2011

Limiting Laws of Coherence of Random Matrices With Applications to Testing Covariance Structure and Construction of Compressed Sensing Matrices

T. Tony Cai
University of Pennsylvania

Tiefeng Jiang
University of Minnesota

Recommended Citation

Cai, T., & Jiang, T. (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. http://dx.doi.org/10.1214/11-AOS879

Limiting Laws of Coherence of Random Matrices With Applications to Testing Covariance Structure and Construction of Compressed Sensing Matrices

Abstract
Testing covariance structure is of significant interest in many areas of statistical analysis and construction of compressed sensing matrices is an important problem in signal processing. Motivated by these applications, we study in this paper the limiting laws of the coherence of an $n \times p$ random matrix in the high-dimensional setting where p can be much larger than n. Both the law of large numbers and the limiting distribution are derived. We then consider testing the bandedness of the covariance matrix of a high-dimensional Gaussian distribution which includes testing for independence as a special case. The limiting laws of the coherence of the data matrix play a critical role in the construction of the test. We also apply the asymptotic results to the construction of compressed sensing matrices.

Keywords
Chen–Stein method, coherence, compressed sensing matrix, covariance structure, law of large numbers, limiting distribution, maxima, moderate deviations, mutual incoherence property, random matrix, sample correlation matrix

Disciplines
Statistics and Probability

The Annals of Statistics
2011, Vol. 39, No. 3, 1496–1525
DOI: 10.1214/11-AOS879

LIMITING LAWS OF COHERENCE OF RANDOM MATRICES WITH APPLICATIONS TO TESTING COVARIANCE STRUCTURE AND CONSTRUCTION OF COMPRESSED SENSING MATRICES

BY T. TONY CAI[1] AND TIEFENG JIANG[2]

University of Pennsylvania and University of Minnesota

Testing covariance structure is of significant interest in many areas of statistical analysis and construction of compressed sensing matrices is an important problem in signal processing. Motivated by these applications, we study in this paper the limiting laws of the coherence of an $n \times p$ random matrix in the high-dimensional setting where p can be much larger than n. Both the law of large numbers and the limiting distribution are derived. We then consider testing the bandedness of the covariance matrix of a high-dimensional Gaussian distribution which includes testing for independence as a special case. The limiting laws of the coherence of the data matrix play a critical role in the construction of the test. We also apply the asymptotic results to the construction of compressed sensing matrices.

1. Introduction. Random matrix theory has been proved to be a powerful tool in a wide range of fields including statistics, high-energy physics, electrical engineering and number theory. Traditionally the primary focus is on the spectral analysis of eigenvalues and eigenvectors. See, for example, Johnstone (2001, 2008), Jiang (2004b) and Bai, Miao and Pan (2007). For general background on random matrix theory, see Bai and Silverstein (2010) and Anderson, Guionnet and Zeitouni (2009).

In statistics, random matrix theory is particularly useful for inference of high-dimensional data which is becoming increasingly available in many areas of scientific investigations. In these applications, the dimension p can be much larger than the sample size n. In such a setting, classical statistical methods and results based on fixed p and large n are no longer applicable. Examples include high-dimensional regression, hypothesis testing concerning high-dimensional parameters and inference on large covariance matrices. See, for example, Bai and Saranadasa (1996), Candes and Tao (2007), Bai et al. (2009), Cai, Wang and Xu (2010a) and Cai, Zhang and Zhou (2010).

Received June 2010; revised January 2011.
[1]Supported in part by NSF FRG Grant DMS-08-54973.
[2]Supported in part by NSF FRG Grant DMS-04-49365.
MSC2010 subject classifications. Primary 62H12, 60F05; secondary 60F15, 62H10.
Key words and phrases. Chen–Stein method, coherence, compressed sensing matrix, covariance structure, law of large numbers, limiting distribution, maxima, moderate deviations, mutual incoherence property, random matrix, sample correlation matrix.

In the present paper, we study the limiting laws of the coherence of an $n \times p$ random matrix, which is defined to be the largest magnitude of the off-diagonal entries of the sample correlation matrix generated from the $n \times p$ random matrix. We are especially interested in the case where $p \gg n$. This is a problem of independent interest. Moreover, we are particularly interested in the applications of the results to testing the covariance structure of a high-dimensional Gaussian variable and the construction of compressed sensing matrices. These three problems are important in their respective fields, one in random matrix theory, one in statistics and one in signal processing. The latter two problems are seemingly unrelated at first sight, but as we shall see later they can both be attacked through the use of the limiting laws of the coherence of random matrices.

1.1. *Limiting laws of the coherence of a random matrix.* Let $X_n = (x_{ij})$ be an $n \times p$ random matrix where the entries x_{ij} are i.i.d. real random variables with mean μ and variance $\sigma^2 > 0$. Let $x_1, x_2, \dots, x_p$ be the p columns of X_n. The sample correlation matrix Γ_n is defined by $\Gamma_n := (\rho_{ij})$ with

$$\rho_{ij} = \frac{(x_i - \bar{x}_i)^T (x_j - \bar{x}_j)}{\|x_i - \bar{x}_i\| \cdot \|x_j - \bar{x}_j\|}, \qquad 1 \le i, j \le p, \tag{1}$$

where $\bar{x}_k = (1/n) \sum_{i=1}^n x_{ik}$ and $\| \cdot \|$ is the usual Euclidean norm in $\mathbb{R}^n$. Here we write $x_i - \bar{x}_i$ for $x_i - \bar{x}_i e$, where $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$. In certain applications such as construction of compressed sensing matrices, the mean μ of the random entries x_{ij} is known (typically $\mu = 0$) and the sample correlation matrix is then defined to be $\tilde{\Gamma}_n := (\tilde{\rho}_{ij})$ with

$$\tilde{\rho}_{ij} = \frac{(x_i - \mu)^T (x_j - \mu)}{\|x_i - \mu\| \cdot \|x_j - \mu\|}, \qquad 1 \le i, j \le p. \tag{2}$$

One of the main objects of interest in the present paper is the largest magnitude of the off-diagonal entries of the sample correlation matrix

$$L_n = \max_{1 \le i < j \le p} |\rho_{ij}| \quad \text{and} \quad \tilde{L}_n = \max_{1 \le i < j \le p} |\tilde{\rho}_{ij}|. \tag{3}$$

In the compressed sensing literature, the quantity $\tilde{L}_n$ is called the *coherence* of the matrix X_n. A matrix is incoherent when $\tilde{L}_n$ is small. See, for example, Donoho, Elad and Temlyakov (2006). With slight abuse of terminology, in this paper we shall call both L_n and $\tilde{L}_n$ *coherence* of the random matrix X_n, the former for the case μ is unknown and the latter for the case μ is known. The first goal of the present paper is to derive the limiting laws of the coherence in the high-dimensional setting.

In the case where p and n are comparable, that is, $n/p \to \gamma \in (0, \infty)$, asymptotic properties of the coherence L_n of the random matrix X_n have been considered by Jiang (2004a), Zhou (2007), Liu, Lin and Shao (2008) and Li, Liu and Rosalsky (2009). In this paper, we focus on the high-dimensional case where p can be as large as e^{n^β} for some $0 < \beta < 1$. This is a case of special interest for the applications considered later.

The results given in Section 2 show that under regularity conditions

$$\sqrt{\frac{n}{\log p}} L_n \xrightarrow{P} 2 \qquad \text{as } n \to \infty,$$

where $\xrightarrow{P}$ denotes convergence in probability. Here and throughout the paper, the log is the natural logarithm $\log_e$. Furthermore, it is shown that $nL_n^2 - 4\log p + \log\log p$ converges weakly to an extreme distribution of type I with distribution function

$$F(y) = e^{-(1/\sqrt{8\pi})e^{-y/2}}, \qquad y \in \mathbb{R}.$$

Same results hold for $\tilde{L}_n$. In contrast to the known results in the literature, here the dimension p can be much larger than n. In the special cases where x_{ij} are either bounded or normally distributed, the results hold as long as $\log p = o(n^{1/3})$.

In addition, motivated by application to testing covariance structure, we also consider the case where the entries of random matrix X_n are correlated. More specifically, let $X_n = (x_{ij})_{1\le i\le n, 1\le j\le p}$, where the n rows are i.i.d. random vectors with distribution $N_p(\mu, \Sigma)$. For a given integer $\tau \ge 1$ (which can depend on n or p), it is of interest in applications to test the hypothesis that the covariance matrix Σ is banded, that is,

$$H_0 : \sigma_{ij} = 0 \qquad \text{for all } |i-j| \ge \tau. \tag{4}$$

Analogous to the definition of L_n and $\tilde{L}_n$, we define

$$L_{n,\tau} = \max_{|i-j|\ge\tau} |\rho_{ij}|, \tag{5}$$

when the mean μ is assumed to be unknown and define

$$\tilde{L}_{n,\tau} = \max_{|i-j|\ge\tau} |\tilde{\rho}_{ij}|, \tag{6}$$

when the mean $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ is assumed to be known. In the latter case, $\tilde{\rho}_{i,j}$ is defined to be

$$\tilde{\rho}_{ij} = \frac{(x_i - \mu_i)^T(x_j - \mu_j)}{\|x_i - \mu_i\| \cdot \|x_j - \mu_j\|}, \qquad 1 \le i, j \le p, \tag{7}$$

where $X_n = (x_1, \dots, x_p)$. We shall call $L_{n,\tau}$ and $\tilde{L}_{n,\tau}$ the τ-*coherence* of the matrix X_n. In Section 2, the limiting distributions of $L_{n,\tau}$ and $\tilde{L}_{n,\tau}$ under the null hypothesis H_0 are derived, and their applications are discussed in Section 3. The study for this case is considerably more difficult than that for the i.i.d. case.

1.2. *Testing covariance structure.* Covariance matrices play a critical role in many areas of statistical inference. Important examples include principal component analysis, regression analysis, linear and quadratic discriminant analysis, and graphical models. In the classical setting of low dimension and large sample size, many methods have been developed for estimating covariance matrices as well as testing specific patterns of covariance matrices. In particular, testing for independence in the Gaussian case is of special interest because many statistical procedures are built upon the assumptions of independence and normality of the observations.

To be more specific, suppose we observe independent and identically distributed p-variate random variables $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ with mean $\mu = \mu_{p\times 1}$, covariance matrix $\Sigma = \Sigma_{p\times p}$ and correlation matrix $R = R_{p\times p}$. In the setting where the dimension p and the sample size n are comparable, that is, $n/p \to \gamma \in (0, \infty)$, testing of the hypotheses $H_0 : \Sigma = I$ versus $H_a : \Sigma \neq I$, assuming $\mu = 0$, has been considered by Johnstone (2001) in the Gaussian case and by Péché (2009) in the more general case where the distribution is assumed to be sub-Gaussian and where the ratio p/n can converge to either a positive number γ, 0 or ∞. The test statistic is based on the largest eigenvalue of the sample covariance matrix and relies on the important results in their papers that the largest eigenvalue of the sample covariance matrix follows the Tracy–Widom distribution asymptotically.

The hypothesis $H_0 : \Sigma = I$ is too restrictive for many applications. An arguably more practically important problem is testing for independence in the Gaussian case. That is, one wishes to test the hypothesis $H_0 : \Sigma$ is diagonal against the hypothesis $H_a : \Sigma$ is not diagonal, or equivalently in terms of the correlation matrix R, one wishes to test $H_0 : R = I$ versus $H_a : R \neq I$. Tests based on the largest eigenvalue of the sample covariance matrix cannot be easily modified for testing these hypotheses.

