

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

Tony Cai¹ and Tiefeng Jiang²

University of Pennsylvania and University of Minnesota

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.

AMS 2000 Subject Classification: Primary 62H12, 60F05; secondary 60F15, 62H10.

¹Statistics Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, tcai@wharton.upenn.edu. The research of Tony Cai was supported in part by NSF FRG Grant DMS-0854973.

²School of Statistics, University of Minnesota, 224 Church Street, MN55455, tjiang@stat.umn.edu. The research of Tiefeng Jiang was supported in part by NSF FRG Grant DMS-0449365.

Contents

1. Introduction
 - 1.1. Limiting Laws of the Coherence of a Random Matrix
 - 1.2. Testing Covariance Structure
 - 1.3. Construction of Compressed Sensing Matrices
 - 1.4. Organization of the Paper
2. Limiting Laws of Coherence of Random Matrix
 - 2.1. The i.i.d. Case
 - 2.2. The Dependent Case
3. Testing the Covariance Structure
4. Construction of Compressed Sensing Matrices
5. Discussion and Comparison with Related Results
- 6 Proofs
 - 6.1. Technical Tools
 - 6.2. Proofs of Theorems 1 and 2
 - 6.3. Proof of Theorem 3
 - 6.4. Proof of Theorem 4
- 7 Appendix

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 and 2008), Bai, Miao and Pan (2007), and Jiang (2004b). For general background on the random matrix theory, see, for example, Bai and Silverstein (2009) and Anderson, Guionnet, and Zeitouni (2009).

In statistics, the 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, Candès and Tao (2007), Cai, Wang and Xu (2010a), Bai and Saranadasa (1996), Bai, Jiang, Yao and Zheng (2009), and Cai, Zhang and Zhou (2010).

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

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

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

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

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,

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

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, i.e., $n/p \rightarrow \gamma \in (0, \infty)$, asymptotic properties of the coherence L_n of 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,

$$\sqrt{n/\log p} L_n \xrightarrow{P} 2 \quad \text{as } n \rightarrow \infty$$

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

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

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 $X_n = (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$, where the n rows are i.i.d. random vectors with distribution $N_p(\mu, \Sigma)$. For a given integer $\tau \geq 1$ (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,

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau. \quad (4)$$

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

$$L_{n,\tau} = \max_{|i-j| \geq \tau} |\rho_{ij}| \quad (5)$$

when the mean μ is assumed to be unknown and define

$$\tilde{L}_{n,\tau} = \max_{|i-j| \geq \tau} |\tilde{\rho}_{ij}| \quad (6)$$

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

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

We shall derive in Section 2 the limiting distribution of $L_{n,\tau}$ and $\tilde{L}_{n,\tau}$ under the null hypothesis H_0 and discuss its application in Section 3. The study for this case is considerably more difficult technically 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, i.e., $n/p \rightarrow \gamma \in (0, \infty)$, 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 P  ch   (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), i.e., 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 $L_{n,\tau}$ developed in the present paper can be readily 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, e.g., Candes and Tao (2007), Bickel, Ritov and Tsybakov (2009), and Candes 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 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, b). 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, e.g., 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 near-optimality properties and 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 of random matrices given in this paper 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 problem 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 of our results with other related work. The main results are proved in Section 6 and the proofs of technical lemmas are given in the Appendix.

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 we consider the limiting distribution of $L_{n,\tau}$ and $\tilde{L}_{n,\tau}$ defined in (5) and (6). We then apply the asymptotic results 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 $\{\xi, x_{ij}, i, j = 1, 2, \dots\}$ are i.i.d. real random variables with mean μ and variance $\sigma^2 > 0$. Let $X_n = (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$ and let x_1, x_2, \dots, x_p be the p columns of X_n . Then $X_n = (x_1, x_2, \dots, x_p)$. 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 $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$. 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 $1 \leq i \leq p$. 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 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 and 2010). We now derive the limiting laws of L_n and \tilde{L}_n .

We shall introduce another quantity that is useful for our technical analysis. Define

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

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

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 $p = p(n) \rightarrow \infty$ and $\log p = o(n^\beta)$ as $n \rightarrow \infty$. Then $\sqrt{n/\log p} L_n \rightarrow 2$ in probability as $n \rightarrow \infty$.*

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 $\beta \rightarrow 1$, 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 $0 < \alpha \leq 2$ and $t_0 > 0$. Set $\beta = \alpha/(4 + \alpha)$. Assume $p = p(n) \rightarrow \infty$ and $\log p = o(n^\beta)$ as $n \rightarrow \infty$. Then $nL_n^2 - 4 \log p + \log \log p$ converges weakly to an extreme distribution of type I with distribution function*

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

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, i.e., $n/p \rightarrow \gamma \in (0, \infty)$, 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, e.g., Li and Rosalsky (2006), Zhou (2007), and Li, Liu and Rosalsky (2009) where the same condition $n/p \rightarrow \gamma \in (0, \infty)$ was imposed. Liu, Lin and Shao (2008) showed that the same results hold for $p \rightarrow \infty$ 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 $p \leq e^{n^\beta}$ for a certain $\beta > 0$, we obtain the strong laws and limiting distributions of the coherence of 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 $n/p \rightarrow \gamma \in (0, \infty)$ and $E|\xi|^{30+\epsilon} < \infty$ for some $\epsilon > 0$, then for any $y \in \mathbb{R}$,

$$P(nL_n^2 - 4 \log n + \log \log n \leq y) \rightarrow e^{-Ke^{-y/2}} \quad (10)$$

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

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

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

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

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 random matrix X_n are drawn independently from a multivariate Gaussian distribution. Let $X_n = (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$, where the n rows are i.i.d. random vectors with distribution $N_p(\mu, \Sigma)$, where $\mu \in \mathbb{R}^p$ is arbitrary 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,

