-dimensional inference problems. In the Gaussian graphical model framework, the difference of two precision matrices $\mathbf{\Omega}_1 - \mathbf{\Omega}_2$ characterizes the differential network, which measures the amount of changes in the network between two states. The first problem to consider is the global detection problem: Is there any change in the two networks? This can be formulated as a global testing problem.

The equality of two precision matrices is mathematically equivalent to the equality of two covariance matrices, which, as discussed earlier, has been well studied. However, in many applications it is often reasonable to assume that $\mathbf{\Delta}$ is sparse under the alternative, while $\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2$ is not. Furthermore, it is of significant interest to identify the locations where $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ differ from each other. So it is essential to work on the precision matrices directly, not the covariance matrices.

Suppose we observe two independent random samples $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ with the precision matrix $\mathbf{\Omega}_1 = (\omega_{i,j,1})_{p \times p} = \mathbf{\Sigma}_1^{-1}$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$ with $\mathbf{\Omega}_2 = (\omega_{i,j,2})_{p \times p} = \mathbf{\Sigma}_2^{-1}$. Let $\mathbf{\Delta} = (\delta_{i,j}) = \mathbf{\Omega}_1 - \mathbf{\Omega}_2$ be the differential network. We wish to test

$$H_0 : \ \mathbf{\Delta} = 0 \quad \text{versus} \quad H_1 : \ \mathbf{\Delta} \neq 0.$$

A testing procedure that is based on a penalized regression approach was recently proposed in Xia et al. (2015) and is shown to be powerful against sparse alternatives. Note that the null hypothesis $H_0 : \mathbf{\Delta} = 0$ is equivalent to the hypothesis

$$H_0 : \max_{1\le i\le j\le p} |\omega_{i,j,1} - \omega_{i,j,2}| = 0,$$

an intuitive approach to test H_0 is to first construct estimators of $\omega_{i,j,d}$, $d = 1$ and 2, and then use the maximum standardized differences as the test statistic.

Compared to testing the equality of covariance matrices or correlation matrices, testing $H_0 : \mathbf{\Omega}_1 = \mathbf{\Omega}_2$ in the high-dimensional setting is technically much more challenging as there is no sample precision matrix that one can use as the starting point. Xia et al. (2015) takes the approach of relating the entries of a precision matrix to the coefficients of a set of regression models, and then constructs test statistics based on the covariances between the residuals of the fitted regression models.

In the Gaussian setting, the precision matrix can be described through linear regression models (Anderson 2003). Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_{n_1})^\intercal$ and $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2})^\intercal$ denote the data matrices. Then one can write

$$X_{k,i} = \alpha_{i,1} + \mathbf{X}_{k,-i}\boldsymbol{\beta}_{i,1} + \epsilon_{k,i,1}, \quad (i = 1, \ldots, p;\ k = 1, \ldots, n_1), \tag{11}$$

$$Y_{k,i} = \alpha_{i,2} + \mathbf{Y}_{k,-i}\boldsymbol{\beta}_{i,2} + \epsilon_{k,i,2}, \quad (i = 1, \ldots, p;\ k = 1, \ldots, n_2), \tag{12}$$

where $\epsilon_{k,i,d} \sim N(0, \sigma_{i,i,d} - \mathbf{\Sigma}_{i,-i,d}\mathbf{\Sigma}^{-1}_{-i,-i,d}\mathbf{\Sigma}_{-i,i,d})$, $d = 1, 2$, are independent of $\mathbf{X}_{k,-i}$ and $\mathbf{Y}_{k,-i}$. The regression coefficients satisfy $\alpha_{i,d} = \mu_{i,d} - \mathbf{\Sigma}_{i,-i,d}\mathbf{\Sigma}^{-1}_{-i,-i,d}\boldsymbol{\mu}_{-i,d}$ and $\beta_{i,d} = -\omega^{-1}_{i,i,d}\mathbf{\Omega}_{-i,i,d}$. To construct the test statistics, it is important to understand the covariances among the error terms $\epsilon_{k,i,d}$,

$$\tau_{i,j,d} \equiv \mathrm{Cov}(\epsilon_{k,i,d}, \epsilon_{k,j,d}) = \frac{\omega_{i,j,d}}{\omega_{i,i,d}\omega_{j,j,d}}.$$

In order to construct the test statistic for testing $H_0 : \mathbf{\Omega}_1 = \mathbf{\Omega}_2$, one first construct estimators of $\tau_{i,j,d}$, and then obtain estimators of $\omega_{i,j,d}$, and finally uses the maximum standardized differences as the test statistic.

Step 1: Construction of the estimators of $\tau_{i,j,d}$. Let $\hat{\boldsymbol{\beta}}_{i,d} = (\hat{\beta}_{1,i,d}, \ldots, \hat{\beta}_{p-1,i,d})^\intercal$ be estimators of $\boldsymbol{\beta}_{i,d}$ satisfying

$$\max_{1\le i\le p} |\hat{\boldsymbol{\beta}}_{i,d} - \boldsymbol{\beta}_{i,d}|_1 = o_{\mathrm{P}}\{(\log p)^{-1}\}, \tag{13}$$

$$\max_{1\le i\le p} |\hat{\boldsymbol{\beta}}_{i,d} - \boldsymbol{\beta}_{i,d}|_2 = o_{\mathrm{P}}\{(n_d \log p)^{-1/4}\}. \tag{14}$$

Estimators $\hat{\boldsymbol{\beta}}_{i,d}$ that satisfy (13) and (14) can be obtained easily using, for example, the Lasso, scale Lasso, or Dantzig selector. Define the residuals by

$$\hat{\epsilon}_{k,i,1} = X_{k,i} - \bar{X}_i - (\mathbf{X}_{k,-i} - \bar{\mathbf{X}}_{\cdot,-i})\hat{\boldsymbol{\beta}}_{i,1}, \quad \hat{\epsilon}_{k,i,2} = Y_{k,i} - \bar{Y}_i - (\mathbf{Y}_{k,-i} - \bar{\mathbf{Y}}_{\cdot,-i})\hat{\boldsymbol{\beta}}_{i,2}.$$

A natural estimator of $\tau_{i,j,d}$ is the sample covariance between the residuals,

$$\tilde{\tau}_{i,j,d} = \frac{1}{n_d}\sum_{k=1}^{n_d} \hat{\epsilon}_{k,i,d}\hat{\epsilon}_{k,j,d}. \tag{15}$$

However, when $i \neq j$, $\tilde{\tau}_{i,j,d}$ tends to be biased due to the correlation induced by the estimated parameters and it is desirable to construct a bias-corrected estimator. For $1 \leq i < j \leq p$, $\beta_{i,j,d} = -\omega_{i,j,d}/\omega_{j,j,d}$ and $\beta_{j-1,i,d} = -\omega_{i,j,d}/\omega_{i,i,d}$. Xia et al. (2015) introduced a bias-corrected estimator of $\tau_{i,j,d}$ as

$$\hat{\tau}_{i,j,d} = -(\tilde{\tau}_{i,j,d} + \tilde{\tau}_{i,i,d}\hat{\beta}_{i,j,d} + \tilde{\tau}_{j,j,d}\hat{\beta}_{j-1,i,d}), \quad 1 \leq i < j \leq p. \tag{16}$$