In this paper, we consider testing more general hypotheses on the covariance structure of a high-dimensional Gaussian distribution which includes testing for independence as a special case. More specifically, we consider testing the hypothesis that Σ is banded with a given bandwidth τ (which may depend on n or p), that is, the variables have nonzero correlations only up to lag τ. In other words, for a given integer $\tau \geq 1$, we wish to test the hypothesis $H_0 : \sigma_{i,j} = 0$ for all $|i - j| \geq \tau$. This problem arises, for example, in econometrics when testing certain economic theories and in time series analysis. See Andrews (1991), Ligeralde and Brown (1995) and references therein. The special case of $\tau = 1$ corresponds to testing for independence. We shall show that the limiting laws of the τ-coherence $L_{n,\tau}$ developed here can be applied to construct a convenient test for the bandedness of the covariance matrix. In the special case of $\tau = 1$, the limiting laws of the coherence of the data matrix $\mathbf{Y}$ play a critical role in the construction of the test.

1.3. *Construction of compressed sensing matrices.* In addition to testing the covariance structure, another important application of our results on the limiting laws of the coherence of a random matrix is to the construction of compressed sensing matrices. Compressed sensing is a fast developing field which provides a novel and efficient data acquisition technique that enables accurate reconstruction of highly undersampled sparse signals. See, for example, Donoho (2006a). It has a wide range of applications including signal processing, medical imaging, and seismology. In addition, the development of the compressed sensing theory also provides crucial insights into high-dimensional regression in statistics. See, for example, Candes and Tao (2007), Bickel, Ritov and Tsybakov (2009) and Candès and Plan (2009).

One of the main goals of compressed sensing is to construct measurement matrices $X_{n\times p}$, with the number of measurements n as small as possible relative to p, such that for any k-sparse signal $\beta \in \mathbb{R}^p$, one can recover β exactly from linear measurements $y = X\beta$ using a computationally efficient recovery algorithm. In compressed sensing it is typical that $p \gg n$, for example, p can be of order e^{n^β} for some $0 < \beta < 1$. In fact, the goal is often to make p as large as possible relative to n. It is now well understood that the method of ℓ_1 minimization provides an effective way for reconstructing a sparse signal in many settings. In order for a recovery algorithm such as ℓ_1 minimization to work well, the measurement matrices $X_{n\times p}$ must satisfy certain conditions. Two commonly used conditions are the so-called restricted isometry property (RIP) and *mutual incoherence property* (MIP). Roughly speaking, the RIP requires subsets of certain cardinality of the columns of X to be close to an orthonormal system and the MIP requires the pairwise correlations among the column vectors of X to be small. See Candes and Tao (2005), Donoho, Elad and Temlyakov (2006) and Cai, Wang and Xu (2010a, 2010b). It is well known that construction of large deterministic measurement matrices that satisfy either the RIP or MIP is difficult. Instead, random matrices are commonly used. Matrices generated by certain random processes have been shown to satisfy the RIP conditions with high probability. See, for example, Baraniuk et al. (2008). A major technical tool used there is the Johnson–Lindenstrauss lemma. Here we focus on the MIP.

The MIP condition can be easily explained. It was first shown by Donoho and Huo (2001), in the setting where X is a concatenation of two square orthogonal matrices, that the condition

$$(2k-1)\tilde{L}_n < 1 \tag{8}$$

ensures the exact recovery of β when β has at most k nonzero entries (such a signal is called k-sparse). This result was then extended by Fuchs (2004) to general matrices. Cai, Wang and Xu (2010b) showed that condition (8) is also sufficient for stable recovery of sparse signal in the noisy case where y is measured with error. In addition, it was shown that this condition is sharp in the sense that there exist matrices X such that it is not possible to recover certain k-sparse signals β based on $y = X\beta$ when $(2k-1)\tilde{L}_n = 1$.

The mutual incoherence property (8) is very desirable. When it is satisfied by the measurement matrix X, the estimator obtained through ℓ_1 minimization satisfies the near-optimality properties and the oracle inequalities. In addition, the technical analysis is particularly simple. See, for example, Cai, Wang and Xu (2010b). Except results on the magnitude and the limiting distribution of $\tilde{L}_n$ when the underlying matrix is Haar-invariant and orthogonal by Jiang (2005), it is, however, unknown in general how likely a random matrix satisfies the MIP (8) in the high-dimensional setting where p can be as large as e^{n^β}. We shall show in Section 4 that the limiting laws of the coherence derived here can readily be applied to compute the probability that random measurement matrices satisfy the MIP condition (8).

1.4. *Organization of the paper.* The rest of the paper is organized as follows. We begin in Section 2 by studying the limiting laws of the coherence of a random matrix in the high-dimensional setting. Section 3 considers the problems of testing for independence and bandedness in the Gaussian case. The test statistic is based on the coherence of the data matrix and the construction of the tests relies heavily on the asymptotic results developed in Section 2. Application to the construction of compressed sensing matrices is considered in Section 4. Section 5 discusses connections and differences with other related work. The main results are proved in Section 6 and the proofs of technical lemmas are given in Cai and Jiang (2010).

2. Limiting laws of coherence of random matrices. In this section, we consider the limiting laws of the coherence of a random matrix with i.i.d. entries. In addition, we also consider the case where each row of the random matrix is drawn independently from a multivariate Gaussian distribution with banded covariance matrix. In the latter case, the limiting distribution of $L_{n,\tau}$ and $\tilde{L}_{n,\tau}$ defined in (5) and (6) is considered. The asymptotic results are then applied to the testing of the covariance structure in Section 3 and the construction of compressed sensing matrices in Section 4.

2.1. *The i.i.d. case.* We begin by considering the case for independence where all entries of the random matrix are independent and identically distributed. Suppose $\{\xi, x_{ij}, i, j = 1, 2, \ldots\}$ are i.i.d. real random variables with mean μ and variance $\sigma^2 > 0$. Let $X_n = (x_{ij})_{1\le i\le n, 1\le j\le p}$ and let $x_1, x_2, \ldots, x_p$ be the p columns of X_n. Then $X_n = (x_1, x_2, \ldots, x_p)$. Let $\bar{x}_k = (1/n)\sum_{i=1}^n x_{ik}$ be the sample average of x_k. We write $x_i - \bar{x}_i$ for $x_i - \bar{x}_i e$, where $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$. Define the Pearson correlation coefficient ρ_{ij} between x_i and x_j as in (1). Then the *sample correlation matrix* generated by X_n is $\Gamma_n := (\rho_{ij})$, which is a p by p symmetric matrix with diagonal entries $\rho_{ii} = 1$ for all $1 \le i \le p$. When the mean μ of the random variables x_{ij} is assumed to be known, we define the sample correlation matrix by $\tilde{\Gamma}_n := (\tilde{\rho}_{ij})$ with $\tilde{\rho}_{ij}$ given as in (2).

In this section, we are interested in the limiting laws of the coherence L_n and $\tilde{L}_n$ of the random matrix X_n, which are defined to be the largest magnitude of the off-diagonal entries of sample correlation matrices Γ_n and $\tilde{\Gamma}_n$, respectively; see (3). The case of $p \gg n$ is of particular interest to us. In such a setting, some simulation studies about the distribution of L_n were made in Cai and Lv (2007), Fan and Lv (2008, 2010). We now derive the limiting laws of L_n and $\tilde{L}_n$. We shall first introduce another quantity that is useful for our technical analysis. Define

$$J_n = \max_{1 \le i < j \le p} \frac{|(x_i - \mu)^T (x_j - \mu)|}{\sigma^2}. \tag{9}$$

We first state the law of large numbers for L_n for the case where the random entries x_{ij} are bounded.

THEOREM 1. *Assume $|x_{11}| \le C$ for a constant $C > 0$, $p = p(n) \to \infty$ and $\log p = o(n)$ as $n \to \infty$. Then $\sqrt{n/\log p}\, L_n \to 2$ in probability as $n \to \infty$.*

We now consider the case where x_{ij} have finite exponential moments.

THEOREM 2. *Suppose $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $\alpha > 0$ and $t_0 > 0$. Set $\beta = \alpha/(4 + \alpha)$. Assume $p = p(n) \to \infty$ and $\log p = o(n^\beta)$ as $n \to \infty$. Then $\sqrt{n/\log p}\, L_n \to 2$ in probability as $n \to \infty$.*

Comparing Theorems 1 and 2, it can be seen that a stronger moment condition gives a higher order of p to make the law of large numbers for L_n valid. Also, based on Theorem 2, if $Ee^{|x_{11}|^\alpha} < \infty$ for any $\alpha > 0$, then $\beta \to 1$, hence the order $o(n^\beta)$ is close to $o(n)$, which is the order in Theorem 1. We now consider the limiting distribution of L_n after suitable normalization.

THEOREM 3. *Suppose $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $0 < \alpha \le 2$ and $t_0 > 0$. Set $\beta = \alpha/(4 + \alpha)$. Assume $p = p(n) \to \infty$ and $\log p = o(n^\beta)$ as $n \to \infty$. Then $nL_n^2 - 4\log p + \log\log p$ converges weakly to an extreme distribution of type* I *with distribution function*

$$F(y) = e^{-(1/\sqrt{8\pi})e^{-y/2}}, \qquad y \in \mathbb{R}.$$

REMARK 2.1. Propositions 6.1, 6.2 and 6.3 show that the above three theorems are still valid if L_n is replaced by either $\tilde{L}_n$ or J_n/n, where $\tilde{L}_n$ is as in (3) and J_n is as in (9).

In the case where n and p are comparable, that is, $n/p \to \gamma \in (0, \infty)$, Jiang (2004a) obtained the strong laws and asymptotic distributions of the coherence L_n of random matrices. Several authors improved the results by sharpening the moment assumptions; see, for example, Li and Rosalsky (2006), Zhou (2007) and Li, Liu and Rosalsky (2009) where the same condition $n/p \to \gamma \in (0, \infty)$ was imposed. Liu, Lin and Shao (2008) showed that the same results hold for $p \to \infty$ and $p = O(n^\alpha)$ where α is a constant.

In this paper, motivated by the applications mentioned earlier, we are particularly interested in the case where both n and p are large and $p = o(e^{n^\beta})$ while the entries of X_n are i.i.d. with a certain moment condition. We also consider the case where the n rows of X_n form a random sample from $N_p(\mu, \Sigma)$ with Σ being a banded matrix. In particular, the entries of X_n are not necessarily independent. As shown in the above theorems and in Section 2.2 later, when $p \le e^{n^\beta}$ for a certain $\beta > 0$, we obtain the strong laws and limiting distributions of the coherence of the random matrix X_n. Presumably the results on high order $p = o(e^{n^\beta})$ need stronger moment conditions than those for the case $p = O(n^\alpha)$. Ignoring the moment conditions, our results cover those in Liu, Lin and Shao (2008) as well as others aforementioned.