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau \quad (11)$$

for a given integer $\tau \geq 1$. In order to construct a test, we study in this section the asymptotic distributions of $L_{n,\tau}$ 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 $0 < \delta < 1$, set

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

THEOREM 4 *Suppose, as $n \rightarrow \infty$,*

(i) $p = p_n \rightarrow \infty$ 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

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

Similar to J_n in (9), we define

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

where we write $x_i - \mu_i$ for $x_i - \mu_i e$ with $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$, $\mu = (\mu_1, \dots, \mu_p)^T$ 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 $\sqrt{\frac{n}{\log p}} L_{n,\tau} \rightarrow 2$ in probability as $n \rightarrow \infty$.*

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. Observe

$$\left\{ (i, j); 1 \leq i < j \leq 2n, |i - j| \geq n \right\} = n + (n - 1) + \dots + 1 = \frac{n(n + 1)}{2} \sim \frac{p^2}{8}$$

as $n \rightarrow \infty$. So $L_{n,\tau}$ is the maximum of roughly $p^2/8$ random variables, and the dependence of any two of such random variables are less than that appeared in L_n in Theorem 3. The result in Theorem 3 can be rewritten as

$$nL_n^2 - 2 \log \frac{p^2}{2} + \log \log \frac{p^2}{2} - \log 8 \quad \text{converges weakly to } F$$

as $n \rightarrow \infty$. Recalling L_n is the maximum of roughly $p^2/2$ weakly dependent random variables, replace L_n with $L_{n,\tau}$ and $p^2/2$ with $p^2/8$ to have $nL_{n,\tau}^2 - 2 \log \frac{p^2}{8} + \log \log \frac{p^2}{8} - \log 8$ converges weakly to F , where F is as in Theorem 3. That is,

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

as $n \rightarrow \infty$ (This can be done rigorously by following the proof of Theorem 3). The difference between (14) and Theorem 4 is evident.

REMARK 2.4 Let $p = mn$ with integer $m \geq 2$. We consider the $p \times p$ matrix $\Sigma = \text{diag}(H_n, \dots, H_n)$ 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. Thus, if $(\zeta_1, \dots, \zeta_p) \sim N_p(0, \Sigma)$, then $\zeta_{ln+1} = \zeta_{ln+2} = \dots = \zeta_{(l+1)n}$ for all $0 \leq l \leq m-1$ and $\zeta_1, \zeta_{n+1}, \dots, \zeta_{(m-1)n+1}$ are i.i.d. $N(0, 1)$ -distributed random variables. Let $\{\zeta_{ij}; 1 \leq i \leq n, 1 \leq j \leq m\}$ be i.i.d. $N(0, 1)$ -distributed random variables. Then

$$\left(\underbrace{(\zeta_{i1}, \dots, \zeta_{i1})}_n, \underbrace{(\zeta_{i2}, \dots, \zeta_{i2})}_n, \dots, \underbrace{(\zeta_{im}, \dots, \zeta_{im})}_n \right)' \in \mathbb{R}^p, \quad 1 \leq i \leq n,$$

are i.i.d. random vectors with distribution $N_p(0, \Sigma)$. Denote the corresponding data matrix by $(x_{ij})_{n \times p}$. Now, take $\tau = n$ and $m = \lceil e^{n^{1/4}} \rceil$. 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. Obviously,

$$L_{n,\tau} = \max_{1 \leq i < j \leq p, |i-j| \geq \tau} |\rho_{ij}| = \max_{1 \leq i < j \leq m} |\hat{\rho}_{ij}|,$$

where $\hat{\rho}_{ij}$ is obtained from $(\zeta_{ij})_{n \times m}$ as in (1) (note that the mn entries of $(\zeta_{ij})_{n \times m}$ are i.i.d. with distribution $N(0, 1)$). By Theorem 3 on $\max_{1 \leq i < j \leq m} |\hat{\rho}_{ij}|$, we have that $nL_{n,\tau}^2 - 4 \log m + \log \log m$ converges weakly to F , which is the same as the F in Theorem 4. Set $\log_2 x = \log \log x$ for $x > 1$. Notice

$$\begin{aligned} nL_{n,\tau}^2 - 4 \log m + \log_2 m &= nL_{n,\tau}^2 - 4 \log p + 4 \log n + \log_2 m \\ &\sim (nL_{n,\tau}^2 - 4 \log p + \log_2 p) + 4 \log n \end{aligned}$$

since $p = mn$ and $\log_2 p - \log_2 m \rightarrow 0$. Further, it is easy to check that $4 \log n - 16 \log_2 p \rightarrow 0$. Therefore, the previous conclusion is equivalent to that

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

as $n \rightarrow \infty$. 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 $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ 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

$$H_0 : \sigma_{i,j} = 0 \text{ for all } |i - j| \geq \tau \text{ versus } H_a : \sigma_{i,j} \neq 0 \text{ for some } |i - j| \geq \tau. \quad (16)$$

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

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

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

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

That is, we reject the null hypothesis H_0 whenever

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

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 λ_{\max} of the sample correlation matrix. However, the limiting distribution of the largest eigenvalue λ_{\max} is unknown even for the case $p/n \rightarrow c$, a finite and 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 $n \times p$ random matrix X whose entries $x_{i,j}$ are iid normal variables

$$x_{i,j} \stackrel{iid}{\sim} N(0, n^{-1}). \quad (19)$$

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

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

or more sparsely by

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

These random matrices are shown to satisfy the RIP conditions with high probability. See Achlioptas (2001) 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

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

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, e.g., 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 Theorem 1 or Theorem 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 Section 6.

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 $P(\tilde{L}_n \geq t) \leq 3p^2 e^{-ng(t)}$ where $g(t) = \min\{I_1(t/2), I_2(1/2)\} > 0$ for any $t > 0$ and*

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

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

We now consider the three particular random matrices mentioned in 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

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

for all $n \geq 2$ and $k \geq 1$. The verification of this example together with the next two are given in the Appendix.