The bias of $\hat{\tau}_{i,j,d}$ is of order $\max\{\tau_{i,j,d}(\log p/n_d)^{1/2}, (n_d \log p)^{-1/2}\}$. For the diagonal entries, note that $\tau_{i,i,d} = 1/\omega_{i,i,d}$ and $\max_{1\leq i\leq p} |\tilde{\tau}_{i,i,d} - \tau_{i,i,d}| = O_{\mathrm{p}}\{(\log p/n_d)^{1/2}\}$, which implies that $\hat{\tau}_{i,i,d} = \tilde{\tau}_{i,i,d}$ is a nearly unbiased estimator of $\tau_{i,i,d}$. A natural estimator of $\omega_{i,j,d}$ can then be defined by

$$F_{i,j,d} = \frac{\hat{\tau}_{i,j,d}}{\hat{\tau}_{i,i,d}\hat{\tau}_{j,j,d}}, \quad 1 \leq i \leq j \leq p. \tag{17}$$

Step 2: Standardization. It is natural to test $H_0 : \mathbf{\Delta} = 0$ based on $F_{i,j,1} - F_{i,j,2}$, $1 \leq i \leq j \leq p$. However, the estimators $F_{i,j,1} - F_{i,j,2}$ are heteroscedastic and possibly have a wide range of variability. Standardization is thus needed in order to compare $F_{i,j,1} - F_{i,j,2}$ for different entries. This requires an estimate of the variance of $F_{i,j,1} - F_{i,j,2}$.

Let $U_{i,j,d} = (1/n_d)\sum_{k=1}^{n_d}\{\epsilon_{k,i,d}\epsilon_{k,j,d} - \mathbb{E}(\epsilon_{k,i,d}\epsilon_{k,j,d})\}$ and $\tilde{U}_{i,j,d} = (\tau_{i,j,d} - U_{i,j,d})/(\tau_{i,i,d}\tau_{j,j,d})$. Xia et al. (2015) showed that, uniformly in $1 \leq i < j \leq p$,

$$|F_{i,j,d} - \tilde{U}_{i,j,d}| = O_{\mathrm{p}}\{(\log p/n_d)^{\frac{1}{2}}\}\tau_{i,j,d} + o_{\mathrm{p}}\{(n_d \log p)^{-\frac{1}{2}}\}.$$

Let $\theta_{i,j,d} = \mathrm{Var}(\tilde{U}_{i,j,d})$. Note that

$$\theta_{i,j,d} = \mathrm{Var}\{\epsilon_{k,i,d}\epsilon_{k,j,d}/(\tau_{i,i,d}\tau_{j,j,d})\}/n_d = (1 + \rho^2_{i,j,d})/(n_d\tau_{i,i,d}\tau_{j,j,d}),$$

where $\rho^2_{i,j,d} = \beta^2_{i,j,d}\tau_{i,i,d}/\tau_{j,j,d}$. The variance $\theta_{i,j,d}$ can then be estimated by

$$\hat{\theta}_{i,j,d} = (1 + \hat{\beta}^2_{i,j,d}\hat{\tau}_{i,i,d}/\hat{\tau}_{j,j,d})/(n_d\hat{\tau}_{i,i,d}\hat{\tau}_{j,j,d}).$$

Define the standardized statistics

$$T_{i,j} = \frac{F_{i,j,1} - F_{i,j,2}}{(\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2})^{1/2}}, \quad 1 \leq i \leq j \leq p. \tag{18}$$

Step 3: The test statistics. Xia et al. (2015) proposed the following test statistic for testing the global null H_0,

$$M_{n,p} = \max_{1\leq i\leq j\leq p} T^2_{i,j} = \max_{1\leq i\leq j\leq p} \frac{(F_{i,j,1} - F_{i,j,2})^2}{\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2}}. \tag{19}$$

Xia et al. (2015) showed that, under regularity conditions, $M_{n,p}$ has asymptotically the extreme value distribution of type I as in (4) which leads to the following test

$$\Psi_\alpha = I(M_{n,p} \geq q_\alpha + 4\log p - \log\log p) \tag{20}$$

where q_α is given in (5). The hypothesis H_0 is rejected whenever $\Psi_\alpha = 1$. Compared with global testing for covariance or correlation matrices, the technical analysis for testing the equality of two precision matrices is more involved. See Xia et al. (2015) for detailed proofs.

3. Large-Scale Multiple Testing for Local Covariance Structures

Multiple testing for local covariance structures has not been as well studied as global testing problems. In this section, we discuss recently proposed methods for simultaneous testing for correlations (Cai and Liu 2015), Gaussian graphical models (Liu 2013), and differential networks (Xia et al. 2015). In each of these problems, the goal is to simultaneously test a large number of hypotheses on local covariance structures, $H_{0,i,j}$, $1 \le i < j \le p$, while controlling the FDP and FDR. The nulls can be, for example, $H_{0,i,j} : r_{i,j} = 0$ where $r_{i,j}$ is the correlation between variables i and j, or $H_{0,i,j} : \omega_{i,j,1} - \omega_{i,j,2} = 0$ where $\omega_{i,j,1}$ and $\omega_{i,j,2}$ are the (i, j)-th entry of two precision matrices corresponding to Gaussian graphical models. These multiple testing procedures share some important features and it is helpful to discuss the main ideas and common features before introducing the specific procedures.

3.1. The Main Ideas and Common Features

Each multiple testing procedure is developed in two stages. The first step is to construct a test statistic $T_{i,j}$ for testing an individual hypothesis $H_{0,i,j}$ and establish its null distribution, and the second step is to develop a multiple testing procedure to account for the multiplicity in testing a large number of hypotheses and the dependency among the $T_{i,j}$'s so that the overall FDP and FDR are controlled. For the three simultaneous testing problems under consideration, the major difference lies in the first step. The second step is common to all three problems. An important component in the second step is the estimation of the proportion of the nulls falsely rejected among all the true nulls at any given threshold level.

Step 1: Construction of Test Statistics For each of these multiple testing problems, a test statistic $T_{i,j}$ is constructed for testing an individual null hypothesis $H_{0,i,j}$. The construction of the test statistics $T_{i,j}$ can be involved and varies significantly from problem to problem. The specific constructions will be discussed in detail separately later. Under the null $H_{0,i,j}$ and regularity conditions, $T_{i,j}$ is shown to have asymptotically standard normal distribution,

$$T_{i,j} \to N(0, 1).$$

The null hypotheses $H_{0,i,j}$ are rejected whenever $|T_{i,j}| \ge t$ for some threshold level $t > 0$. The choice of t is important.