Theorem 1.2 in Jiang (2004a) states that if $n/p \to \gamma \in (0, \infty)$ and $E|\xi|^{30+\varepsilon} < \infty$ for some $\varepsilon > 0$, then for any $y \in \mathbb{R}$,

$$P(nL_n^2 - 4\log n + \log\log n \le y) \to e^{-Ke^{-y/2}}, \tag{10}$$

where $K = (\gamma^2\sqrt{8\pi})^{-1}$, as $n \to \infty$. It is not difficult to see that Theorem 3 implies Theorem 1.2 in Jiang (2004a) under condition that $n/p \to \gamma$ and $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $0 < \alpha \le 2$ and $t_0 > 0$. In fact, write

$$\begin{aligned} &nL_n^2 - 4\log n + \log\log n \\ &\qquad = (nL_n^2 - 4\log p + \log\log p) + 4\log\frac{p}{n} + (\log\log n - \log\log p). \end{aligned}$$

Theorem 3 yields that $nL_n^2 - 4\log p + \log\log p$ converges weakly to $F(y) = e^{-(1/\sqrt{8\pi})e^{-y/2}}$. Note that since $n/p \to \gamma$,

$$4\log\frac{p}{n} \to -4\log\gamma \quad \text{and} \quad \log\log n - \log\log p \to 0.$$

Now it follows from Slutsky's theorem that $nL_n^2 - 4\log n + \log\log n$ converges weakly to $F(y + 4\log\gamma)$, which is exactly (10) from Theorem 1.2 in Jiang (2004a).

2.2. *The dependent case.* We now consider the case where the rows of the random matrix X_n are drawn independently from a multivariate Gaussian distribution. Let $X_n = (x_{ij})_{1\le i\le n, 1\le j\le p}$, where the n rows are i.i.d. random vectors with distribution $N_p(\mu, \Sigma)$, where $\mu \in \mathbb{R}^p$ is arbitrary and Σ does not have to be nonsingular in this section unless otherwise specified. Let $(r_{ij})_{p\times p}$ be the correlation matrix obtained from $\Sigma = (\sigma_{ij})_{p\times p}$. As mentioned in the Introduction, it is of interest to test the hypothesis that the covariance matrix Σ is banded, that is,

$$H_0 : \sigma_{ij} = 0 \qquad \text{for all } |i - j| \ge \tau \tag{11}$$

for a given integer $\tau \geq 1$. In order to construct a test, we study in this section the asymptotic distributions of the τ-*coherence* $L_{n,\tau}$ and $\tilde{L}_{n,\tau}$ defined in (5) and (6), respectively, assuming the covariance matrix Σ has desired banded structure under the null hypothesis. This case is much harder than the i.i.d. case considered in Section 2.1 because of the dependence.

For any given $0 < \delta < 1$, set

$$\Gamma_{p,\delta} = \{1 \leq i \leq p; |r_{ij}| > 1 - \delta \text{ for some } 1 \leq j \leq p \text{ with } j \neq i\}. \tag{12}$$

THEOREM 4. *Suppose, as $n \to \infty$*:

(i) $p = p_n \to \infty$ *with* $\log p = o(n^{1/3})$;
(ii) $\tau = o(p^t)$ *for any* $t > 0$;
(iii) *for some* $\delta \in (0, 1)$, $|\Gamma_{p,\delta}| = o(p)$, *which is particularly true if* $\max_{1 \leq i < j \leq p < \infty} |r_{ij}| \leq 1 - \delta$.

Then, under H_0, $nL_{n,\tau}^2 - 4\log p + \log\log p$ converges weakly to an extreme distribution of type I *with distribution function*

$$F(y) = e^{-(1/\sqrt{8\pi})e^{-y/2}}, \qquad y \in \mathbb{R}.$$

Similar to J_n in (9), we define

$$U_{n,\tau} = \max_{1 \leq i < j \leq p, |i-j| \geq \tau} \frac{|(x_i - \mu_i)^T(x_j - \mu_j)|}{\sigma_i \sigma_j}, \tag{13}$$

where we write $x_i - \mu_i$ for $x_i - \mu_i e$ with $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, $\mu = (\mu_1, \ldots, \mu_p)^T$ and σ_i^2's are diagonal entries of Σ.

REMARK 2.2. From Proposition 6.4, we know Theorem 4 still holds if $L_{n,\tau}$ is replaced with $U_{n,\tau}$ defined in (13). In fact, by the first paragraph in the proof of Theorem 4, to see if Theorem 4 holds for $U_{n,\tau}$, we only need to consider the problem by assuming, w.l.o.g., $\mu = 0$ and σ_i's, the diagonal entries of Σ, are all equal to 1. Thus, by Proposition 6.4, Theorem 4 holds when $L_{n,\tau}$ is replaced by $U_{n,\tau}$.

Theorem 4 implies immediately the following result.

COROLLARY 2.1. *Suppose the conditions in Theorem 4 hold, then*

$$\sqrt{\frac{n}{\log p}} L_{n,\tau} \xrightarrow{P} 2 \qquad \text{as } n \to \infty.$$

The assumptions (ii) and (iii) in Theorem 4 are both essential. If one of them is violated, the conclusion may fail. The following two examples illustrate this point.

REMARK 2.3. Consider $\Sigma = I_p$ with $p = 2n$ and $\tau = n$. So conditions (i) and (iii) in Theorem 4 hold, but (ii) does not. By following the proof of Theorem 3, we have

$$(nL_{n,\tau}^2 - 4\log p + \log\log p) + \log 16 \qquad \text{converges weakly to } F \tag{14}$$

as $n \to \infty$. The difference between (14) and Theorem 4 is evident.

The details of this and the next remark are given in Cai and Jiang (2010).

REMARK 2.4. Let $p = mn$ with integer $m \geq 2$. We consider the $p \times p$ matrix $\Sigma = \mathrm{diag}(H_n, \ldots, H_n)$ where there are m H_n's in the diagonal of Σ and all of the entries of the $n \times n$ matrix H_n are equal to 1. Take $\tau = n$ and $m = [e^{n^{1/4}}]$. Notice $\Gamma_{p,\delta} = p$ for any $\delta > 0$. Since $p = mn$, both (i) and (ii) in Theorem 4 are satisfied, but (iii) does not. It is not very hard to see that

$$(nL_{n,\tau}^2 - 4\log p + \log\log p) + 16\log\log p \qquad \text{converges weakly to } F \tag{15}$$

as $n \to \infty$. This is different from the conclusion of Theorem 4.

3. Testing the covariance structure. The limiting laws derived in the last section have immediate statistical applications. Testing the covariance structure of a high-dimensional random variable is an important problem in statistical inference. In particular, as aforementioned, in econometrics when testing certain economic theories and in time series analysis in general it is of significant interest to test the hypothesis that the covariance matrix Σ is banded. That is, the variables have nonzero correlations only up to a certain lag τ. The limiting distribution of $L_{n,\tau}$ obtained in Section 2 can be readily used to construct a test for the bandedness of the covariance matrix in the Gaussian case.

Suppose we observe independent and identically distributed p-variate Gaussian variables $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ with mean $\mu_{p\times 1}$, covariance matrix $\Sigma_{p\times p} = (\sigma_{ij})$ and correlation matrix $R_{p\times p} = (r_{ij})$. For a given integer $\tau \geq 1$ and a given significant level $0 < \alpha < 1$, we wish to test the hypotheses

$$\begin{aligned} &H_0 : \sigma_{i,j} = 0 \qquad \text{for all } |i-j| \geq \tau \quad \text{vs.} \\ &H_a : \sigma_{i,j} \neq 0 \qquad \text{for some } |i-j| \geq \tau. \end{aligned} \tag{16}$$

A case of special interest is $\tau = 1$, which corresponds to testing independence of the Gaussian random variables. The asymptotic distribution of $L_{n,\tau}$ derived in Section 2.2 can be used to construct a convenient test statistic for testing the hypotheses in (16).

Based on the asymptotic result given in Theorem 4 that

$$P(nL_{n,\tau}^2 - 4\log p + \log\log p \leq y) \to e^{-(1/\sqrt{8\pi})e^{-y/2}}, \tag{17}$$

we define a test for testing the hypotheses in (16) by

$$T = I\big(L_{n,\tau}^2 \geq n^{-1}\big(4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}\big)\big). \tag{18}$$

That is, we reject the null hypothesis H_0 whenever

$$L_{n,\tau}^2 \geq n^{-1}\big(4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}\big).$$

Note that for $\tau = 1$, $L_{n,\tau}$ reduces to L_n and the test is then based on the coherence L_n.

THEOREM 5. *Under the conditions of Theorem 4, the test T defined in (18) has size α asymptotically.*

This result is a direct consequence of (17).

REMARK 3.1. For testing independence, another natural approach is to build a test based on the largest eigenvalue $\lambda_{\max}$ of the sample correlation matrix. However, the limiting distribution of the largest eigenvalue $\lambda_{\max}$ is unknown even for the case $p/n \to c$, a positive constant. For $\tau \geq 2$, the eigenvalues are not useful for testing bandedness of the covariance matrix.

4. Construction of compressed sensing matrices. As mentioned in the Introduction, an important problem in compressed sensing is the construction of measurement matrices $X_{n\times p}$ which enables the precise recovery of a sparse signal β from linear measurements $y = X\beta$ using an efficient recovery algorithm. Such a measurement matrix X is difficult to construct deterministically. It has been shown that randomly generated matrix X can satisfy the so called RIP condition with high probability.

The best known example is perhaps the $n \times p$ random matrix X whose entries $x_{i,j}$ are i.i.d. normal variables

$$x_{i,j} \overset{\text{i.i.d.}}{\sim} N(0, n^{-1}). \tag{19}$$

Other examples include generating $X = (x_{i,j})$ by Bernoulli random variables

$$x_{i,j} = \begin{cases} 1/\sqrt{n}, & \text{with probability } \frac{1}{2}; \\ -1/\sqrt{n}, & \text{with probability } \frac{1}{2}, \end{cases} \tag{20}$$

or more sparsely by

$$x_{i,j} = \begin{cases} \sqrt{3/n}, & \text{with probability } 1/6; \\ 0, & \text{with probability } 2/3; \\ -\sqrt{3/n}, & \text{with probability } 1/6. \end{cases} \tag{21}$$

These random matrices are shown to satisfy the RIP conditions with high probability. See Achlioptas (2003) and Baraniuk et al. (2008).

In addition to RIP, another commonly used condition is the mutual incoherence property (MIP) which requires the pairwise correlations among the column vectors of X to be small. In compressed sensing, $\tilde{L}_n$ (instead of L_n) is commonly used. It has been shown that the condition

$$(2k-1)\tilde{L}_n < 1 \tag{22}$$

ensures the exact recovery of k-sparse signal β in the noiseless case where $y = X\beta$, and stable recovery of sparse signal in the noisy case where

$$y = X\beta + z.$$

Here z is an error vector, not necessarily random. The MIP (22) is a very desirable property. When the measurement matrix X satisfies (22), the constrained ℓ_1 minimizer can be shown to be exact in the noiseless case and near-optimal in the noisy case. Under the MIP condition, the analysis of ℓ_1 minimization methods is also particularly simple. See, for example, Cai, Wang and Xu (2010b).

The results given in Theorems 1 and 2 can be used to show how likely a random matrix satisfies the MIP condition (22). Under the conditions of either Theorems 1 or 2,

$$\tilde{L}_n \sim 2\sqrt{\frac{\log p}{n}}.$$

So in order for the MIP condition (22) to hold, roughly the sparsity k should satisfy

$$k < \frac{1}{4}\sqrt{\frac{n}{\log p}}.$$

In fact, we have the following more precise result which is proved in Cai and Jiang (2010).