Example 2. Let x_{11} be such that $P(x_{11} = \pm 1/\sqrt{n}) = 1/2$ as in (20). In this case, we have

$$P\left((2k-1)\tilde{L}_n < 1\right) \geq 1 - 3p^2 \exp\left\{-\frac{n}{12(2k-1)^2}\right\} \quad (24)$$

for all $n \geq 2$ and $k \geq 1$.

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

$$P\left((2k-1)\tilde{L}_n < 1\right) \geq 1 - 3p^2 \exp\left\{-\frac{n}{12(2k-1)^2}\right\} \quad (25)$$

for all $n \geq 2$ and $k \geq 2$.

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 6.8.

REMARK 4.2 Here we would like to point out an error on pp. 801 of Donoho (2006b) and pp. 2147 of Candes 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, e.g., 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 (2004), 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 $n \rightarrow \infty$, 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 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. 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 6.10, 6.11 and 6.12.

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 $c_1 n^\alpha \leq p \leq c_2 n^\alpha$ 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 and 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 $\|A\| = \max_{1 \leq i \neq j \leq n} |a_{i,j}|$; 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 the Appendix.

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

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

where

$$b_{n,1} = \max_{1 \leq i \leq p} |h_i - 1|, \quad W_n = \max_{1 \leq i < j \leq p} |x_i^T x_j|, \quad b_{n,3} = \min_{1 \leq i \leq p} h_i, \quad b_{n,4} = \max_{1 \leq i \leq p} |\bar{x}_i|.$$

The following Poisson approximation result is essentially a special case of Theorem 1 from Arratia et al. (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

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

where

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t)P(\eta_\beta > t), \\ b_2 &= \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} P(\eta_\alpha > t, \eta_\beta > t), \\ b_3 &= \sum_{\alpha \in I} E|P(\eta_\alpha > t | \sigma(\eta_\beta, \beta \notin B_\alpha)) - P(\eta_\alpha > t)|, \end{aligned}$$

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 et al (2008).

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

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

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

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

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

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 ξ_1, ξ_2, \dots are i.i.d. r.v.'s with $E\xi_1 = 0$ and $E\xi_1^2 = 1$. Put $S_n = \sum_{i=1}^n \xi_i$.*

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

$$\lim_{n \rightarrow \infty} \frac{1}{a_n^2} \log P\left(\frac{S_n}{\sqrt{na_n}} \geq u\right) = -\frac{u^2}{2} \quad (26)$$

for any $u > 0$.

(ii) Let $0 < \alpha < 1$ and $\{a_n; n \geq 1\}$ satisfy that $a_n \rightarrow +\infty$ and $a_n = O(n^{\frac{\alpha}{2(2-\alpha)}})$. If $Ee^{t|\xi_1|^\alpha} < \infty$ for all $t > 0$, then (26) also holds.

6.2 Proofs of Theorems 1 and 2

Recall that a sequence of random variables $\{X_n; n \geq 1\}$ are said to be *tight* if, for any $\epsilon > 0$, there is a constant $K > 0$ such that $\sup_{n \geq 1} P(|X_n| \geq K) < \epsilon$. Obviously, $\{X_n; n \geq 1\}$ are tight if for some $K > 0$, $\lim_{n \rightarrow \infty} P(|X_n| \geq K) \rightarrow 0$. It is easy to check that

if $\{X_n; n \geq 1\}$ are tight, then for any sequence of constants $\{\epsilon_n; n \geq 1\}$ with $\lim_{n \rightarrow \infty} \epsilon_n = 0$, we have $\epsilon_n X_n \rightarrow 0$ in probability as $n \rightarrow \infty$. (27)

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

LEMMA 6.5 *Let $\{x_{ij}; i \geq 1, j \geq 1\}$ be i.i.d. random variables with $Ex_{11} = 0$ and $Ex_{11}^2 = 1$. Then, $b_{n,3} \rightarrow 1$ in probability as $n \rightarrow \infty$, 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) $|x_{11}| \leq C$ for some constant $C > 0$, $p_n \rightarrow \infty$ and $\log p_n = o(n)$ as $n \rightarrow \infty$;*
- (ii) $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $0 < \alpha \leq 2$ and $t_0 > 0$, and $p_n \rightarrow \infty$ and $\log p_n = o(n^\beta)$ as $n \rightarrow \infty$, where $\beta = \alpha/(4 - \alpha)$.*

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

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

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

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

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

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

Lemmas 6.5, 6.6, and 6.7 are proved in the Appendix.

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 $W_n = \max_{1 \leq i < j \leq p} |x_i^T x_j| = \max_{1 \leq i < j \leq p} \left| \sum_{k=1}^n x_{ki} x_{kj} \right|$. Then

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

in probability as $n \rightarrow \infty$.

Proof. We first prove

$$\lim_{n \rightarrow \infty} P \left(\frac{W_n}{\sqrt{n \log p}} \geq 2 + 2\epsilon \right) = 0 \quad (28)$$

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

$$P(W_n \geq (2 + 2\epsilon) \sqrt{n \log p}) \leq \binom{p}{2} \cdot P \left(\left| \sum_{k=1}^n x_{k1} x_{k2} \right| \geq (2 + 2\epsilon) \sqrt{n \log p} \right) \quad (29)$$

for any $\epsilon > 0$. Notice $E(|x_{11} x_{12}|^2) = E(|x_{11}|^2) \cdot E(|x_{12}|^2) = 1$. By (i) of Lemma 6.4, using conditions $E e^{|x_{11} x_{12}|} < \infty$ and $\log p = o(n)$ as $n \rightarrow \infty$, we obtain

$$P \left(\left| \sum_{k=1}^n x_{k1} x_{k2} \right| \geq (2 + 2\epsilon) \sqrt{n \log p} \right) \leq \exp \left(-\frac{(2 + \epsilon)^2}{2} \log p \right) \leq \frac{1}{p^{2 + \epsilon}} \quad (30)$$

as n is sufficiently large. The above two assertions conclude

$$P(W_n \geq (2 + 2\epsilon)\sqrt{n \log p}) \leq \frac{1}{p^\epsilon} \rightarrow 0 \quad (31)$$

as $n \rightarrow \infty$. Thus (28) holds. Now, to finish the proof, we only need to show

$$\lim_{n \rightarrow \infty} P\left(\frac{W_n}{\sqrt{n \log p}} \leq 2 - \epsilon\right) = 0 \quad (32)$$

for any $\epsilon > 0$ small enough.