Step 2: Construction of a Multiple Testing Procedure Once the test statistics $T_{i,j}$ are constructed and the null distribution of $T_{i,j}$ is obtained, the next step is to develop a multiple testing procedure that accounts for the multiplicity and dependency among the $T_{i,j}$'s. This step is common to all three multiple testing problems under consideration. Let t be the threshold level such that the null hypotheses $H_{0,i,j}$ are rejected when $|T_{i,j}| \ge t$. For any given t, the total number of rejections is

$$R(t) = \sum_{1 \le i < j \le p} I\{|T_{i,j}| \ge t\} \tag{21}$$

and the total number of false rejections is

$$R_0(t) = \sum_{(i,j) \in \mathcal{H}_0} I\{|T_{i,j}| \ge t\} \tag{22}$$

where $\mathcal{H}_0 = \{(i,j) : 1 \le i < j \le p,\ H_{0,i,j} \text{ is true}\}$ is the set of true null hypotheses. The false discovery proportion (FDP) and false discovery rate (FDR) are defined as

$$\mathrm{FDP}(t) = \frac{R_0(t)}{R(t) \vee 1} \quad \text{and} \quad \mathrm{FDR}(t) = \mathbb{E}[\mathrm{FDP}(t)]. \tag{23}$$

The fact that the asymptotic null distribution of $T_{i,j}$ is standard normal implies that

$$\mathbb{P}\left(\max_{(i,j)\in\mathcal{H}_0} |T_{i,j}| \ge 2\sqrt{\log p}\right) \to 0 \quad \text{as } (n,p) \to \infty. \tag{24}$$

An ideal choice of t is thus

$$t_* = \inf\left\{0 \le t \le \sqrt{2\log p} : \frac{R_0(t)}{R(t) \vee 1} \le \alpha\right\}.$$

With this choice of the threshold, the multiple testing procedure would reject as many true positives as possible while controlling the FDR at the given level α. However, the total number of false positives, $R_0(t)$, is unknown as the set $\mathcal{H}_0$ is unknown. Thus this ideal choice of t is not available and a data-driven choice of the threshold needs to be developed.

An important step in constructing the multiple testing procedure is estimating the proportion of the nulls falsely rejected by the procedure among all the true nulls at the threshold level t,

$$G_0(t) = \frac{R_0(t)}{|\mathcal{H}_0|}. \tag{25}$$

It can be shown that, under regularity conditions,

$$\sup_{0 \le t \le b_p} \left| \frac{G_0(t)}{G(t)} - 1 \right| \to 0 \tag{26}$$

in probability as $(n,p) \to \infty$, where $G(t) = 2 - 2\Phi(t)$ is the tail probability of the standard normal distribution, $b_p = \sqrt{4\log p - a_p}$ and $a_p = 2\log(\log p)$. The upper bound b_p is near-optimal for (26) to hold in the sense that a_p cannot be replaced by any constant. Equation (26) shows that $G_0(t)$ is well approximated by $G(t)$ and one can estimate the total number of false rejections $R_0(t)$ by $G(t)(p^2 - p)/2$.

This analysis leads to the following multiple testing procedure for covariance structures.

Algorithm 1 Large-scale multiple testing for local covariance structures

1: Calculate the test statistics $T_{i,j}$ for $1 \le i < j \le p$.

2: For given $0 \le \alpha \le 1$, calculate

$$\hat{t} = \inf\left\{0 \le t \le b_p : \frac{G(t)(p^2 - p)/2}{R(t) \vee 1} \le \alpha\right\}. \tag{27}$$

If (27) does not exist, then set $\hat{t} = 2\sqrt{\log p}$.

3: Reject $H_{0,i,j}$ whenever $|T_{i,j}| \ge \hat{t}$.

It is helpful to explain the algorithm in more detail.

1. The above multiple testing procedure uses $(p^2 - p)/2$ as the estimate for the number of the true nulls. In many applications, the number of the true significant alternatives is relatively small. In such a sparse setting, one has $|\mathcal{H}_0|/((p^2 - p)/2) \approx 1$ and the FDP and FDR levels of the testing procedure would be close to the nominal level α.

2. Note that $G(t)$ is used to estimate $G_0(t)$ in the above procedure only in the range $0 \leq t \leq b_p$. This is done for an important reason. For $t \geq \sqrt{4\log p - \log(\log p) + O(1)}$, $G(t)$ is not a consistent estimator of $G_0(t)$ since $p^2 G(t)$ is bounded. Thus, when $\hat{t}$ in (27) does not exist, the test statistic $|T_{i,j}|$ is thresholded at $2\sqrt{\log p}$ directly to control the FDP.
3. It is also interesting to compare the proposed procedure with the well-known Benjamini-Hochberg (B-H) procedure. The B-H procedure with p-values $G(|T_{i,j}|)$ is equivalent to rejecting $H_{0,i,j}$ if $|T_{i,j}| \geq \hat{t}_{\mathrm{BH}}$, where $\hat{t}_{\mathrm{BH}}$ is defined as

$$\hat{t}_{\mathrm{BH}} = \inf\left\{t \geq 0 : \frac{G(t)(p^2 - p)/2}{R(t) \vee 1} \leq \alpha\right\}. \tag{28}$$

Note that the difference between (27) and (28) lies in the range for t. It is important to restrict the range of t to $[0,\ b_p]$ in (27). The B-H procedure uses $G(t)$ to estimate $G_0(t)$ for all $t \geq 0$. As a result, when the number of true alternatives $|\mathcal{H}_0^c|$ is fixed as $p \to \infty$, with some positive probability, the B-H method is unable to control the FDP, even in the independent case. See more discussions in Cai and Liu (2015) and Liu and Shao (2014).

3.2. Multiple Testing for Correlations

Correlation detection is an important problem and has a number of applications. Cai and Liu (2015) considered multiple testing for correlations in both the one-sample and two-sample settings. We discuss below these two settings separately.

3.2.1. One-Sample Testing. Let us begin with the relatively simple case of one-sample testing, where the goal is to detect significant correlations among the variables by simultaneously testing the hypotheses $H_{0,i,j} : r_{i,j} = 0$ versus $H_{1,i,j} : r_{i,j} \neq 0$ for $1 \leq i < j \leq p$. Note that this is equivalent to multiple testing for the covariances,

$$H_{0,i,j}: \quad \sigma_{i,j} = 0 \quad \text{versus} \quad H_{1,i,j}: \quad \sigma_{i,j} \neq 0, \quad \text{for} \quad 1 \leq i < j \leq p. \tag{29}$$

The first step in developing the multiple testing procedure is to construct a test statistic for testing each hypothesis $H_{0,i,j}$. A natural starting point for testing the hypotheses in (29) is the sample covariance matrix $\hat{\mathbf{\Sigma}} = (\hat{\sigma}_{i,j})$ defined in (1). As mentioned in Section 2, the sample covariances $\hat{\sigma}_{i,j}$ are heteroscedastic and may have a wide range of variability, it is necessary to standardize $\hat{\sigma}_{i,j}$ before comparing them. Let $\theta_{i,j} = \mathrm{Var}((X_i - \mu_i)(X_j - \mu_j))$. As noted in Cai and Liu (2011), a consistent estimate of $\theta_{i,j}$ is given by

$$\hat{\theta}_{i,j} = \frac{1}{n}\sum_{k=1}^{n}[(X_{k,i} - \bar{X}_i)(X_{k,j} - \bar{X}_j) - \hat{\sigma}_{i,j}]^2.$$