PROPOSITION 4.1. *Let $X_n = (x_{ij})_{n\times p}$ where x_{ij}'s are i.i.d. random variables with mean μ, variance $\sigma^2 > 0$ and $Ee^{t_0|x_{11}|^2} < \infty$ for some $t_0 > 0$. Let $\tilde{L}_n$ be as in (3). Then $P(\tilde{L}_n \ge t) \le 3p^2e^{-ng(t)}$ where $g(t) = \min\{I_1(\frac{t}{2}), I_2(\frac{1}{2})\} > 0$ for any $t > 0$ and*

$$I_1(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \log Ee^{\theta\xi\eta}\} \quad \textit{and} \quad I_2(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \log Ee^{\theta\xi^2}\},$$

and $\xi, \eta, (x_{11}-\mu)/\sigma$ are i.i.d.

We now consider the three particular random matrices mentioned at the beginning of this section.

EXAMPLE 1. Let $x_{11} \sim N(0, n^{-1})$ as in (19). In this case, according to the above proposition, we have

$$P\big((2k-1)\tilde{L}_n < 1\big) \ge 1 - 3p^2 \exp\left\{-\frac{n}{12(2k-1)^2}\right\} \tag{23}$$

for all $n \ge 2$ and $k \ge 1$. The verification of this example together with the next two are given in Cai and Jiang (2010).

EXAMPLE 2. Let x_{11} be such that $P(x_{11} = \pm 1/\sqrt{n}) = 1/2$ as in (20). In this case, (23) holds for all $n \ge 2$ and $k \ge 1$.

EXAMPLE 3. Let x_{11} be such that $P(x_{11} = \pm\sqrt{3/n}) = 1/6$ and $P(x_{11} = 0) = 2/3$ as in (21). Then (23) holds for all $n \ge 2$ and $k \ge 2$.

REMARK 4.1. One can see from the above that (23) is true for all of the three examples with different restrictions on k. In fact this is always the case as long as $Ee^{t_0|x_{11}|^2} < \infty$ for some $t_0 > 0$, which can be seen from Lemma 0.1 in Cai and Jiang (2010).

REMARK 4.2. Here we would like to point out an error on page 801 of Donoho (2006b) and page 2147 of Candès and Plan (2009) that the coherence of a random matrix with i.i.d. Gaussian entries is about $2\sqrt{\frac{\log p}{n}}$, not $\sqrt{\frac{2\log p}{n}}$.

5. Discussion and comparison with related results. This paper studies the limiting laws of the largest magnitude of the off-diagonal entries of the sample correlation matrix in the high-dimensional setting. Entries of other types of random matrices have been studied in the literature; see, for example, Diaconis, Eaton and Lauritzen (1992) and Jiang (2004a, 2005, 2006, 2009). Asymptotic properties of the eigenvalues of the sample correlation matrix have also been studied when both p and n are large and proportional to each other. For instance, it is proved in Jiang (2004b) that the empirical distributions of the eigenvalues of the sample correlation matrices converge to the Marchenko–Pastur law; the largest and smallest eigenvalues satisfy certain law of large numbers. However, the high-dimensional case of $p \gg n$ remains an open problem.

The motivations of our current work consist of the applications to testing covariance structure and construction of compressed sensing matrices in the ultra-high-dimensional setting where the dimension p can be as large as e^{n^β} for some $0 < \beta < 1$. The setting is different from those considered in the earlier literature such as Jiang (2010a, 2010b), Zhou (2007), Liu, Lin and Shao (2008) and Li, Liu and Rosalsky (2009). Our main theorems and techniques are different from those mentioned above in the following two aspects:

(a) Given $n \to \infty$, we push the size of p as large as we can to make the law of large numbers and limiting results on L_n and $\tilde{L}_n$ valid. Our current theorems say that, under some moment conditions, these results hold as long as $\log p = o(n^\beta)$ for a certain $\beta > 0$.

(b) We study L_n and $\tilde{L}_n$ when the p coordinates of the underlying multivariate distribution are not i.i.d. Instead, the p coordinates follow a multivariate normal distribution $N_p(\mu, \Sigma)$ with Σ being banded and μ arbitrary. Obviously, the p coordinates are dependent in this case. The proofs of our theorems are more subtle and involved than those in the earlier papers. In fact, we have to consider the dependence structure of Σ in detail, which is more complicated than the independent case. See Lemmas 6.9, 6.10 and 6.11.

Liu, Lin and Shao (2008) introduced a statistic for testing independence that is different from L_n and $\tilde{L}_n$ to improve the convergence speed of the two statistics under the constraint $c_1 n^\alpha \le p \le c_2 n^\alpha$ for some constants $c_1, c_2, \alpha > 0$. In this paper, while pushing the order of p as large as possible to have the limit theorems, we focus on the behavior of L_n and $\tilde{L}_n$ only. This is because L_n and $\tilde{L}_n$ are specifically used in some applications such as compressed sensing. On the other hand, we also consider a more general testing problem where one wishes to test the bandedness of the covariance matrix Σ in $N_p(\mu, \Sigma)$ while allowing μ to be arbitrary. We propose the statistic $L_{n,\tau}$ in (5) and derive its law of large numbers and its limiting distribution. To our knowledge, this is new in the literature. It is interesting to explore the possibility of improving the convergence speed by modifying $L_{n,\tau}$ as that of L_n in Liu, Lin and Shao (2008). We leave this as future work.

6. Proofs. In this section we prove Theorems 1–4. The letter C stands for a constant that may vary from place to place throughout this section. Also, we sometimes write p for p_n if there is no confusion. For any square matrix $A = (a_{i,j})$, define $|||A||| = \max_{1\le i\ne j\le n}|a_{i,j}|$; that is, the maximum of the absolute values of the off-diagonal entries of A.

We begin by collecting a few essential technical lemmas in Section 6.1 without proof. Other technical lemmas used in the proofs of the main results are proved in Cai and Jiang (2010).

6.1. *Technical tools.*

LEMMA 6.1 [Lemma 2.2 from Jiang (2004a)]. *Recall x_i and Γ_n in* (1). *Let $h_i = \|x_i - \bar{x}_i\|/\sqrt{n}$ for each i. Then*

$$|||n\Gamma_n - X_n^T X_n||| \le (b_{n,1}^2 + 2b_{n,1})W_n b_{n,3}^{-2} + n b_{n,3}^{-2} b_{n,4}^2,$$

where

$$b_{n,1} = \max_{1\le i\le p}|h_i - 1|, \qquad W_n = \max_{1\le i<j\le p}|x_i^T x_j|,$$

$$b_{n,3} = \min_{1\le i\le p} h_i, \qquad b_{n,4} = \max_{1\le i\le p}|\bar{x}_i|.$$

The following Poisson approximation result is essentially a special case of Theorem 1 from Arratia, Goldstein and Gordon (1989).

LEMMA 6.2. *Let I be an index set and $\{B_\alpha, \alpha \in I\}$ be a set of subsets of I, that is, $B_\alpha \subset I$ for each $\alpha \in I$. Let also $\{\eta_\alpha, \alpha \in I\}$ be random variables. For a given $t \in \mathbb{R}$, set $\lambda = \sum_{\alpha \in I} P(\eta_\alpha > t)$. Then*

$$\left|P\Big(\max_{\alpha \in I} \eta_\alpha \le t\Big) - e^{-\lambda}\right| \le (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3),$$

where

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t) P(\eta_\beta > t), \qquad b_2 = \sum_{\alpha \in I} \sum_{\alpha \ne \beta \in B_\alpha} P(\eta_\alpha > t, \eta_\beta > t),$$

$$b_3 = \sum_{\alpha \in I} E\big|P\big(\eta_\alpha > t | \sigma(\eta_\beta, \beta \notin B_\alpha)\big) - P(\eta_\alpha > t)\big|,$$

and $\sigma(\eta_\beta, \beta \notin B_\alpha)$ is the σ-algebra generated by $\{\eta_\beta, \beta \notin B_\alpha\}$. In particular, if η_α is independent of $\{\eta_\beta, \beta \notin B_\alpha\}$ for each α, then $b_3 = 0$.

The following conclusion is Example 1 from Sakhanenko (1991). See also Lemma 6.2 from Liu, Lin and Shao (2008).

LEMMA 6.3. *Let ξ_i, $1 \le i \le n$, be independent random variables with $E\xi_i = 0$. Set*

$$s_n^2 = \sum_{i=1}^n E\xi_i^2, \qquad \varrho_n = \sum_{i=1}^n E|\xi_i|^3, \qquad S_n = \sum_{i=1}^n \xi_i.$$

Assume $\max_{1 \le i \le n} |\xi_i| \le c_n s_n$ *for some* $0 < c_n \le 1$. *Then*

$$P(S_n \ge x s_n) = e^{\gamma(x/s_n)}\big(1 - \Phi(x)\big)\big(1 + \theta_{n,x}(1+x)s_n^{-3}\varrho_n\big)$$

for $0 < x \le 1/(18c_n)$, *where* $|\gamma(x)| \le 2x^3\varrho_n$ *and* $|\theta_{n,x}| \le 36$.

The following are moderate deviation results from Chen (1990); see also Chen (1991), Dembo and Zeitouni (1998) and Ledoux (1992). They are a special type of large deviations.

LEMMA 6.4. *Suppose $\xi_1, \xi_2, \dots$ are i.i.d. r.v.'s with $E\xi_1 = 0$ and $E\xi_1^2 = 1$. Set $S_n = \sum_{i=1}^n \xi_i$.*

(i) *Let $0 < \alpha \le 1$ and $\{a_n; n \ge 1\}$ satisfy that $a_n \to +\infty$ and $a_n = o(n^{\alpha/(2(2-\alpha))})$. If $Ee^{t_0|\xi_1|^\alpha} < \infty$ for some $t_0 > 0$, then*

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log P\left(\frac{S_n}{\sqrt{n}a_n} \ge u\right) = -\frac{u^2}{2} \tag{24}$$

for any $u > 0$.

(ii) *Let $0 < \alpha < 1$ and $\{a_n; n \ge 1\}$ satisfy that $a_n \to +\infty$ and $a_n = O(n^{\alpha/(2(2-\alpha))})$. If $Ee^{t|\xi_1|^\alpha} < \infty$ for all $t > 0$, then (24) also holds.*

6.2. *Proofs of Theorems 1 and 2.* The following is known:

(25) if $\{X_n\}$ are tight, then for any sequence of constants $\{\varepsilon_n\}$ with $\lim_{n\to\infty}\varepsilon_n = 0$, we have $\varepsilon_n X_n \to 0$ in probability as $n \to \infty$.

Reviewing the notation $b_{n,i}$'s defined in Lemma 6.1, we have the following properties.

LEMMA 6.5. *Let $\{x_{ij}; i \ge 1, j \ge 1\}$ be i.i.d. random variables with $Ex_{11} = 0$ and $Ex_{11}^2 = 1$. Then, $b_{n,3} \to 1$ in probability as $n \to \infty$, and $\{\sqrt{n/\log p}\, b_{n,1}\}$ and $\{\sqrt{n/\log p}\, b_{n,4}\}$ are tight provided one of the following conditions holds*:

(i) $|x_{11}| \le C$ *for some constant* $C > 0$, $p_n \to \infty$ *and* $\log p_n = o(n)$ *as* $n \to \infty$;

(ii) $Ee^{t_0|x_{11}|^\alpha} < \infty$ *for some* $\alpha > 0$ *and* $t_0 > 0$ *and* $p_n \to \infty$ *with* $\log p_n = o(n^\beta)$ *as* $n \to \infty$, *where* $\beta = \alpha/(4+\alpha)$.