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

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

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

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

$$P(W_n \leq a_n) \leq e^{-\lambda_n} + b_{1,n} + b_{2,n} \quad (33)$$

where

$$\lambda_n = \frac{p(p-1)}{2} P(A_{12}), \quad b_{1,n} \leq 2p^3 P(A_{12})^2 \quad \text{and} \quad b_{2,n} \leq 2p^3 P(A_{12}A_{13}). \quad (34)$$

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 $n \rightarrow \infty$, we know

$$\lim_{n \rightarrow \infty} \frac{1}{\log p} \log P(A_{12}) = -\frac{(2 - \epsilon)^2}{2} \quad (35)$$

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

$$\frac{1}{p^{2-\epsilon}} \leq P(A_{12}) \leq \frac{1}{p^{2-2\epsilon}} \quad (36)$$

as n is sufficiently large. This implies

$$e^{-\lambda_n} \leq e^{-p^\epsilon/3} \quad \text{and} \quad b_{1,n} \leq \frac{2}{p^{1-4\epsilon}} \quad (37)$$

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

$$\begin{aligned} P(A_{12}A_{13}) &= P(|y_{12}^{(n)}| > a_n, |y_{13}^{(n)}| > a_n) \\ &= E\{P^1(|\sum_{k=1}^n x_{k1}x_{k2}| > a_n)^2\} \end{aligned} \quad (38)$$

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

$$P(A_{12}A_{13}) \leq \frac{1}{p^{4-4\epsilon}} \quad (39)$$

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

$$b_{2,n} \leq 2p^3 P(A_{12}A_{13}) \leq \frac{2}{p^{1-4\epsilon}} \rightarrow 0 \quad (40)$$

as $n \rightarrow \infty$. This together with (33) and (37) concludes (32). \blacksquare

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

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

in probability as $n \rightarrow \infty$.

The proof of Proposition 6.2 is similar to that of Proposition 6.1. Details are given in the Appendix.

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 $\{x_{ij}; 1 \leq i \leq n, 1 \leq j \leq p\}$ are i.i.d. random variables with mean zero and variance 1.

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

$$\frac{W_n}{\sqrt{n \log p}} \rightarrow 2 \quad (41)$$

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

$$\frac{nL_n - W_n}{\sqrt{n \log p}} \rightarrow 0 \quad (42)$$

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

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

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

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

which concludes (42) by (27). \blacksquare

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. \blacksquare

Proof of Proposition 4.1. Recall the definition of \tilde{L}_n in (3), to prove the conclusion, w.l.o.g., we assume $\mu = 0$ and $\sigma^2 = 1$. Evidently, by the i.i.d. assumption,

$$\begin{aligned} P(\tilde{L}_n \geq t) &\leq \frac{p^2}{2} P\left(\frac{|x'_1 x_2|}{\|x_1\| \cdot \|x_2\|} \geq t\right) \\ &\leq \frac{p^2}{2} P\left(\frac{|x'_1 x_2|}{n} \geq \frac{t}{2}\right) + \frac{p^2}{2} \cdot 2P\left(\frac{\|x_1\|^2}{n} \leq \frac{1}{2}\right) \end{aligned} \quad (43)$$

where the event $\{\|x_{11}\|^2/n > 1/2, \|x_{12}\|^2/n > 1/2\}$ and its complement are used to get the last inequality. Since $\{x_{ij}; i \geq 1, j \geq 1\}$ are i.i.d., the condition $Ee^{t_0|x_{11}|^2} < \infty$ implies $Ee^{t'_0|x_{11}x_{12}|} < \infty$ for some $t'_0 > 0$. By the Chernoff bound (see, e.g., p. 27 from Dembo and Zeitouni (1998)) and noting that $E(x_{11}x_{12}) = 0$ and $Ex_{11}^2 = 1$, we have

$$P\left(\frac{|x'_1 x_2|}{n} \geq \frac{t}{2}\right) \leq 2e^{-nI_1(t/2)} \text{ and } P\left(\frac{\|x_1\|^2}{n} \leq \frac{1}{2}\right) \leq 2e^{-nI_2(1/2)}$$

for any $n \geq 1$ and $t > 0$, where the following facts about rate functions $I_1(x)$ and $I_2(y)$ are used:

- (i) $I_1(x) = 0$ if and only if $x = 0$; $I_2(y) = 0$ if and only if $y = 1$;
- (ii) $I_1(x)$ is non-decreasing on $A := [0, \infty)$ and non-increasing on A^c . This is also true for $I_2(y)$ with $A = [1, \infty)$.