It is thus natural to normalize the sample covariances as

$$T_{i,j} = \frac{\hat{\sigma}_{i,j}}{\sqrt{\hat{\theta}_{i,j}}} = \frac{\sum_{k=1}^{n}(X_{k,i} - \bar{X}_i)(X_{k,j} - \bar{X}_j)}{\sqrt{n\hat{\theta}_{i,j}}}, \tag{30}$$

and use them as the test statistics for simultaneous testing of $H_{0,i,j}$. The null distribution of $T_{i,j}$ is relatively easy to establish. It follows from the central limit theorem and the law of large numbers that under $H_{0,i,j}$ and the moment condition $\mathbb{E}(X_i - \mu_i)^4/\sigma_{ii}^2 < \infty$,

$$T_{i,j} \to N(0, 1).$$

Once the test statistics $T_{i,j}$ are constructed and their null distributions are established, one can apply Step 2 discussed in Section 3.1 to construct a multiple testing procedure that accounts for the multiplicity and dependency among the $T_{i,j}$'s. The large-scale simultaneous testing procedure for correlations is given as in Algorithm 1 with $T_{i,j}$ computed from (30).

Remark 1. The normal approximation is suitable when the sample size is large. In the case of small sample size, a bootstrap procedure can be used to improve the accuracy of the approximation to $G_0(t)$. See Cai and Liu (2015) for more details.

Under mild regularity conditions, the multiple testing procedure given in Algorithm 1 with $T_{i,j}$ computed from (30) controls both the FDP and FDR asymptotically,

$$\overline{\lim_{(n,p)\to\infty}} \text{FDR} \le \alpha \quad \text{and} \quad \lim_{(n,p)\to\infty} \mathbb{P}(\text{FDP} \le \alpha + \varepsilon) = 1 \tag{31}$$

for any $\varepsilon > 0$. If in addition the number of significant true alternatives is at least of order $\sqrt{\log(\log p)}$, then the FDP and FDR will converge to $\alpha\tau$ where $\tau = |\mathcal{H}_0|/((p^2 - p)/2)$, i.e.,

$$\lim_{(n,p)\to\infty} \frac{\text{FDR}}{\alpha\tau} = 1$$

and $\frac{\text{FDP}}{\alpha\tau} \to 1$ in probability as $(n,p) \to \infty$. See Cai and Liu (2015) for a detailed proof.

3.2.2. Two-Sample Testing. We now turn to the two-sample case where one wishes to simultaneously test the entries of the differential correlation matrix $\mathbf{R}_1 - \mathbf{R}_2 = (r_{i,j,1} - r_{i,j,2})$,

$$H_{0,i,j}: \quad r_{i,j,1} - r_{i,j,2} = 0 \quad \text{versus} \quad H_{1,i,j}: \quad r_{i,j,1} - r_{i,j,2} \ne 0, \quad \text{for} \quad 1 \le i < j \le p. \tag{32}$$

This multiple testing problem is technically more involved than the one-sample case as each null $H_{0,i,j}$ is a composite hypothesis and it cannot be translated into a simple hypothesis on the covariances. As in the one-sample case, the first step in developing the multiple testing procedure is the construction of a suitable test statistic for the individual hypothesis and establish its null distribution.

We begin by constructing a test statistic for testing the equality of each individual pair of correlations, $H_{0,i,j} : r_{i,j,1} = r_{i,j,2}$. A natural starting point is the sample correlations

$$\hat{r}_{i,j,1} = \frac{\sum_{k=1}^{n_1}(X_{k,i} - \bar{X}_i)(X_{k,j} - \bar{X}_j)}{\sqrt{\sum_{k=1}^{n_1}(X_{k,i} - \bar{X}_i)^2 \sum_{k=1}^{n_1}(X_{k,j} - \bar{X}_j)^2}}, \ \hat{r}_{i,j,2} = \frac{\sum_{k=1}^{n_2}(Y_{k,i} - \bar{Y}_i)(Y_{k,j} - \bar{Y}_j)}{\sqrt{\sum_{k=1}^{n_2}(Y_{k,i} - \bar{Y}_i)^2 \sum_{k=1}^{n_2}(Y_{k,j} - \bar{Y}_j)^2}},$$

where $\bar{X}_i = \frac{1}{n_1}\sum_{k=1}^{n_1} X_{k,i}$ and $\bar{Y}_i = \frac{1}{n_2}\sum_{k=1}^{n_2} Y_{k,i}$. Similar to the sample covariances, $\hat{r}_{i,j,1}$ and $\hat{r}_{i,j,2}$ are heteroscedastic and are thus not directly comparable. Cai and Liu (2015) assumes a moment condition on the distribution, which is satisfied by the class of the elliptically contoured distributions. This is clearly a much larger class than the class of multivariate normal distributions. Under regularity conditions,

$$\frac{\hat{r}_{i,j,1} - \hat{r}_{i,j,2}}{\sqrt{\frac{\kappa_1}{n_1}(1 - r_{i,j,1}^2)^2 + \frac{\kappa_2}{n_2}(1 - r_{i,j,2}^2)^2}} \to N(0,1) \tag{33}$$

with $\kappa_1 \equiv \frac{1}{3}\frac{\mathbb{E}(X_i - \mu_{i,1})^4}{[\mathbb{E}(X_i - \mu_{i,1})^2]^2}$ and $\kappa_2 \equiv \frac{1}{3}\frac{\mathbb{E}(Y_i - \mu_{i,2})^4}{[\mathbb{E}(Y_i - \mu_{i,2})^2]^2}$. Note that $\kappa_1 = \kappa_2 = 1$ for multivariate normal distributions.

In general, the parameters $r_{i,j,1}$, $r_{i,j,2}$, κ_1 and κ_2 in the denominator of (33) are unknown and need to be estimated. κ_1 and κ_2 can be estimated respectively by

$$\hat{\kappa}_1 = \frac{1}{3p}\sum_{i=1}^{p}\frac{n_1\sum_{k=1}^{n_1}(X_{k,i}-\bar{X}_i)^4}{[\sum_{k=1}^{n_1}(X_{k,i}-\bar{X}_i)^2]^2} \quad \text{and} \quad \hat{\kappa}_2 = \frac{1}{3p}\sum_{i=1}^{p}\frac{n_2\sum_{k=1}^{n_2}(Y_{k,i}-\bar{Y}_i)^4}{[\sum_{k=1}^{n_2}(Y_{k,i}-\bar{Y}_i)^2]^2}.$$

Taking into account of possible sparsity of the correlation matrices, a thresholded version of the sample correlation coefficients can be used to estimate $r_{i,j,1}$ and $r_{i,j,2}$,

$$\tilde{r}_{i,j,d} \ = \ \hat{r}_{ijl} I\left\{\frac{|\hat{r}_{i,j,d}|}{\sqrt{\frac{\hat{\kappa}_d}{n_d}(1-\hat{r}^2_{i,j,d})^2}} \geq 2\sqrt{\frac{\log p}{n_d}}\right\}, \quad d = 1,\ 2.$$