LEMMA 6.6. *Let $\{x_{ij}; i \ge 1, j \ge 1\}$ be i.i.d. random variables with $|x_{11}| \le C$ for a finite constant $C > 0$, $Ex_{11} = 0$ and $E(x_{11}^2) = 1$. Assume $p = p(n) \to \infty$ and $\log p = o(n)$ as $n \to \infty$. Then, for any $\varepsilon > 0$ and a sequence of positive numbers $\{t_n\}$ with limit $t > 0$,*

$$\Psi_n := E\left\{P^1\left(\left|\sum_{k=1}^n x_{k1}x_{k2}\right| > t_n\sqrt{n\log p}\right)^2\right\} = O\left(\frac{1}{p^{t^2-\varepsilon}}\right)$$

as $n \to \infty$, where P^1 stands for the conditional probability given $\{x_{k1}, 1 \le k \le n\}$.

LEMMA 6.7. *Suppose $\{x_{ij}; i \ge 1, j \ge 1\}$ are i.i.d. random variables with $Ex_{11} = 0$, $E(x_{11}^2) = 1$ and $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $t_0 > 0$ and $\alpha > 0$. Assume $p = p(n) \to \infty$ and $\log p = o(n^\beta)$ as $n \to \infty$, where $\beta = \alpha/(4+\alpha)$. Then, for any $\varepsilon > 0$ and a sequence of positive numbers $\{t_n\}$ with limit $t > 0$,*

$$\Psi_n := E\left\{P^1\left(\left|\sum_{k=1}^n x_{k1}x_{k2}\right| > t_n\sqrt{n\log p}\right)^2\right\} = O\left(\frac{1}{p^{t^2-\varepsilon}}\right)$$

as $n \to \infty$, where P^1 stands for the conditional probability given $\{x_{k1}, 1 \le k \le n\}$.

Lemmas 6.5, 6.6 and 6.7 are proved in Cai and Jiang (2010).

PROPOSITION 6.1. *Suppose the conditions in Lemma 6.6 hold with $X_n = (x_{ij})_{n\times p} = (x_1, \dots, x_p)$. Define $W_n = \max_{1\le i<j\le p}|x_i^T x_j|$. Then*

$$\frac{W_n}{\sqrt{n\log p}} \to 2$$

in probability as $n \to \infty$.

PROOF. We first prove

$$\lim_{n\to\infty} P\left(\frac{W_n}{\sqrt{n\log p}} \ge 2+2\varepsilon\right) = 0 \tag{26}$$

for any $\varepsilon > 0$. First, since $\{x_{ij}; i \ge 1, j \ge 1\}$ are i.i.d., we have

$$P\big(W_n \ge (2+2\varepsilon)\sqrt{n\log p}\big) \le \binom{p}{2} \cdot P\left(\left|\sum_{k=1}^{n} x_{k1}x_{k2}\right| \ge (2+2\varepsilon)\sqrt{n\log p}\right) \tag{27}$$

for any $\varepsilon > 0$. Notice $E(|x_{11}x_{12}|^2) = E(|x_{11}|^2) \cdot E(|x_{12}|^2) = 1$. It follows from Lemma 6.4(i) and the conditions $Ee^{|x_{11}x_{12}|} < \infty$ and $\log p = o(n)$ that

$$P\left(\left|\sum_{k=1}^{n} x_{k1}x_{k2}\right| \ge (2+2\varepsilon)\sqrt{n\log p}\right) \le e^{-((2+\varepsilon)^2/2)\log p} \le \frac{1}{p^{2+\varepsilon}} \tag{28}$$

for sufficiently large n. The above two assertions conclude

$$P\big(W_n \ge (2+2\varepsilon)\sqrt{n\log p}\big) \le \frac{1}{p^{\varepsilon}} \to 0 \tag{29}$$

as $n \to \infty$. Thus, (26) holds. Now, to finish the proof, we only need to show

$$\lim_{n\to\infty} P\left(\frac{W_n}{\sqrt{n\log p}} \le 2-\varepsilon\right) = 0 \tag{30}$$

for any $\varepsilon > 0$ small enough.

Set $a_n = (2-\varepsilon)\sqrt{n\log p}$ for $0 < \varepsilon < 2$ and

$$y_{ij}^{(n)} = \sum_{k=1}^{n} x_{ki}x_{kj}$$

for $1 \le i, j \le n$. Then $W_n = \max_{1\le i<j\le p}|y_{ij}^{(n)}|$ for all $n \ge 1$.

Take $I = \{(i, j); 1 \le i < j \le p\}$. For $u = (i, j) \in I$, set $B_u = \{(k,l) \in I$; one of k and $l = i$ or j, but $(k,l) \ne u\}$, $\eta_u = |y_{ij}^{(n)}|$, $t = a_n$ and $A_u = A_{ij} = \{|y_{ij}^{(n)}| > a_n\}$. By the i.i.d. assumption on $\{x_{ij}\}$ and Lemma 6.2,

$$P(W_n \le a_n) \le e^{-\lambda_n} + b_{1,n} + b_{2,n}, \tag{31}$$

where

$$\begin{aligned} &\lambda_n = \frac{p(p-1)}{2}P(A_{12}), \qquad b_{1,n} \le 2p^3P(A_{12})^2 \quad \text{and} \\ &b_{2,n} \le 2p^3P(A_{12}A_{13}). \end{aligned} \tag{32}$$

Remember that $y_{12}^{(n)}$ is a sum of i.i.d. bounded random variables with mean 0 and variance 1. By (i) of Lemma 6.4, using conditions $Ee^{t|x_{11}x_{12}|} < \infty$ for any $t > 0$ and $\log p = o(n)$ as $n \to \infty$, we know

$$\lim_{n\to\infty} \frac{1}{\log p}\log P(A_{12}) = -\frac{(2-\varepsilon)^2}{2} \tag{33}$$

for any $\varepsilon \in (0, 2)$. Noticing $2 - 2\varepsilon < (2-\varepsilon)^2/2 < 2 - \varepsilon$ for $\varepsilon \in (0, 1)$, we have

$$\frac{1}{p^{2-\varepsilon}} \le P(A_{12}) \le \frac{1}{p^{2-2\varepsilon}} \tag{34}$$

as n is sufficiently large. This implies

$$e^{-\lambda_n} \le e^{-p^{\varepsilon}/3} \quad \text{and} \quad b_{1,n} \le \frac{2}{p^{1-4\varepsilon}} \tag{35}$$

for $\varepsilon \in (0, 1/4)$ as n is large enough. On the other hand, by independence

$$\begin{aligned} P(A_{12}A_{13}) &= P\bigl(|y_{12}^{(n)}| > a_n, |y_{13}^{(n)}| > a_n\bigr) \\ &= E\left\{P^1\left(\left|\sum_{k=1}^{n} x_{k1}x_{k2}\right| > a_n\right)^2\right\}, \end{aligned} \tag{36}$$

where P^1 stands for the conditional probability given $\{x_{k1}, 1 \le k \le n\}$. By Lemma 6.6,

$$P(A_{12}A_{13}) \le \frac{1}{p^{4-4\varepsilon}} \tag{37}$$

for any $\varepsilon > 0$ as n is sufficiently large. Therefore, taking $\varepsilon \in (0, 1/4)$, we have

$$b_{2,n} \le 2p^3 P(A_{12}A_{13}) \le \frac{2}{p^{1-4\varepsilon}} \to 0 \tag{38}$$

as $n \to \infty$. This together with (31) and (35) concludes (30). $\square$

PROPOSITION 6.2. *Suppose the conditions in Lemma 6.7 hold. Let W_n be as in Lemma 6.1. Then*

$$\frac{W_n}{\sqrt{n \log p}} \to 2$$

in probability as $n \to \infty$.

The proof of Proposition 6.2 is similar to that of Proposition 6.1. Details are given in Cai and Jiang (2010).

PROOF OF THEOREM 1. First, for constants $\mu_i \in \mathbb{R}$ and $\sigma_i > 0$, $i = 1, 2, \dots, p$, it is easy to see that matrix $X_n = (x_{ij})_{n\times p} = (x_1, x_2, \dots, x_p)$ and $(\sigma_1 x_1 + \mu_1 e, \sigma_2 x_2 + \mu_2 e, \dots, \sigma_p x_p + \mu_p e)$ generate the same sample correlation matrix $\Gamma_n = (\rho_{ij})$, where ρ_{ij} is as in (1) and $e = (1, \dots, 1)' \in \mathbb{R}^n$. Thus, w.l.o.g., we prove the theorem next by assuming that $\{x_{ij}; 1 \le i \le n, 1 \le j \le p\}$ are i.i.d. random variables with mean zero and variance 1.

By Proposition 6.1, under condition $\log p = o(n)$,

$$\frac{W_n}{\sqrt{n \log p}} \to 2 \tag{39}$$

in probability as $n \to \infty$. Thus, to prove the theorem, it is enough to show

$$\frac{nL_n - W_n}{\sqrt{n \log p}} \to 0 \tag{40}$$

in probability as $n \to \infty$. From Lemma 6.1,

$$|nL_n - W_n| \le \||n\Gamma_n - X_n^T X_n\|| \le (b_{n,1}^2 + 2b_{n,1})W_n b_{n,3}^{-2} + nb_{n,3}^{-2} b_{n,4}^2.$$

By (i) of Lemma 6.5, $b_{n,3} \to 1$ in probability as $n \to \infty$, $\{\sqrt{n/\log p}\, b_{n,1}\}$ and $\{\sqrt{n/\log p}\, b_{n,4}\}$ are all tight. Set $b'_{n,1} = \sqrt{n/\log p}\, b_{n,1}$ and $b'_{n,4} = \sqrt{n/\log p}\, b_{n,4}$ for all $n \ge 1$. Then $\{b'_{n,1}\}$ and $\{b'_{n,4}\}$ are both tight. It follows that

$$\frac{|nL_n - W_n|}{\sqrt{n \log p}} \le \sqrt{\frac{\log p}{n}} \left(\sqrt{\frac{\log p}{n}} b'^2_{n,1} + 2b'_{n,1} \right) \cdot \frac{W_n}{\sqrt{n \log p}} \cdot b_{n,3}^{-2} + \sqrt{\frac{\log p}{n}} b_{n,3}^{-2} b'^2_{n,4},$$

which concludes (40) by (25). $\square$

PROOF OF THEOREM 2. In the proof of Theorem 1, replace "Proposition 6.1" with "Proposition 6.2" and "(i) of Lemma 6.5" with "(ii) of Lemma 6.5," keep all other statements the same, we then get the desired result. $\square$

6.3. *Proof of Theorem 3.*

LEMMA 6.8. *Let $\xi_1, \dots, \xi_n$ be i.i.d. random variables with $E\xi_1 = 0$, $E\xi_1^2 = 1$ and $Ee^{t_0|\xi_1|^\alpha} < \infty$ for some $t_0 > 0$ and $0 < \alpha \le 1$. Put $S_n = \sum_{i=1}^n \xi_i$ and $\beta = \alpha/(2+\alpha)$. Then, for any $\{p_n; n \ge 1\}$ with $0 < p_n \to \infty$ and $\log p_n = o(n^\beta)$ and $\{y_n; n \ge 1\}$ with $y_n \to y > 0$,*

$$P\left(\frac{S_n}{\sqrt{n \log p_n}} \ge y_n \right) \sim \frac{p_n^{-y_n^2/2} (\log p_n)^{-1/2}}{\sqrt{2\pi}\, y}$$

as $n \to \infty$.