These and (43) conclude

$$P(\tilde{L}_n \geq t) \leq p^2 e^{-nI_1(t/2)} + 2p^2 e^{-nI_2(1/2)} \leq 3p^2 e^{-ng(t)}$$

where $g(t) = \min\{I_1(t/2), I_2(1/2)\}$ for any $t > 0$. Obviously, $g(t) > 0$ for any $t > 0$ from (i) and (ii) above. \blacksquare

LEMMA 6.8 *Let Z be a random variable with $EZ = 0$, $EZ^2 = 1$ and $Ee^{t_0|Z|} < \infty$ for some $t_0 > 0$. Choose $\alpha > 0$ such that $E(Z^2 e^{\alpha|Z|}) \leq 3/2$. Set $I(x) = \sup_{t \in \mathbb{R}} \{tx - \log Ee^{tZ}\}$. Then $I(x) \geq x^2/3$ for all $0 \leq x \leq 3\alpha/2$.*

Proof. By the Taylor expansion, for any $x \in \mathbb{R}$, $e^x = 1 + x + \frac{x^2}{2}e^{\theta x}$ for some $\theta \in [0, 1]$. It follows from $EZ = 0$ that

$$Ee^{tZ} = 1 + \frac{t^2}{2}E(Z^2 e^{\theta tZ}) \leq 1 + \frac{t^2}{2}E(Z^2 e^{t|Z|}) \leq 1 + \frac{3}{4}t^2$$

for all $0 \leq t \leq \alpha$. Use the inequality $\log(1+x) \leq x$ for all $x > -1$ to see that $\log Ee^{tZ} \leq 3t^2/4$ for every $0 \leq t \leq \alpha$. Take $t_0 = 2x/3$ with $x > 0$. Then $0 \leq t_0 \leq \alpha$ for all $0 \leq x \leq 3\alpha/2$. It follows that

$$I(x) \geq t_0x - \frac{3}{4}t_0^2 = \frac{x^2}{3}. \quad \blacksquare$$

6.3 Proof of Theorem 3

LEMMA 6.9 *Let ξ_1, \dots, ξ_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 $0 < \alpha \leq 1$. Put $S_n = \sum_{i=1}^n \xi_i$ and $\beta = \alpha/(2 + \alpha)$. Then, for any $\{p_n; n \geq 1\}$ with $0 < p_n \rightarrow \infty$ and $\log p_n = o(n^\beta)$ and $\{y_n; n \geq 1\}$ with $y_n \rightarrow y > 0$,*

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

as $n \rightarrow \infty$.

PROPOSITION 6.3 *Let $\{x_{ij}; i \geq 1, j \geq 1\}$ 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 $0 < \alpha \leq 2$ and $t_0 > 0$. Set $\beta = \alpha/(4 + \alpha)$. Assume $p = p(n) \rightarrow \infty$ and $\log p = o(n^\beta)$ as $n \rightarrow \infty$. Then*

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

as $n \rightarrow \infty$ 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

$$P\left(\max_{1 \leq i < j \leq p} |y_{ij}| \leq \sqrt{\alpha_n + nz}\right) \rightarrow e^{-Ke^{-z/2}}, \quad (44)$$

where $y_{ij} = \sum_{k=1}^n x_{ki}x_{kj}$. We now apply Lemma 6.2 to prove (44). Take $I = \{(i, j); 1 \leq i < j \leq p\}$. For $u = (i, j) \in I$, set $X_u = |y_{ij}|$ and $B_u = \{(k, l) \in I; \text{one of } k \text{ and } l = i \text{ or } j, \text{ but } (k, l) \neq u\}$. 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,

$$|P(W_n \leq a_n) - e^{-\lambda_n}| \leq b_{1,n} + b_{2,n} \quad (45)$$

where

$$\lambda_n = \frac{p(p-1)}{2} P(A_{12}), \quad b_{1,n} \leq 2p^3 P(A_{12})^2 \quad \text{and} \quad b_{2,n} \leq 2p^3 P(A_{12}A_{13}). \quad (46)$$

We first calculate λ_n . Write

$$\lambda_n = \frac{p^2 - p}{2} P\left(\frac{|y_{12}|}{\sqrt{n}} > \sqrt{\frac{\alpha_n}{n} + z}\right) \quad (47)$$

and $y_{12} = \sum_{i=1}^n \xi_i$, where $\{\xi_i; 1 \leq i \leq n\}$ 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 $\alpha_1 := \alpha/2 \leq 1$. We then have

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

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 $y_n \rightarrow 2$ as $n \rightarrow \infty$. By Lemma 6.9,

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

as $n \rightarrow \infty$. Considering $E x_{ij} = 0$, it is easy to see that the above also holds if y_{12} is replaced by $-y_{12}$. These and (47) imply that

$$\lambda_n \sim \frac{p^2 - p}{2} \cdot 2 \cdot \frac{e^{-z/2}}{\sqrt{8\pi}} \cdot \frac{1}{p^2} \sim \frac{e^{-z/2}}{\sqrt{8\pi}} \quad (48)$$

as $n \rightarrow \infty$.

Recall (45) and (46), to complete the proof, we have to verify that $b_{1,n} \rightarrow 0$ and $b_{2,n} \rightarrow 0$ as $n \rightarrow \infty$. By (46), (47) and (48),

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

as $n \rightarrow \infty$. Also, by (46),

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

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

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 Slutsky lemma, it suffices to show

$$\frac{n^2 L_n^2 - W_n^2}{n} \rightarrow 0 \quad (49)$$

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

$$|n^2 L_n^2 - W_n^2| = |nL_n - W_n| \cdot |nL_n + W_n| \leq \Delta_n \cdot (\Delta_n + 2W_n). \quad (50)$$

It is easy to see from Proposition 6.3 that

$$\frac{W_n}{\sqrt{n \log p}} \rightarrow 2 \quad (51)$$

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

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

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

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

which combining with (51) yields that

$$\left\{ \frac{\Delta_n}{\log p} \right\} \text{ is tight.} \quad (52)$$

This and (51) 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 (50) and then (27),

$$\begin{aligned} \frac{|n^2 L_n^2 - W_n^2|}{n} &\leq \frac{(\log p) \Delta'_n \left\{ (\log p) \Delta'_n + 2\sqrt{n \log p} W'_n \right\}}{n} \\ &\leq 2\sqrt{\frac{(\log p)^3}{n}} \left(\sqrt{\frac{\log p}{n}} \Delta'_n + W'_n \right) \rightarrow 0 \end{aligned} \quad (53)$$

in probability as $n \rightarrow \infty$ since $\log p = o(n^{1/3})$. This gives (49). \blacksquare

6.4 Proof of Theorem 4

We begin to prove the Theorem 4 by stating three technical lemmas which are proved in the Appendix.