Let $\tilde{r}^2_{i,j} = \max\{\tilde{r}^2_{i,j,1}, \tilde{r}^2_{i,j,2}\}$. Cai and Liu (2015) proposed the following test statistic by replacing $r^2_{i,j,1}$ and $r^2_{i,j,2}$ in the denominator of (33) with $\tilde{r}^2_{i,j}$,

$$T_{i,j} = \frac{\hat{r}_{i,j,1} - \hat{r}_{i,j,2}}{\sqrt{\frac{\hat{\kappa}_1}{n_1}(1-\tilde{r}^2_{i,j})^2 + \frac{\hat{\kappa}_2}{n_2}(1-\tilde{r}^2_{i,j})^2}}, \tag{34}$$

for testing the hypothesis $H_{0,i,j} : r_{i,j,1} = r_{i,j,2}$. Note that under $H_{0,i,j}$, $\tilde{r}^2_{i,j}$ is a consistent estimator of $r_{i,j,1}$ and $r_{i,j,2}$. On the other hand, under the alternative $H_{1,i,j}$, $\sqrt{\frac{\hat{\kappa}_1}{n_1}(1-\tilde{r}^2_{i,j})^2 + \frac{\hat{\kappa}_2}{n_2}(1-\tilde{r}^2_{i,j})^2} \leq \sqrt{\frac{\hat{\kappa}_1}{n_1}(1-\tilde{r}^2_{i,j,1})^2 + \frac{\hat{\kappa}_2}{n_2}(1-\tilde{r}^2_{i,j,2})^2}$.

The test statistic $T_{i,j}$ has standard normal distribution asymptotically under the null hypothesis $H_{0,i,j}$ and it is robust against a class of non-normal population distributions of $\mathbf{X}$ and $\mathbf{Y}$. More precisely, under $H_{0,i,j}$ and some regularity conditions,

$$\sup_{0\leq t\leq b\sqrt{\log p}}\left|\frac{\mathbb{P}(|T_{i,j}| \geq t)}{G(t)} - 1\right| \to 0 \quad \text{as } (n_1, n_2) \to \infty$$

uniformly in $1 \leq i < j \leq p$ and $p \leq n^k$ for any $b > 0$ and $k > 0$.

After the test statistics $T_{i,j}$ for testing the individual hypothesis $H_{0,i,j} : r_{i,j,1} = r_{i,j,2}$ are constructed and their null distributions are obtained, one can again apply Step 2 discussed in Section 3.1. The multiple testing procedure for differential correlations is then given by Algorithm 1 with $T_{i,j}$ computed from (34). The properties of the proposed procedure are studied both theoretically and numerically in Cai and Liu (2015). It is shown that, under regularity conditions, the multiple testing procedure controls the overall FDP and FDR at the pre-specified level asymptotically. The proposed procedure works well even when the components of the random vectors are strongly dependent and hence provides theoretical guarantees for a large class of correlation matrices.

Remark 2. The multiple testing procedure for the one-sample case given in Section 3.2.1 was developed by first transforming the problem of testing the simple null hypotheses $H_{0,i,j}$: $r_{i,j} = 0$ to testing equivalent hypotheses on the covariances $H_{0,i,j} : \sigma_{i,j} = 0$, $1 \leq i < j \leq p$. A different multiple testing procedure can be constructed by using the same argument as in the two-sample case and treating the correlations directly. Under similar regularity conditions, the following test statistic can be used for testing each $H_{0,i,j} : r_{i,j} = 0$,

$$\tilde{T}_{i,j} = \frac{|\hat{r}_{i,j}|}{\sqrt{\frac{\hat{\kappa}}{n}}\ (1-\hat{r}^2_{i,j})},$$

where $\hat{\kappa} = \frac{1}{3p} \sum_{i=1}^{p} \frac{n \sum_{k=1}^{n} (X_{k,i} - \bar{X}_i)^4}{(\sum_{k=1}^{n} (X_{k,i} - \bar{X}_i)^2)^2}$ is an estimate of $\kappa \equiv \frac{1}{3} \frac{\mathbb{E}(X_i - \mu_i)^4}{[\mathbb{E}(X_i - \mu_i)^2]^2}$. It can be shown that the asymptotic null distribution of $\tilde{T}_{i,j}$ is standard normal. This combined with the same second step leads to a multiple testing procedure that controls the FDP and FDR.

Remark 3. As in the one-sample case, the normal approximation is suitable in the two-sample setting when the sample sizes are large. A similar bootstrap procedure can be used to improve the accuracy of the approximation when the sample sizes are small. See Cai and Liu (2015) for more details.

3.3. Multiple Testing for A Gaussian Graphical Model

Motivated by a range of applications for Gaussian graphical models, Liu (2013) considers multiple testing for the partial correlations under a Gaussian graphical model framework and introduced a testing procedure with the FDP and FDR control. Suppose we observe a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} = (\omega_{i,j})$ and wish to simultaneously test the hypotheses

$$H_{0,i,j} : \omega_{i,j} = 0 \quad \text{versus} \quad H_{1,i,j} : \omega_{i,j} \neq 0, \quad 1 \leq i < j \leq p \tag{35}$$

while controlling the FDP and FDR asymptotically at a pre-specified level $0 < \alpha < 1$.

As mentioned in Section 2.5, the precision matrix $\boldsymbol{\Omega}$ can be described through linear regression models. Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^\intercal$ denote the data matrix. Then one can write

$$X_{k,i} = \alpha_i + \mathbf{X}_{k,-i}\boldsymbol{\beta}_i + \epsilon_{k,i}, \quad (i = 1, \ldots, p;\ k = 1, \ldots, n), \tag{36}$$

where $\epsilon_{k,i} \sim N(0, \sigma_{i,i} - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i})$ are independent of $\mathbf{X}_{k,-i}$. The regression coefficients satisfy $\alpha_i = \mu_i - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\mu}_{-i}$ and $\beta_i = -\omega_{i,i}^{-1}\boldsymbol{\Omega}_{-i,i}$. Important quantities for the construction of the test statistics are the covariances among the residuals $\epsilon_{k,i}$ and $\epsilon_{k,j}$,

$$\tau_{i,j} \equiv \text{Cov}(\epsilon_{k,i}, \epsilon_{k,j}) = \frac{\omega_{i,j}}{\omega_{i,i}\omega_{j,j}}, \quad 1 \leq i < j \leq p.$$

The test statistic $T_{i,j}$ for testing the hypothesis $H_{0,i,j} : \omega_{i,j} = 0$ is obtained by first constructing an estimate of $\tau_{i,j}$, and then an estimate of $\omega_{i,j}$. We begin by constructing the estimators for $\tau_{i,j}$. Let $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{1,i}, \ldots, \hat{\beta}_{p-1,i})^\intercal$ be estimators of $\boldsymbol{\beta}_i$ satisfying

$$\max_{1 \leq i \leq p} |\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i|_1 = o_{\mathrm{P}}\{(\log p)^{-\frac{1}{2}}\}, \tag{37}$$

and

$$\min\left\{\lambda_{\max}^{\frac{1}{2}}(\boldsymbol{\Sigma}) \max_{1 \leq i \leq p} |\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i|_2, \quad \max_{1 \leq i \leq p} \sqrt{(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)^\intercal \hat{\boldsymbol{\Sigma}}_{-i,-i}^{-1} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)}\right\} = o_{\mathrm{P}}(n^{-\frac{1}{4}}). \tag{38}$$