The proof of this lemma is given at Cai and Jiang (2010).

PROPOSITION 6.3. *Let $\{x_{ij}; i \ge 1, j \ge 1\}$ be i.i.d. random variables with $Ex_{11} = 0$, $E(x_{11}^2) = 1$ and $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $0 < \alpha \le 2$ and $t_0 > 0$. Set $\beta = \alpha/(4+\alpha)$. Assume $p = p(n) \to \infty$ and $\log p = o(n^\beta)$ as $n \to \infty$. Then*

$$P\left(\frac{W_n^2 - \alpha_n}{n} \le z \right) \to e^{-Ke^{-z/2}}$$

as $n \to \infty$ for any $z \in \mathbb{R}$, where $\alpha_n = 4n \log p - n \log(\log p)$ and $K = (\sqrt{8\pi})^{-1}$.

PROOF. It suffices to show that

$$P\Big(\max_{1\le i<j\le p}|y_{ij}|\le\sqrt{\alpha_n+nz}\Big)\to e^{-Ke^{-z/2}},\tag{41}$$

where $y_{ij}=\sum_{k=1}^n x_{ki}x_{kj}$. We now apply Lemma 6.2 to prove (41). Take $I=\{(i,j);1\le i<j\le p\}$. For $u=(i,j)\in I$, set $X_u=|y_{ij}|$ and $B_u=\{(k,l)\in I$; one of k and $l=i$ or j, but $(k,l)\ne u\}$. Let $a_n=\sqrt{\alpha_n+nz}$ and $A_{ij}=\{|y_{ij}|>a_n\}$. Since $\{y_{ij};(i,j)\in I\}$ are identically distributed, by Lemma 6.2,

$$|P(W_n\le a_n)-e^{-\lambda_n}|\le b_{1,n}+b_{2,n},\tag{42}$$

where

$$\begin{aligned}&\lambda_n=\frac{p(p-1)}{2}P(A_{12}),\qquad b_{1,n}\le 2p^3P(A_{12})^2\quad\text{and}\\&b_{2,n}\le 2p^3P(A_{12}A_{13}).\end{aligned}\tag{43}$$

We first calculate λ_n. Write

$$\lambda_n=\frac{p^2-p}{2}P\left(\frac{|y_{12}|}{\sqrt{n}}>\sqrt{\frac{\alpha_n}{n}+z}\right)\tag{44}$$

and $y_{12}=\sum_{i=1}^n\xi_i$, where $\{\xi_i;1\le i\le n\}$ are i.i.d. random variables with the same distribution as that of $x_{11}x_{12}$. In particular, $E\xi_1=0$ and $E\xi_1^2=1$. Note $\alpha_1:=\alpha/2\le 1$. We then have

$$|x_{11}x_{12}|^{\alpha_1}\le\left(\frac{x_{11}^2+x_{12}^2}{2}\right)^{\alpha_1}\le\frac{1}{2^{\alpha_1}}(|x_{11}|^\alpha+|x_{12}|^\alpha).$$

Hence, by independence,

$$Ee^{t_0|\xi_1|^{\alpha_1}}=Ee^{t_0|x_{11}x_{12}|^{\alpha_1}}<\infty.$$

Let $y_n=\sqrt{(\frac{\alpha_n}{n}+z)/\log p}$. Then $y_n\to 2$ as $n\to\infty$. By Lemma 6.8,

$$\begin{aligned}P\left(\frac{y_{12}}{\sqrt{n}}>\sqrt{\frac{\alpha_n}{n}+z}\right)&=P\left(\frac{\sum_{i=1}^n\xi_i}{\sqrt{n\log p}}>y_n\right)\sim\frac{p^{-y_n^2/2}(\log p)^{-1/2}}{2\sqrt{2\pi}}\\&\sim\frac{e^{-z/2}}{\sqrt{8\pi}}\cdot\frac{1}{p^2}\end{aligned}$$

as $n\to\infty$. Considering $Ex_{ij}=0$, it is easy to see that the above also holds if y_{12} is replaced by $-y_{12}$. These and (44) imply that

$$\lambda_n\sim\frac{p^2-p}{2}\cdot 2\cdot\frac{e^{-z/2}}{\sqrt{8\pi}}\cdot\frac{1}{p^2}\sim\frac{e^{-z/2}}{\sqrt{8\pi}}\tag{45}$$

as $n\to\infty$.

Recall (42) and (43), to complete the proof, we have to verify that $b_{1,n} \to 0$ and $b_{2,n} \to 0$ as $n \to \infty$. By (43), (44) and (45),

$$b_{1,n} \le 2p^3 P(A_{12})^2 = \frac{8p^3\lambda_n^2}{(p^2-p)^2} = O\left(\frac{1}{p}\right)$$

as $n \to \infty$. Also, by (43),

$$\begin{aligned} b_{2,n} &\le 2p^3 P\big(|y_{12}| > \sqrt{\alpha_n + nz}, |y_{13}| > \sqrt{\alpha_n + nz}\big) \\ &= 2p^3 E\left\{P^1\left(\left|\sum_{k=1}^n x_{k1}x_{k2}\right| > t_n\sqrt{n\log p}\right)^2\right\}, \end{aligned}$$

where P^1 stands for the conditional probability given $\{x_{k,1}; 1 \le k \le n\}$ and $t_n := \sqrt{\alpha_n + nz}/\sqrt{n\log p} \to 2$. By Lemma 6.7, the above expectation is equal to $O(p^{\varepsilon-4})$ as $n \to \infty$ for any $\varepsilon > 0$. Now choose $\varepsilon \in (0,1)$, then $b_{2,n} = O(p^{\varepsilon-1}) \to 0$ as $n \to \infty$. The proof is then completed. $\square$

PROOF OF THEOREM 3. By the first paragraph in the proof of Theorem 1, w.l.o.g., assume $\mu = 0$ and $\sigma = 1$. From Proposition 6.3 and the Slusky lemma, it suffices to show

$$\frac{n^2L_n^2 - W_n^2}{n} \to 0 \tag{46}$$

in probability as $n \to \infty$. Let $\Delta_n = |nL_n - W_n|$ for $n \ge 1$. Observe that

$$|n^2L_n^2 - W_n^2| = |nL_n - W_n| \cdot |nL_n + W_n| \le \Delta_n \cdot (\Delta_n + 2W_n). \tag{47}$$

It is easy to see from Proposition 6.3 that

$$\frac{W_n}{\sqrt{n\log p}} \to 2 \tag{48}$$

in probability as $n \to \infty$. By Lemma 6.1,

$$\Delta_n \le \|n\Gamma_n - X_n^T X_n\| \le (b_{n,1}^2 + 2b_{n,1})W_n b_{n,3}^{-2} + n b_{n,3}^{-2} b_{n,4}^2.$$

By (ii) of Lemma 6.5, $b_{n,3} \to 1$ in probability as $n \to \infty$, $\{\sqrt{n/\log p}\, b_{n,1}\}$ and $\{\sqrt{n/\log p}\, b_{n,4}\}$ are tight. Set $b'_{n,1} = \sqrt{n/\log p}\, b_{n,1}$ and $b'_{n,4} = \sqrt{n/\log p}\, b_{n,4}$ for all $n \ge 1$. Then $\{b'_{n,1}\}$ and $\{b'_{n,4}\}$ are tight. It follows that

$$\frac{\Delta_n}{\log p} \le \left(\sqrt{\frac{\log p}{n}} b_{n,1}'^2 + 2b'_{n,1}\right) \cdot \frac{W_n}{\sqrt{n\log p}} \cdot b_{n,3}^{-2} + b_{n,3}^{-2} b_{n,4}'^2,$$

which combining with (48) yields that

$$\left\{\frac{\Delta_n}{\log p}\right\} \text{ is tight.} \tag{49}$$

This and (48) imply that $\{\Delta_n'\}$ and $\{W_n'\}$ are tight, where $\Delta_n' := \Delta_n/\log p$ and $W_n' := W_n/\sqrt{n \log p}$. From (47) and then (25),

$$
\begin{aligned}
\frac{|n^2 L_n^2 - W_n^2|}{n} &\le \frac{(\log p)\Delta_n'\{(\log p)\Delta_n' + 2\sqrt{n\log p}\, W_n'\}}{n} \\
&\le 2\sqrt{\frac{(\log p)^3}{n}}\left(\sqrt{\frac{\log p}{n}}\Delta_n' + W_n'\right) \to 0
\end{aligned} \tag{50}
$$

in probability as $n \to \infty$ since $\log p = o(n^{1/3})$. This gives (46). $\square$

6.4. *Proof of Theorem 4*. We begin to prove Theorem 4 by stating three technical lemmas which are proved in Cai and Jiang (2010).

LEMMA 6.9. *Let $\{(u_{k1}, u_{k2}, u_{k3}, u_{k4})^T; 1 \le i \le n\}$ be a sequence of i.i.d. random vectors with distribution $N_4(0, \Sigma_4)$ where*

$$
\Sigma_4 = \begin{pmatrix} 1 & 0 & r & 0 \\ 0 & 1 & 0 & 0 \\ r & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad |r| \le 1.
$$

Set $a_n = (4n\log p - n\log(\log p) + ny)^{1/2}$ for $n \ge e^e$ and $y \in \mathbb{R}$. Suppose $n \to \infty$, $p \to \infty$ with $\log p = o(n^{1/3})$. Then

$$
\sup_{|r|\le 1} P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n, \left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) = O\left(\frac{1}{p^{4-\varepsilon}}\right) \tag{51}
$$

for any $\varepsilon > 0$.

LEMMA 6.10. *Let $\{(u_{k1}, u_{k2}, u_{k3}, u_{k4})^T; 1 \le i \le n\}$ be a sequence of i.i.d. random vectors with distribution $N_4(0, \Sigma_4)$ where*

$$
\Sigma_4 = \begin{pmatrix} 1 & 0 & r_1 & 0 \\ 0 & 1 & r_2 & 0 \\ r_1 & r_2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad |r_1| \le 1, |r_2| \le 1.
$$

Set $a_n = (4n\log p - n\log(\log p) + ny)^{1/2}$ for $n \ge e^e$ and $y \in \mathbb{R}$. Suppose $n \to \infty$, $p \to \infty$ with $\log p = o(n^{1/3})$. Then, as $n \to \infty$,

$$
\sup_{|r_1|,|r_2|\le 1} P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n, \left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) = O(p^{-8/3+\varepsilon})
$$

for any $\varepsilon > 0$.