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

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

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

$$\sup_{|r| \leq 1} P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n, \left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) = O\left(\frac{1}{p^{4-\epsilon}}\right) \quad (54)$$

for any $\epsilon > 0$.

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

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

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

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

for any $\epsilon > 0$.

LEMMA 6.12 Let $\{(u_{k1}, u_{k2}, u_{k3}, u_{k4})^T; 1 \leq i \leq n\}$ be a sequence of i.i.d. random vectors with distribution $N_4(0, \Sigma_4)$ where

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

Set $a_n = (4n \log p - n \log(\log p) + ny)^{1/2}$ for $n \geq e^e$ and $y \in \mathbb{R}$. Suppose $n \rightarrow \infty$, $p \rightarrow \infty$ with $\log p = o(n^{1/3})$. Then, for any $\delta \in (0, 1)$, there exists $\epsilon_0 = \epsilon(\delta) > 0$ such that

$$\sup_{|r_1|, |r_2| \leq 1-\delta} P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n, \left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) = O\left(p^{-2-\epsilon_0}\right) \quad (55)$$

as $n \rightarrow \infty$.

Recall notation τ , $\Sigma = (\sigma_{ij})_{p \times p}$ and $X_n = (x_{ij})_{n \times p} \sim N_p(\mu, \Sigma)$ above (11).

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

$$V_n = V_{n,\tau} = \max_{1 \leq i < j \leq p, |j-i| \geq \tau} |x_i^T x_j|. \quad (56)$$

Suppose $n \rightarrow \infty$, $p = p_n \rightarrow \infty$ 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 $n \rightarrow \infty$. Then, under H_0 in (11),

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

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

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

$$\Lambda_p = \left\{ (i, j); 1 \leq i < j \leq p, j - i \geq \tau, \max_{1 \leq k \neq i \leq p} \{|r_{ik}|\} \leq 1 - \delta, \max_{1 \leq k \neq j \leq p} \{|r_{jk}|\} \leq 1 - \delta \right\},$$

$$V'_n = \max_{(i,j) \in \Lambda_p} \left| \sum_{k=1}^n x_{ki} x_{kj} \right|. \quad (57)$$

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

$$\lim_{n \rightarrow \infty} P(V'_n \leq a_n) = e^{-Ke^{-y/2}} \quad (58)$$

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

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

$$\lim_{n \rightarrow \infty} P(V_n > a_n) = 1 - e^{-Ke^{-y/2}} \quad (59)$$

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

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

where the sum runs over all pair (i, j) such that $1 \leq i < j \leq p$ and one of i and j is in $\Gamma_{p,\delta}$. Note that $|x_{11} x_{1\tau+1}| \leq (x_{11}^2 + x_{1\tau+1}^2)/2$, it follows that $Ee^{|x_{11} x_{1\tau+1}|/2} < \infty$ by independence and $E \exp(N(0, 1)^2/4) < \infty$. Since $\{x_{k1}, x_{k\tau+1}; 1 \leq k \leq n\}$ are i.i.d. with mean zero and variance one, and $y_n := a_n/\sqrt{n \log p} \rightarrow 2$ as $n \rightarrow \infty$, taking $\alpha = 1$ in Lemma 6.9, we get

$$\begin{aligned} & P\left(\frac{1}{\sqrt{n \log p}} \left| \sum_{k=1}^n x_{k1} x_{k\tau+1} \right| > \frac{a_n}{\sqrt{n \log p}}\right) \\ & \sim 2 \cdot \frac{p^{-y_n^2/2} (\log p)^{-1/2}}{2\sqrt{2\pi}} \sim \frac{e^{-y/2}}{\sqrt{2\pi}} \cdot \frac{1}{p^2} \end{aligned} \quad (60)$$

as $n \rightarrow \infty$. Moreover, note that the total number of such pairs is no more than $2p|\Gamma_{p,\delta}|$. Therefore,

$$\begin{aligned} P(V'_n > a_n) & \leq P(V_n > a_n) \leq P(V'_n > a_n) + 2p|\Gamma_{p,\delta}| \cdot P\left(\left| \sum_{k=1}^n x_{k1} x_{k\tau+1} \right| > a_n\right) \\ & \leq P(V'_n > a_n) + o(p^2) \cdot O\left(\frac{1}{p^2}\right) \end{aligned} \quad (61)$$

by the assumption on $\Gamma_{p,\delta}$ and (60). Thus, this joint with (59) gives (58).

Step 2. We now apply Lemma 6.2 to prove (58). Take $I = \Lambda_p$. For $(i, j) \in I$, set $Z_{ij} = |\sum_{k=1}^n x_{ki}x_{kj}|$,

$$B_{i,j} = \{(k, l) \in \Lambda_p; |s - t| < \tau \text{ for some } s \in \{k, l\} \text{ and some } t \in \{i, j\}, \text{ but } (k, l) \neq (i, j)\},$$

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

It is easy to see that $|B_{i,j}| \leq 2 \cdot (2\tau + 2\tau)p = 8\tau p$ 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,

$$|P(V_n \leq a_n) - e^{-\lambda_n}| \leq b_{1,n} + b_{2,n} \quad (62)$$

where

$$\lambda_n = |\Lambda_p| \cdot P(A_{1\tau+1}), \quad b_{1,n} \leq \sum_{d \in \Lambda_p} \sum_{d' \in B_a} P(A_{12})^2 = 8\tau p^3 P(A_{1\tau+1})^2 \text{ and} \quad (63)$$

$$b_{2,n} \leq \sum_{d \in \Lambda_p} \sum_{d \neq d' \in B_a} P(Z_d > t, Z_{d'} > t) \quad (64)$$

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

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

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

$$|\Lambda_p| \sim \frac{p^2}{2} \quad (65)$$

as $n \rightarrow \infty$. It then follows from (60) that

$$\lambda_n \sim \frac{p^2}{2} \cdot \frac{e^{-y/2}}{\sqrt{2\pi}} \cdot \frac{1}{p^2} \sim \frac{e^{-y/2}}{\sqrt{8\pi}} \quad (66)$$

as $n \rightarrow \infty$.