Estimators $\hat{\boldsymbol{\beta}}_i$ that satisfy (37) and (38) can be obtained easily using, for example, the Lasso, scale Lasso, or Dantzig selector. Define the residuals by

$$\hat{\epsilon}_{k,i} = X_{k,i} - \bar{X}_i - (\mathbf{X}_{k,-i} - \bar{\mathbf{X}}_{\cdot,-i})\hat{\boldsymbol{\beta}}_i.$$

A natural estimator of $\tau_{i,j}$ is the sample covariance between the residuals, $\tilde{\tau}_{i,j} = \frac{1}{n}\sum_{k=1}^{n} \hat{\epsilon}_{k,i}\hat{\epsilon}_{k,j}$. However, when $i \neq j$, $\tilde{\tau}_{i,j}$ tends to be biased due to the correlation induced by the estimated parameters. For $1 \le i < j \le p$, $\beta_{i,j} = -\omega_{i,j}/\omega_{j,j}$ and $\beta_{j-1,i} = -\omega_{i,j}/\omega_{i,i}$. Liu (2013) introduced a debiased estimator of $\tau_{i,j}$ as

$$\hat{\tau}_{i,j} = \tilde{\tau}_{i,j} + \tilde{\tau}_{i,i}\hat{\beta}_{i,j} + \tilde{\tau}_{j,j}\hat{\beta}_{j-1,i}, \quad 1 \le i < j \le p \tag{39}$$

and for $i = j$, define $\hat{\tau}_{i,i} = \tilde{\tau}_{i,i}$. Let $b_{i,j} = \omega_{i,i}\hat{\sigma}_{i,i,\epsilon} + \omega_{j,j}\hat{\sigma}_{j,j,\epsilon} - 1$, where $(\hat{\sigma}_{i,j,\epsilon})$ is the sample covariance matrix of the residual vectors $\boldsymbol{\epsilon}_k = (\epsilon_{k,1}, ..., \epsilon_{k,p})^\intercal$ for $k = 1, ..., n$. It is shown in Liu (2013) that

$$\sqrt{\frac{n}{\hat{\tau}_{i,i}\hat{\tau}_{j,j}}}(\hat{\tau}_{i,j} + b_{i,j}\tau_{i,j}) \to N(0, 1 + \omega_{i,j}\tau_{i,j}).$$

In particular, since $\omega_{i,j} = 0$ implies $\tau_{i,j} = 0$, under the null $H_{0,i,j}$, $\sqrt{n}\frac{\hat{\tau}_{i,j}}{\sqrt{\hat{\tau}_{i,i}\hat{\tau}_{j,j}}} \to N(0, 1)$. It is then natural to use

$$T_{i,j} = \sqrt{n} \cdot \frac{\hat{\tau}_{i,j}}{\sqrt{\hat{\tau}_{i,i}\hat{\tau}_{j,j}}}, \quad 1 \le i < j \le p \tag{40}$$

as the test statistics for simultaneously testing the hypotheses in (35). The multiple testing procedure for a Gaussian graphical model is then given by Algorithm 1 with $T_{i,j}$ computed from (40). It is shown in Liu (2013) that the procedure controls both the FDP and FDR.

3.4. Multiple Testing for Differential Networks

Section 2.5 considered the problem of testing the global null hypothesis $H_0 : \boldsymbol{\Delta} \equiv \boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2 = 0$. If the global null $H_0 : \boldsymbol{\Delta} = 0$ is rejected, it is often of significant interest to investigate further the local structures of the differential network $\boldsymbol{\Delta} = (\delta_{i,j})$ with $\delta_{i,j} = \omega_{i,j,1} - \omega_{i,j,2}$. A natural approach is to carry out multiple testing on the entries of $\boldsymbol{\Delta}$ with FDP and FDR control. Motivated by the problem of identifying gene-by-gene interactions associated with a binary trait, Xia et al. (2015) considered simultaneous testing of $(p^2 - p)/2$ hypotheses

$$H_{0,i,j} : \delta_{i,j} = 0 \quad \text{versus} \quad H_{1,i,j} : \delta_{i,j} \neq 0, \quad 1 \le i < j \le p. \tag{41}$$

We will follow the same notation and use the results developed in Section 2.5.

The test statistic for a given null hypothesis $H_{0,i,j}$ has in fact already been constructed in Section 2.5 where the goal was testing the global hypothesis $H_0 : \boldsymbol{\Delta} = 0$ for a differential network. Let the test statistics $T_{i,j}$ be defined as in (18). The null distribution of $T_{i,j}$ is more complicated than in the case of testing correlations in Section 3.2 or a single Gaussian graphical model in Section 3.3. Under the null $H_{0,i,j}$, $\omega_{i,j,1} = \omega_{i,j,2}$, the distribution of $T_{i,j}$ depends on the common value $\omega_{i,j,1}$and $\omega_{i,j,2}$. When the common value is large, the distribution of $T_{i,j}$ can be far away from the standard normal distribution. Xia et al. (2015) introduced a sparsity condition on the individual precision matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ to ensure that, for "most" of the entries $\delta_{i,j}$, the asymptotic null distribution of the test statistic $T_{i,j}$ under $H_{0,i,j}$ is standard normal,

$$T_{i,j} \to N(0, 1).$$

Due to this complication, the technical analysis is more challenging than multiple testing problems discussed in the earlier sections.

The multiple testing procedure for a differential network is then given by Algorithm 1 with $T_{i,j}$ computed from (18). It can be shown that, under regularity conditions, this multiple testing procedure controls the FDP and FDR at the pre-specified level α asymptotically. See Xia et al. (2015) for the detailed technical analysis.

4. Discussion and Future Issues

In this expository paper, we discussed a collection of recently developed methods for both global testing and large-scale multiple testing for high-dimensional covariance structures. These testing problems are technically challenging due to the intrinsic heteroscedasticity and dependency as well as multiplicity. This is only a selective survey of recent results on the topic. For reasons of space, we have to exclude the discussion on a number of related testing problems. In this section, we discuss first a related detection problem and then a few interesting open problems.