LEMMA 6.11. *Let* $\{(u_{k1}, u_{k2}, u_{k3}, u_{k4})^T; 1 \le i \le n\}$ *be a sequence of i.i.d. random vectors with distribution* $N_4(0, \Sigma_4)$ *where*

$$\Sigma_4 = \begin{pmatrix} 1 & 0 & r_1 & 0 \\ 0 & 1 & 0 & r_2 \\ r_1 & 0 & 1 & 0 \\ 0 & r_2 & 0 & 1 \end{pmatrix}, \qquad |r_1| \le 1, |r_2| \le 1.$$

Set $a_n = (4n \log p - n \log(\log p) + ny)^{1/2}$ *for* $n \ge e^e$ *and* $y \in \mathbb{R}$. *Suppose* $n \to \infty$, $p \to \infty$ *with* $\log p = o(n^{1/3})$. *Then, for any* $\delta \in (0, 1)$, *there exists* $\varepsilon_0 = \varepsilon(\delta) > 0$ *such that*

$$\sup_{|r_1|, |r_2| \le 1-\delta} P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n, \left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) = O(p^{-2-\varepsilon_0}). \tag{52}$$

Recall the paragraph above (11) on notation τ, $\Sigma = (\sigma_{ij})_{p\times p}$ and the assumption that the n rows of $X_n = (x_{ij})_{n\times p}$ are i.i.d. with distribution $N_p(\mu, \Sigma)$.

PROPOSITION 6.4. *Assume* $\mu = 0$ *and* $\sigma_{ii} = 1$ *for all* $1 \le i \le p$. *Define*

$$V_n = V_{n,\tau} = \max_{1\le i<j\le p, |j-i|\ge\tau} |x_i^T x_j|. \tag{53}$$

Suppose $n \to \infty$, $p = p_n \to \infty$ *with* $\log p = o(n^{1/3})$, $\tau = o(p^t)$ *for any* $t > 0$, *and for some* $\delta \in (0, 1)$, $|\Gamma_{p,\delta}| = o(p)$ *as* $n \to \infty$. *Then, under* H_0 *in* (11),

$$P\left(\frac{V_n^2 - \alpha_n}{n} \le y\right) \to e^{-Ke^{-y/2}}$$

as $n \to \infty$ *for any* $y \in \mathbb{R}$, *where* $\alpha_n = 4n \log p - n \log(\log p)$ *and* $K = (\sqrt{8\pi})^{-1}$.

PROOF. Set $a_n = (4n \log p - n \log(\log p) + ny)^{1/2}$,

$$\begin{aligned} \Lambda_p &= \left\{(i, j) : 1 \le i < j \le p, j - i \ge \tau, \max_{1\le k\ne i, k\ne j\le p} \{|r_{ik}|, |r_{jk}|\} \le 1 - \delta\right\}, \\ V_n' &= \max_{(i,j)\in\Lambda_p} \left|\sum_{k=1}^n x_{ki}x_{kj}\right|. \end{aligned} \tag{54}$$

Step 1. We claim that, to prove the proposition, it suffices to show

$$\lim_{n\to\infty} P(V_n' \le a_n) = e^{-Ke^{-y/2}} \tag{55}$$

for any $y \in \mathbb{R}$.

In fact, to prove the theorem, we need to show that

$$\lim_{n\to\infty} P(V_n > a_n) = 1 - e^{-Ke^{-y/2}} \tag{56}$$

for every $y \in \mathbb{R}$. Notice $\{x_{ki}, x_{kj}; 1 \le k \le n\}$ are $2n$ i.i.d. standard normals if $|j - i| \ge \tau$. Then

$$P(V_n > a_n) \le P(V_n' > a_n) + \sum P\left(\left|\sum_{k=1}^{n} x_{k1}x_{k\tau+1}\right| > a_n\right),$$

where the sum runs over all pair (i, j) such that $1 \le i < j \le p$ and one of i and j is in $\Gamma_{p,\delta}$. Note that $|x_{11}x_{1\tau+1}| \le (x_{11}^2 + x_{1\tau+1}^2)/2$, it follows that $Ee^{|x_{11}x_{1\tau+1}|/2} < \infty$ by independence and $E\exp(N(0,1)^2/4) < \infty$. Since $\{x_{k1}, x_{k\tau+1}; 1 \le k \le n\}$ are i.i.d. with mean zero and variance one, and $y_n := a_n/\sqrt{n \log p} \to 2$ as $n \to \infty$, taking $\alpha = 1$ in Lemma 6.8, we get

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{n\log p}}\left|\sum_{k=1}^{n} x_{k1}x_{k\tau+1}\right| > \frac{a_n}{\sqrt{n\log p}}\right) \\
&\quad \sim 2 \cdot \frac{p^{-y_n^2/2}(\log p)^{-1/2}}{2\sqrt{2\pi}} \sim \frac{e^{-y/2}}{\sqrt{2\pi}} \cdot \frac{1}{p^2}
\end{aligned} \tag{57}$$

as $n \to \infty$. Moreover, note that the total number of such pairs is no more than $2p|\Gamma_{p,\delta}|$. Therefore, $P(V_n' > a_n) \le P(V_n > a_n)$ and

$$\begin{aligned}
P(V_n > a_n) &\le P(V_n' > a_n) + 2p|\Gamma_{p,\delta}| \cdot P\left(\left|\sum_{k=1}^{n} x_{k1}x_{k\tau+1}\right| > a_n\right) \\
&\le P(V_n' > a_n) + o(p^2) \cdot O\left(\frac{1}{p^2}\right)
\end{aligned} \tag{58}$$

by the assumption on $\Gamma_{p,\delta}$ and (57). Thus, this joint with (56) gives (55).

Step 2. We now apply Lemma 6.2 to prove (55). Take $I = \Lambda_p$. For $(i, j) \in I$, set $Z_{ij} = |\sum_{k=1}^{n} x_{ki}x_{kj}|$,

$$\begin{aligned}
B_{i,j} = \bigl\{(k,l) \in \Lambda_p; |s - t| < \tau \text{ for some } s \in \{k,l\} \text{ and some } t \in \{i, j\}, \\
\text{but } (k,l) \ne (i,j)\bigr\},
\end{aligned}$$

$$a_n = \sqrt{\alpha_n + ny} \quad \text{and} \quad A_{ij} = \{|Z_{ij}| > a_n\}.$$

It is easy to see that $|B_{i,j}| \le 2 \cdot (2\tau + 2\tau)p = 8\tau p$ and that Z_{ij} are independent of $\{Z_{kl}; (k,l) \in \Lambda_p \setminus B_{i,j}\}$ for any $(i, j) \in \Lambda_p$. By Lemma 6.2,

$$|P(V_n \le a_n) - e^{-\lambda_n}| \le b_{1,n} + b_{2,n}, \tag{59}$$

where

$$\lambda_n = |\Lambda_p| \cdot P(A_{1\tau+1}), \qquad b_{1,n} \le \sum_{d\in\Lambda_p}\sum_{d'\in B_a} P(A_{12})^2 = 8\tau p^3 P(A_{1\tau+1})^2 \tag{60}$$

and

$$b_{2,n} \le \sum_{d\in\Lambda_p}\sum_{d\ne d'\in B_a} P(Z_d > t, Z_{d'} > t) \tag{61}$$

from the fact that $\{Z_{ij}; (i,j) \in \Lambda_p\}$ are identically distributed. We first calculate λ_n. By definition

$$\begin{aligned}\frac{p^2}{2} > |\Lambda_p| &\geq |\{(i,j); 1 \leq i < j \leq p, j-i \geq \tau\}| - 2p \cdot |\Gamma_{p,\delta}| \\ &= \sum_{i=1}^{p-\tau}(p-\tau-i+1) - 2p \cdot |\Gamma_{p,\delta}|.\end{aligned}$$

Now the sum above is equal to $\sum_{j=1}^{p-\tau} j = (p-\tau)(p-\tau+1)/2 \sim p^2/2$ since $\tau = o(p)$. By assumption $|\Gamma_{p,\delta}| = o(p)$ we conclude that

$$|\Lambda_p| \sim \frac{p^2}{2} \tag{62}$$

as $n \to \infty$. It then follows from (57) that

$$\lambda_n \sim \frac{p^2}{2} \cdot \frac{e^{-y/2}}{\sqrt{2\pi}} \cdot \frac{1}{p^2} \sim \frac{e^{-y/2}}{\sqrt{8\pi}} \qquad \text{as } n \to \infty. \tag{63}$$

Recall (59) and (63), to complete the proof, we have to verify that $b_{1,n} \to 0$ and $b_{2,n} \to 0$ as $n \to \infty$. Clearly, by the first expression in (60), we get from (63) and then (62) that

$$b_{1,n} \leq 8\tau p^3 P(A_{1\tau+1})^2 = \frac{8\tau p^3 \lambda_n^2}{|\Lambda_p|^2} = O\left(\frac{\tau}{p}\right) \to 0$$

as $n \to \infty$ by the assumption on τ.

Step 3. Now we consider $b_{2,n}$. Write $d = (d_1, d_2) \in \Lambda_p$ and $d' = (d_3, d_4) \in \Lambda_p$ with $d_1 < d_2$ and $d_3 < d_4$. It is easy to see from (61) that

$$b_{2,n} \leq 2 \sum P(Z_d > a_n, Z_{d'} > a_n),$$

where the sum runs over every pair (d, d') satisfying

$$d, d' \in \Lambda_p, \qquad d \neq d', \qquad d_1 \leq d_3 \quad \text{and} \quad |d_i - d_j| < \tau \tag{64}$$

for some $i \in \{1,2\}$ and some $j \in \{3,4\}$. Geometrically, there are three cases for the locations of $d = (d_1, d_2)$ and $d' = (d_3, d_4)$:

$$(1) \quad d_2 \leq d_3; \qquad (2) \quad d_1 \leq d_3 < d_4 \leq d_2; \qquad (3) \quad d_1 \leq d_3 \leq d_2 \leq d_4. \tag{65}$$

Let Ω_j be the subset of index (d, d') with restrictions (64) and (j) for $j = 1, 2, 3$. Then

$$b_{2,n} \leq 2 \sum_{i=1}^{3} \sum_{(d,d') \in \Omega_i} P(Z_d > a_n, Z_{d'} > a_n). \tag{66}$$

We next analyze each of the three sums separately. Recall all diagonal entries of Σ in $N_p(0,\Sigma)$ are equal to 1. Let random vector

$$(w_1, w_2, \dots, w_p) \sim N_p(0,\Sigma). \tag{67}$$

Then every w_i has the distribution of $N(0,1)$.

Case (1). Evidently, (64) and (1) of (65) imply that $0 \le d_3 - d_2 < \tau$. Hence, $|\Omega_1| \le \tau p^3$. Further, for $(d,d') \in \Omega_1$, the covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is equal to

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \gamma & 0 \\ 0 & \gamma & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

for some $\gamma \in [-1,1]$. Thus, the covariance matrix of $(w_{d_2}, w_{d_1}, w_{d_3}, w_{d_4})$ is equal to

$$\begin{pmatrix} 1 & 0 & \gamma & 0 \\ 0 & 1 & 0 & 0 \\ \gamma & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Recall $Z_d = Z_{d_1,d_2} = Z_{d_2,d_1} = |\sum_{k=1}^n x_{kd_1} x_{kd_2}|$ defined at the beginning of Step 2. By Lemma 6.9, for some $\varepsilon > 0$ small enough,

$$\begin{aligned} \sum_{(d,d')\in\Omega_1} P(Z_d > a_n, Z_{d'} > a_n) &= \sum_{(d,d')\in\Omega_1} P(Z_{d_2,d_1} > a_n, Z_{d_3,d_4} > a_n) \\ &\le \tau p^3 \cdot O\left(\frac{1}{p^{4-\varepsilon}}\right) = O\left(\frac{\tau}{p^{1-\varepsilon}}\right) \to 0 \end{aligned} \tag{68}$$

as $n \to \infty$ since $\tau = o(p^t)$ for any $t > 0$.