Recall (62) and (66), to complete the proof, we have to verify that $b_{1,n} \rightarrow 0$ and $b_{2,n} \rightarrow 0$ as $n \rightarrow \infty$. Clearly, by the first expression in (63), we get from (66) and then (65) that

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

as $n \rightarrow \infty$ 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 (64) 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

$$d, d' \in \Lambda_p, d \neq d', d_1 \leq d_3 \text{ and } |d_i - d_j| < \tau \text{ for some } i \in \{1, 2\} \text{ and some } j \in \{3, 4\}. \quad (67)$$

Geometrically, there are three cases for the locations of $d = (d_1, d_2)$ and $d' = (d_3, d_4)$:

$$(1) d_2 \leq d_3; \quad (2) d_1 \leq d_3 < d_4 \leq d_2; \quad (3) d_1 \leq d_3 \leq d_2 \leq d_4. \quad (68)$$

Let Ω_j be the subset of index (d, d') with restrictions (67) and (j) for $j = 1, 2, 3$. Then

$$b_{2,n} \leq 2 \sum_{i=1}^3 \sum_{(d,d') \in \Omega_i} P(Z_d > a_n, Z_{d'} > a_n). \quad (69)$$

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

$$(w_1, w_2, \dots, w_p) \sim N_p(0, \Sigma). \quad (70)$$

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

Case (1). Evidently, (67) and (1) of (68) imply that $0 \leq d_3 - d_2 < \tau$. Hence, $|\Omega_1| \leq \tau p^3$. 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 6.10, for some $\epsilon > 0$ small enough,

$$\begin{aligned} \sum_{(d,d') \in \Omega_1} P(Z_d > a_n, Z_{d'} > a_n) &= \sum_{(d,d') \in \Omega_1} P(Z_{d_2, d_1} > a_n, Z_{d_3, d_4} > a_n) \\ &\leq \tau p^3 \cdot O\left(\frac{1}{p^{4-\epsilon}}\right) = O\left(\frac{\tau}{p^{1-\epsilon}}\right) \rightarrow 0 \end{aligned} \quad (71)$$

as $n \rightarrow \infty$ 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 $|d_2 - d_4| \geq \tau$; (III): $|d_1 - d_3| \geq \tau$ and $|d_2 - d_4| < \tau$. The case that $|d_1 - d_3| \geq \tau$ and $|d_2 - d_4| \geq \tau$ is excluded by (67).

Let $\Omega_{2,I}$ be the subset of $(d, d') \in \Omega_2$ satisfying (I), and $\Omega_{2,II}$ and $\Omega_{2,III}$ be defined similarly. It is easy to check that $|\Omega_{2,I}| \leq \tau^2 p^2$. 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 6.12,

$$\sum_{(d, d') \in \Omega_{2,I}} P(Z_d > a_n, Z_{d'} > a_n) = O\left(\frac{\tau^2}{p^{\epsilon_0}}\right) \rightarrow 0 \quad (72)$$

as $n \rightarrow \infty$.

Observe $|\Omega_{2,II}| \leq \tau p^3$. The covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is equal to

$$\begin{pmatrix} 1 & 0 & \gamma & 0 \\ 0 & 1 & 0 & 0 \\ \gamma & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad |\gamma| \leq 1.$$

By Lemma 6.10, take $\epsilon > 0$ small enough to get

$$\sum_{(d, d') \in \Omega_{2,II}} P(Z_d > a_n, Z_{d'} > a_n) = O\left(\frac{\tau}{p^{1-\epsilon}}\right) \rightarrow 0 \quad (73)$$

as $n \rightarrow \infty$.

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

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \gamma \\ 0 & 0 & 1 & 0 \\ 0 & \gamma & 0 & 1 \end{pmatrix}, \quad |\gamma| \leq 1.$$

Thus, the covariance matrix of $(w_{d_2}, w_{d_1}, w_{d_4}, w_{d_3})$ is equal to Σ_4 in Lemma 6.10. Then, by the same argument as that in the equality in (71) we get

$$\sum_{(d, d') \in \Omega_{2,III}} P(Z_d > a_n, Z_{d'} > a_n) = O\left(\frac{\tau}{p^{1-\epsilon}}\right) \rightarrow 0 \quad (74)$$

as $n \rightarrow \infty$ by taking $\epsilon > 0$ small enough. Combining (72), (73) and (74), we conclude

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

as $n \rightarrow \infty$. This and (71) together with (69) say that, to finish the proof of this proposition, it suffices to verify

$$\sum_{(d,d') \in \Omega_3} P(Z_d > a_n, Z_{d'} > a_n) \rightarrow 0 \quad (75)$$

as $n \rightarrow \infty$. The next lemma confirms this. The proof is then completed. \blacksquare

LEMMA 6.13 *Let the notation be as in the proof of Proposition 6.4, then (75) holds.*

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 $X_n = (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$ 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

$$b_{n,3} \rightarrow 1 \text{ in probability as } n \rightarrow \infty, \\ \left\{ \sqrt{\frac{n}{\log p}} b_{n,1} \right\} \text{ and } \left\{ \sqrt{\frac{n}{\log p}} b_{n,4} \right\} \text{ are tight} \quad (76)$$

as $n \rightarrow \infty, p \rightarrow \infty$ with $\log p = o(n)$, 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 (56). It is seen from Proposition 6.4 that

$$\frac{V_{n,\tau}}{\sqrt{n \log p}} \rightarrow 2 \quad (77)$$

in probability as $n \rightarrow \infty, p \rightarrow \infty$ 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

$$\Delta_n := \max_{1 \leq i < j \leq p, |i-j| \geq \tau} |n\rho_{ij} - v_{ij}| \leq (b_{n,1}^2 + 2b_{n,1})V_{n,\tau}b_{n,3}^{-2} + nb_{n,3}^{-2}b_{n,4}^2. \quad (78)$$