Berthet and Rigollet (2013) studied a rank-one detection problem which can be stated as testing

$$H_0 : \mathbf{\Sigma} = \mathbf{I}, \quad \text{versus} \quad H_1 : \mathbf{\Sigma} = \mathbf{I} + \lambda \mathbf{v}\mathbf{v}', \quad \mathbf{v} \in \mathbb{U}_k^p, \tag{42}$$

where $\mathbb{U}_k^p$ denotes the set of k-sparse unit vectors in $\mathbb{R}^p$. This problem of detecting low-rank structure in the covariance matrix is related to the global testing problems considered in Section 2.1. Berthet and Rigollet (2013) obtained a detection boundary and showed that it is optimal when $k \ll \sqrt{p}$. Cai et al. (2015) considered more general rank r detection and the results resolved a gap between the lower and upper bounds in Berthet and Rigollet (2013) in the rank-one case when $k \gtrsim \sqrt{p}$. A particularly interesting result established in Berthet and Rigollet (2013) is that there is a fundamental gap between what is statistically possible and what is computational feasible. More precisely, it was shown that there is a region for the signal strength λ over which it is statistically possible to detect the signal with high probability but that there is no computationally efficient algorithm (randomized polynomial-time algorithm) to do so, assuming the commonly-believed Planted Clique Hypothesis. This phenomenon has since been observed with rigorous justifications in a number of other high-dimensional statistical inference problems that have combinatorial structures. It would be interesting to consider the interplay between statistical accuracy and computational efficiency for some of the testing problems discussed in this paper, under certain sparsity constraints.

Although much recent progress has been made on testing for high-dimensional covariance structures, there are still many open problems. We conclude the paper with a brief discussion on testing for submatrices of a large covariance or precision matrix and hypothesis testing with incomplete data. These are interesting problems for future research.

A common feature of the three large-scale multiple testing problems discussed in Section 3 is that the objects of direct interest are the individual entries of the corresponding matrices. In some important applications in genomics and other fields, the objects of direct interest are submatrices of a large covariance/correlation/precision matrix. Such is the case in studying the between pathway interactions in genomics where each pathway consists of a group of genes. Another example is studying the relationships between a large number of disease phenotypes and some candidate genomic markers in PheWAS. Making simultaneous inference for a large number of submatrices of a high-dimensional covariance/correlation/precision matrix is technically difficult due to the complex dependence structures. It is challenging to construct test statistics for individual hypotheses and establish their null distribution as well as develop a multiple testing procedure that provides accurate FDR or FDP control.

Missing data is a common problem in many applications and statistical inference with incomplete data has been well studied in classical statistics. The problem of missing data also occurs frequently in high-dimensional settings, e.g., in GWAS and PheWAS. Discarding samples with any missingness is clearly inefficient and possibly induces bias due to non-random missingness. It is of significant interest to develop global and multiple testing methods for high-dimensional covariance structures in the presence of missing data.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The research was supported in part by NSF Grants DMS-1208982 and DMS-1403708, and NIH Grant R01 CA127334.

LITERATURE CITED

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.). Wiley.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Bai, Z., D. Jiang, J.-F. Yao, and S. Zheng (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics 37*(6B), 3822–3840.

Baik, J., G. Ben Arous, and S. Péché (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability 33*(5), 1643–1697.

Bandyopadhyay, S., M. Mehta, D. Kuo, M.-K. Sung, R. Chuang, E. J. Jaehnig, B. Bodenmiller, K. Licon, W. Copeland, M. Shales, D. Fiedler, J. Dutkowski, A. Guénolé, H. van Attikum, K. M. Shokat, R. D. Kolodner, W.-K. Huh, R. Aebersold, M.-C. Keogh, N. J. Krogan, and T. Ideker (2010). Rewiring of genetic networks in response to dna damage. *Sci. Signal. 330*, 1385–1389.

Banerjee, O., L. El Ghaoui, and A. d'Aspremont (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *The Journal of Machine Learning Research 9*, 485–516.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*, 289–300.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics 29*, 1165–1188.

Berthet, Q. and P. Rigollet (2013). Optimal detection of sparse principal components in high dimension. *The Annals of Statistics 41*(4), 1780–1815.

Bickel, P. J. and E. Levina (2008a). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*(1), 199–227.

Bickel, P. J. and E. Levina (2008b). Covariance regularization by thresholding. *The Annals of Statistics 36*(6), 2577–2604.

Birke, M. and H. Dette (2005). A note on testing the covariance matrix for large dimension. *Statistics & Probability Letters 74*(3), 281–289.

Cai, T. T. and T. Jiang (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics 39*(3), 1496–1525.

Cai, T. T. and T. Jiang (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *Journal of Multivariate Analysis 107*, 24–39.

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

Cai, T. T. and W. Liu (2015). Large-scale multiple testing of correlations. *Journal of the American Statistical Association 110*(to appear).

Cai, T. T., W. Liu, and X. Luo (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association 106*(494), 594–607.

Cai, T. T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association 108*(501), 265–277.

Cai, T. T., W. Liu, and H. H. Zhou (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics 44*, 455–488.

Cai, T. T. and Z. Ma (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli 19*(5B), 2359–2388.

Cai, T. T., Z. Ma, and Y. Wu (2013). Sparse PCA: Optimal rates and adaptive estimation. *The Annals of Statistics 41*(6), 3074–3110.

Cai, T. T., Z. Ma, and Y. Wu (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability Theory and Related Fields 161*(3-4), 781–815.

Cai, T. T., Z. Ren, and H. H. Zhou (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probability Theory and Related Fields 156*(1-2), 101–143.

Cai, T. T., Z. Ren, and H. H. Zhou (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation (with discussion). *Electronic Journal of Statistics 10*, 1–59.

Cai, T. T. and M. Yuan (2012). Adaptive covariance matrix estimation through block thresholding. *The Annals of Statistics 40*(4), 2014–2042.

Cai, T. T. and A. Zhang (2016). Inference on high-dimensional differential correlation matrices. *Journal Multivariate Analysis 143*, 107–126.

Cai, T. T., C. H. Zhang, and H. H. Zhou (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics 38*(4), 2118–2144.

Cai, T. T. and H. H. Zhou (2012). Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics 40*(5), 2389–2420.

Carter, S. L., C. M. Brechbühler, M. Griffin, and A. T. Bond (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics 20*(14), 2242–2250.

Chandrasekaran, V., P. A. Parrilo, and A. S. Willsky (2012). Latent variable graphical model selection via convex optimization. *The Annals of Statistics 40*(4), 1935–1967.

Chen, S. X., L. X. Zhang, and P. S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

Danaher, P., P. Wang, and D. M. Witten (2014). 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.

d'Aspremont, A., O. Banerjee, and L. El Ghaoui (2008). First-order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Applications 30*(1), 56–66.

de la Fuente, A. (2010). From "differential expression" to "differential networking" – identification of dysfunctional regulatory networks in diseases. *Trends Genet. 26*(7), 326–333.

Dubois, P. C., G. Trynka, L. Franke, K. A. Hunt, J. Romanos, A. Curtotti, A. Zhernakova, G. A. Heap, R. Ádány, A. Aromaa, et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. *Nature genetics 42*(4), 295–302.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistics Association 99*, 96–104.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association 102*, 93–103.

El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability 35*(2), 663–714.

El Karoui, N. (2008). Operator norm consistent estimation of large dimensional sparse covariance matrices. *The Annals of Statistics 36*(6), 2717–2756.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics 39*(6), 3320–3356.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scandinavian Journal of Statistics 34*(2), 275–297.

Fukushima, A. (2013). Diffcorr: an r package to analyze and visualize differential correlations in biological networks. *Gene 518*(1), 209–214.