Case (2). For any $(d,d') \in \Omega_2$, there are three possibilities.

(I): $|d_1 - d_3| < \tau$ and $|d_2 - d_4| < \tau$; (II): $|d_1 - d_3| < \tau$ and $|d_2 - d_4| \ge \tau$; (III): $|d_1 - d_3| \ge \tau$ and $|d_2 - d_4| < \tau$. The case that $|d_1 - d_3| \ge \tau$ and $|d_2 - d_4| \ge \tau$ is excluded by (64).

Let $\Omega_{2,\mathrm{I}}$ be the subset of $(d,d') \in \Omega_2$ satisfying (I), and $\Omega_{2,\mathrm{II}}$ and $\Omega_{2,\mathrm{III}}$ be defined similarly. It is easy to check that $|\Omega_{2,\mathrm{I}}| \le \tau^2 p^2$. The covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is equal to

$$\begin{pmatrix} 1 & 0 & \gamma_1 & 0 \\ 0 & 1 & 0 & \gamma_2 \\ \gamma_1 & 0 & 1 & 0 \\ 0 & \gamma_2 & 0 & 1 \end{pmatrix}$$

for some $\gamma_1, \gamma_2 \in [-1,1]$. By Lemma 6.11,

$$\sum_{(d,d')\in\Omega_{2,\mathrm{I}}} P(Z_d > a_n, Z_{d'} > a_n) = O\left(\frac{\tau^2}{p^{\varepsilon_0}}\right) \to 0 \qquad \text{as } n \to \infty. \tag{69}$$

Easily, $|\Omega_{2,\mathrm{II}}| \le \tau p^3$. The covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is

$$\begin{pmatrix} 1 & 0 & \gamma & 0 \\ 0 & 1 & 0 & 0 \\ \gamma & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad |\gamma| \le 1.$$

By Lemma 6.9, take $\varepsilon > 0$ small enough to get

$$\sum_{(d,d')\in\Omega_{2,\mathrm{II}}} P(Z_d > a_n, Z_{d'} > a_n) = O\left(\frac{\tau}{p^{1-\varepsilon}}\right) \to 0 \tag{70}$$

as $n \to \infty$.

The third case is similar to the second one. In fact, $|\Omega_{2,\mathrm{III}}| \le \tau p^3$. The covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is equal to

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \gamma \\ 0 & 0 & 1 & 0 \\ 0 & \gamma & 0 & 1 \end{pmatrix}, \qquad |\gamma| \le 1.$$

Thus, the covariance matrix of $(w_{d_2}, w_{d_1}, w_{d_4}, w_{d_3})$ is equal to Σ_4 in Lemma 6.9. Then, by the same argument as that in the equality in (68) we get

$$\sum_{(d,d')\in\Omega_{2,\mathrm{III}}} P(Z_d > a_n, Z_{d'} > a_n) = O\left(\frac{\tau}{p^{1-\varepsilon}}\right) \to 0 \tag{71}$$

as $n \to \infty$ by taking $\varepsilon > 0$ small enough. Combining (69), (70) and (71), we conclude

$$\sum_{(d,d')\in\Omega_2} P(Z_d > a_n, Z_{d'} > a_n) \to 0$$

as $n \to \infty$. This and (68) together with (66) say that, to finish the proof of this proposition, it suffices to verify

$$\sum_{(d,d')\in\Omega_3} P(Z_d > a_n, Z_{d'} > a_n) \to 0 \tag{72}$$

as $n \to \infty$. The next lemma confirms this and the proof is complete. $\square$

LEMMA 6.12. *Let the notation be as in the proof of Proposition 6.4, then (72) holds.*

The proof of this lemma is given at Cai and Jiang (2010).

PROOF OF THEOREM 4. By the first paragraph in the proof of Theorem 1, w.l.o.g., we prove the theorem by assuming that the n rows of $X_n = (x_{ij})_{1\le i\le n, 1\le j\le p}$ are i.i.d. random vectors with distribution $N_p(0, \Sigma)$ where all of the diagonal entries of Σ are equal to 1. Consequently, by the assumption on Σ, for any subset $E = \{i_1, i_2, \dots, i_m\}$ of $\{1, 2, \dots, p\}$ with $|i_s - i_t| \geq \tau$ for all $1 \leq s < t \leq m$, we know that $\{x_{ki}; 1 \leq k \leq n, i \in E\}$ are mn i.i.d. $N(0, 1)$-distributed random variables.

Reviewing the proof of Lemma 6.5, the argument is only based on the distribution of each column of $\{x_{ij}\}_{n\times p}$; the joint distribution of any two different columns are irrelevant. In current situation, the entries in each column are i.i.d. standard normals. Thus, take $\alpha = 2$ in the lemma to have

$$\begin{gathered} b_{n,3} \xrightarrow{P} 1 \qquad \text{as } n \to \infty, \\ \left\{\sqrt{\frac{n}{\log p}} b_{n,1}\right\} \quad \text{and} \quad \left\{\sqrt{\frac{n}{\log p}} b_{n,4}\right\} \text{ are tight} \end{gathered} \tag{73}$$

provided $\log p = o(n^{1/3})$, where $b_{n,1}$, $b_{n,3}$ and $b_{n,4}$ are as in Lemma 6.5. Let $V_n = V_{n,\tau} = (v_{ij})_{p\times p}$ be as in (53). It is seen from Proposition 6.4 that

$$\frac{V_{n,\tau}}{\sqrt{n \log p}} \to 2 \tag{74}$$

in probability as $n \to \infty$, $p \to \infty$ and $\log p = o(n^{1/3})$. Noticing the differences in the indices of $\max_{1\leq i<j\leq p} |\rho_{ij}|$ and $\max_{1\leq i<j\leq p, |i-j|\geq\tau} |\rho_{ij}| = L_{n,\tau}$, checking the proof of Lemma 2.2 from Jiang (2004a), it is easy to see that

$$\Delta_n := \max_{1\leq i<j\leq p, |i-j|\geq \tau} |n\rho_{ij} - v_{ij}| \leq (b_{n,1}^2 + 2b_{n,1}) V_{n,\tau} b_{n,3}^{-2} + n b_{n,3}^{-2} b_{n,4}^2. \tag{75}$$

Now, using (73), (74) and (75), replacing W_n with $V_{n,\tau}$ and L_n with $L_{n,\tau}$ in the proof of Theorem 3, and repeating the whole proof again, we obtain that $(n^2 L_{n,\tau}^2 - V_{n,\tau}^2)/n \to 0$ in probability as $n \to \infty$. This joint with Proposition 6.4 and the Slusky lemma yields the desired limiting result for $L_{n,\tau}$. $\square$

Acknowledgment. The authors thank Danning Li for helpful discussions on the proofs of the results in this paper.

SUPPLEMENTARY MATERIAL

Additional technical proofs (DOI: 10.1214/11-AOS879SUPP; .pdf). We give complete proofs for some technical lemmas used in the proofs of the main results.

REFERENCES

ACHLIOPTAS, D. (2003). Database-friendly random projections: Johnson–Lindenstrauss with binary coins. *J. Comput. Syst. Sci.* **66** 671–687.

ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2009). *An Introduction to Random Matrices*. Cambridge Univ. Press, Cambridge.

ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

BAI, Z. D., MIAO, B. Q. and PAN, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35** 1532–1572. MR2330979

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

BARANIUK, R., DAVENPORT, M., DEVORE, R. and WAKIN, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constr. Approx.* **28** 253–263. MR2453366

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

CAI, T. and JIANG, T. (2010). Supplement to "Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices." DOI:10.1214/11-AOS879SUPP.

CAI, T. T. and LV, J. (2007). Discussion: "The Dantzig selector: Statistical estimation when p is much larger than n," by E. Candes and T. Tao. *Ann. Statist.* **35** 2365–2369. MR2382647

CAI, T. T., WANG, L. and XU, G. (2010a). Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.* **58** 1300–1308. MR2730209

CAI, T. T., WANG, L. and XU, G. (2010b). Stable recovery of sparse signals and an oracle inequality. *IEEE Trans. Inform. Theory* **56** 3516–3522.

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885

CANDÈS, E. J. and PLAN, Y. (2009). Near-ideal model selection by ℓ_1 minimization. *Ann. Statist.* **37** 2145–2177. MR2543688

CANDES, E. J. and TAO, T. (2005). Decoding by linear programming. *IEEE Trans. Inform. Theory* **51** 4203–4215. MR2243152

CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n (with discussion). *Ann. Statist.* **35** 2313–2351. MR2382644

CHEN, X. (1990). Moderate deviations of B-valued independent random vectors. *Chinese Ann. Math. Ser. A* **11** 621–629. MR1092054

CHEN, X. (1991). Moderate deviations of independent random vectors in a Banach space. *Chinese J. Appl. Probab. Statist.* **7** 24–32. MR1204537

DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036

DIACONIS, P. W., EATON, M. L. and LAURITZEN, S. L. (1992). Finite de Finetti theorems in linear models and multivariate analysis. *Scand. J. Statist.* **19** 289–315. MR1211786

DONOHO, D. L. (2006a). Compressed sensing. *IEEE Trans. Inform. Theory* **52** 1289–1306. MR2241189

DONOHO, D. L. (2006b). For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.* **59** 797–829. MR2217606

DONOHO, D. L. and HUO, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory* **47** 2845–2862. MR1872845

DONOHO, D. L., ELAD, M. and TEMLYAKOV, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory* **52** 6–18. MR2237332

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 849–911. MR2530322

FAN, J. and LV, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica* **20** 101–148. MR2640659

FUCHS, J.-J. (2004). On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory* **50** 1341–1344. MR2094894

JIANG, T. (2004a). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906

JIANG, T. (2004b). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66** 35–48. MR2082906

JIANG, T. (2005). Maxima of entries of Haar distributed matrices. *Probab. Theory Related Fields* **131** 121–144. MR2105046

JIANG, T. (2006). How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.* **34** 1497–1529. MR2257653

JIANG, T. (2009). The entries of circular orthogonal ensembles. *J. Math. Phys.* **50** 063302, 13. MR2536107

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010

LEDOUX, M. (1992). Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. *Ann. Inst. H. Poincaré Probab. Statist.* **28** 267–280. MR1162575

LI, D. and ROSALSKY, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **16** 423–447. MR2209348

LI, D., LIU, W. D. and ROSALSKY, A. (2009). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148** 5–35.

LIGERALDE, A. and BROWN, B. (1995). Band covariance matrix estimation using restricted residuals: A Monte Carlo analysis. *Internat. Econom. Rev.* **36** 751–767.

LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539

PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670

SAKHANENKO, A. I. (1991). Estimates of Berry–Esseen type for the probabilities of large deviations. *Sibirsk. Mat. Zh.* **32** 133–142, 228. MR1142075

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345–5363. MR2327033

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: tcai@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~tcai

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
224 CHURCH STREET
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: tjiang@stat.umn.edu
URL: http://www.stat.umn.edu/~tjiang