Fuller, T. F., A. Ghazalpour, J. E. Aten, T. A. Drake, A. J. Lusis, and S. Horvath (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. *Mammalian Genome 18*(6-7), 463–472.

Furrer, R. and T. Bengtsson (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis 98*(2), 227–255.

Genovese, C. and L. Wasserman (2004). A stochastic process approach to false discovery control. *The Annals of Statistics 32*, 1035–1061.

Gupta, A. and J. Tang (1984). Distribution of likelihood ratio statistic for testing equality of covariance matrices of multivariate gaussian models. *Biometrika 71*(3), 555–559.

Gupta, S. D. and N. Giri (1973). Properties of tests concerning covariance matrices of normal distributions. *The Annals of Statistics 1*(6), 1222–1224.

Hirai, M. Y., K. Sugiyama, Y. Sawada, T. Tohge, T. Obayashi, A. Suzuki, R. Araki, N. Sakurai, H. Suzuki, K. Aoki, et al. (2007). Omics-based identification of arabidopsis myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proceedings of the National Academy of Sciences 104*(15), 6478–6483.

Huang, J. Z., N. Liu, M. Pourahmadi, and L. Liu (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika 93*(1), 85–98.

Hudson, N. J., A. Reverter, and B. P. Dalrymple (2009). A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput. Biol. 5*(5), e1000382.

Ideker, T. and N. Krogan (2012). Differential network biology. *Mol. Syst. Biol. 8*(1), 565.

Jiang, D., T. Jiang, and F. Yang (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference 142*(8), 2241–2256.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability 14*(2), 865–880.

John, S. (1971). Some optimal multivariate tests. *Biometrika 58*(1), 123–127.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *The Annals of Statistics 29*(2), 295–327.

Johnstone, I. M. and A. Y. Lu (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association 104*(486), 682–693.

Lam, C. and J. Fan (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *The Annals of Statistics 37*(6B), 4254–4278.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford: Oxford University Press.

Ledoit, O. and M. Wolf (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics 30*(4), 1081–1102.

Lee, H. K., A. K. Hsu, J. Sajdak, J. Qin, and P. Pavlidis (2004). Coexpression analysis of human genes across many microarray data sets. *Genome research 14*(6), 1085–1094.

Li, J. and S. X. Chen (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics 40*(2), 908–940.

Li, K.-C., A. Palotie, S. Yuan, D. Bronnikov, D. Chen, X. Wei, O.-W. Choi, J. Saarela, and L. Peltonen (2007). Finding disease candidate genes by liquid association. *Genome Biol. 8*(10), R205.

Ligeralde, A. V. and B. W. Brown (1995). Band covariance matrix estimation using restricted residuals: A monte carlo analysis. *International Economic Review 36*, 751–767.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics 41*(6), 2948–2978.

Liu, W. and Q.-M. Shao (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *The Annals of Statistics 42*(5), 2003–2025.

Meinshausen, N. and P. Bühlmann (2006). High dimensional graphs and variable selection with the Lasso. *The Annals of Statistics 34*(3), 1436–1462.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Nagao, H. (1973). On some test criteria for covariance matrix. *The Annals of Statistics 1*(4), 700–709.

Onatski, A., M. Moreira, and M. Hallin (2013). Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics 41*(3), 1204–1231.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica 17*(4), 1617.

Peche, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields 143*(3-4), 481–516.

Perlman, M. D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *The Annals of Statistics 8*, 247–263.

Qiu, X., L. Klebanov, and A. Yakovlev (2005). Correlation between gene expression levels and limitations of the empirical bayes methodology for finding differentially expressed genes. *Statistical Applications in Genetics and Molecular Biology 4*(1).

Qiu, Y. and S. X. Chen (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics 40*(3), 1285–1314.

Raizada, R. D., T. L. Richards, A. Meltzoff, and P. K. Kuhl (2008). Socioeconomic status predicts hemispheric specialisation of the left inferior frontal gyrus in young children. *Neuroimage 40*(3), 1392–1401.

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

Ren, Z., T. Sun, C.-H. Zhang, and H. H. Zhou (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical model. *The Annals of Statistics 43*(3), 991–1026.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics 2*, 494–515.

Rothman, A. J., E. Levina, and J. Zhu (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association 104*(485), 177–186.

Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics & Data Analysis 51*(12), 6535–6542.

Shao, Q.-M. and W.-X. Zhou (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *The Annals of Probability 42*(2), 623–648.

Shaw, P., D. Greenstein, J. Lerch, L. Clasen, R. Lenroot, N. Gogtay, A. Evans, J. Rapoport, and J. Giedd (2006). Intellectual ability and cortical development in children and adolescents. *Nature 440*(7084), 676–679.

Shedden, K. and J. Taylor (2005). Differential correlation detects complex associations between gene expression and clinical outcomes in lung adenocarcinomas. In *Methods of Microarray Data Analysis*, pp. 121–131. Springer.

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics 108*(5-6), 1033–1056.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society 35*(2), 251–272.

Srivastava, M. S. and H. Yanagihara (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis 101*(6), 1319–1329.

Storey, J., J. Taylor, and D. Siegmund (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B 66*, 187–205.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* ***64***, 479–498.

Sugiura, N. and H. Nagao (1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. *The Annals of Mathematical Statistics 39*(5), 1686–1692.

Sun, T. and C.-H. Zhang (2013). Sparse matrix inversion with scaled Lasso. *The Journal of Machine Learning Research 14*(1), 3385–3418.

Sun, W. and T. T. Cai (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association 102*, 901–912.

Sun, W. and T. T. Cai (2009). Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society: Series B 71*, 393–424.

Sun, W., B. J. Reich, T. T. Cai, M. Guindani, and A. Schwartzman (2015). False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B 77*, 59–83.

Vu, V. Q. and J. Lei (2013). Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics 41*(6), 2905–2947.

Wu, W. B. (2008). On false discovery control under dependence. *The Annals of statistics*, 364–380.

Wu, W. B. and M. Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica 19*(4), 1755–1768.

Xia, Y., T. Cai, and T. T. Cai (2015). Testing differential networks with applications to detecting gene-by-gene interactions. *Biometrika 102*, 247–266.

Xiao, H. and W. B. Wu (2011). Simultaneous inference of covariances. *arXiv preprint arXiv:1109.0524*.

Yuan, M. (2010). Sparse inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research 11*, 2261–2286.

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

Zhang, J., J. Li, and H. Deng (2008). Class-specific correlations of gene expressions: identification and their effects on clustering analyses. *The American Journal of Human Genetics 83*(2), 269–277.

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

Zheng, S., Z. Bai, and J. Yao (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics 43*(2), 546–591.

Zhu, D., A. O. Hero, Z. S. Qin, and A. Swaroop (2005). High throughput screening of co-expressed gene pairs with controlled false discovery rate (fdr) and minimum acceptable strength (mas). *Journal of Computational Biology 12*(7), 1029–1045.

Zou, H., T. Hastie, and R. Tibshirani (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics 15*(2), 265–286.