Now, using (76), (77) and (78), 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

$$\frac{n^2 L_{n,\tau}^2 - V_{n,\tau}^2}{n} \rightarrow 0$$

in probability as $n \rightarrow \infty$. This joint with Proposition 6.4 and the Slutsky lemma yields the desired limiting result for $L_{n,\tau}$. \blacksquare

References

- [1] Achlioptas, D. (2001). Database-friendly random projections. In *Proc. ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems*, 274-281.
- [2] Anderson, G. W., Guionnet, A. and Zeitouni, O. (2009). *An Introduction to Random Matrices*. Cambridge University Press.
- [3] Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817-858.
- [4] Arratia, R. and Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximation: The Chen-Stein method. *Ann. Probab.* **17**, 9-25.
- [5] Bai, Z. D., Jiang, D., Yao, J.-F. and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37**, 3822-3840.
- [6] Bai, Z., Miao, B. Q. and Pan, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35**, 1532-1572.
- [7] Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension comparison of significance tests for a high-dimensional two sample problem. *Statist. Sinica* **6**, 311-329.
- [8] Bai, Z. and Silverstein, J. W. (2009). *Spectral Analysis of Large Dimensional Random Matrices*. Second Edition, Springer.
- [9] 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.
- [10] Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37**, 1705-1732.
- [11] Cai, T. and Lv, J. (2007). Discussion: The Dantzig selector: statistical estimation when p is much larger than n . *Ann. Statist.* **35**, 2365-2369.
- [12] Cai, T. T., Wang, L. and Xu, G. (2010a). Shifting inequality and recovery of sparse signals, *IEEE Transactions on Signal Processing* **58**, 1300-1308.
- [13] Cai, T. T., Wang, L. and Xu, G. (2010b). Stable recovery of sparse signals and an oracle inequality, *IEEE Trans. Inf. Theory*, to appear.
- [14] Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.*, to appear.
- [15] Candès, E. J. and Plan, Y. (2009). Near-ideal model selection by ℓ_1 minimization. *Ann. Statist.* **37**, 2145-2177.

- [16] Candès, E. J. and Tao, T. (2005). Decoding by linear programming, *IEEE Trans. Inf. Theory* 51, 4203-4215.
- [17] Candès, E. J. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n (with discussion). *Ann. Statist.* 35, 2313-2351.
- [18] Chen, X. (1990). Probabilities of moderate deviations for B-valued independent random vectors. *Chinese Ann. Mathematics* 11, 621-629.
- [19] Chen, X. (1991). Probabilities of moderate deviations for independent random vectors in a Banach space. *Chinese J. of Appl. Probab. and Statist.* 7, 24-32.
- [20] Chow, Y. S. and Teicher, H. (1997). *Probability Theory, Independence, Interchangeability, Martingales*. Springer Texts in Statistics, Third edition.
- [21] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Springer, Second edition.
- [22] Diaconis, P., Eaton, M. and Lauritzen, L. (1992). Finite deFinetti theorem in linear models and multivariate analysis. *Scand. J. Statist.* 19, 289-315.
- [23] Donoho, D. L. and Huo, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory* 47, 2845-2862.
- [24] Donoho, D. (2006a). Compressed sensing. *IEEE Trans. Inf. Theory* 52, 1289-1306.
- [25] Donoho, D. (2006b). For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.* 59, 797-829.
- [26] Donoho, D. L., Elad, M. and Temlyakov, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* 52, 6-18.
- [27] Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of the Royal Statistical Society Series B* 70, 849-911.
- [28] Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* 20, 101-148.
- [29] Fristedt, B. and Gray, L. (1996). *A Modern Approach to Probability Theory*. Birkhäuser Boston, 1 edition.
- [30] Fuchs, J.-J. (2004). On sparse representations in arbitrary redundant bases, *IEEE Trans. Inf. Theory* 50, 1341-1344.

- [31] Jiang, T. (2004a). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* 14, 865-880.
- [32] Jiang, T. (2004b). The limiting distribution of eigenvalues of sample correlation matrices. *Sankhya* 66, 35-48.
- [33] Jiang, T. (2005). Maxima of Entries of Haar Distributed Matrices. *Probability Theory and Related Fields* 131, 121-144.
- [34] Jiang, T. (2006). How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.* 34, 1497-1529.
- [35] Jiang, T. (2009). The entries of circular orthogonal ensembles. *Journal of Mathematical Physics* 50, 063302.
- [36] Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.* 29, 295-327.
- [37] Johnstone, I. (2008). Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Stat.* 36, 2638-2716.
- [38] Ledoux, M. (1992). On moderate deviations of sums of i.i.d. vector random variables. *Ann. Inst. H. Poincaré Probab. Statist.*, 28, 267-280.
- [39] 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 Relat. Fields.*
- [40] Li, D. and Rosalsky, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices. *Ann. Appl. Probab.* 16, 423-447.
- [41] Ligeralde, A. and Brown, B. (1995). Band covariance matrix estimation using restricted residuals: A Monte Carlo analysis. *International Economic Review* **36**, 751-767.
- [42] Liu, W. D., Lin, Z. Y. 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.
- [43] Péché, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Relat. Fields* **143**, 481C51.
- [44] Sakhanenko, A. I. (1991). Estimates of Berry-Esseen type for the probabilities of large deviations. *Sibirsk. Mat. Zh.* 32, 133-142, 228.
- [45] Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transaction of American Mathematical Society* 359, 5345-5363.

7 Appendix

In this appendix we prove Proposition 6.2 and verify the three examples given in Section 4. We then prove Lemmas 6.5 - 6.7 and Lemmas 6.9 - 6.13 which are used in the proof of the main results.

Proof of Proposition 6.2. We prove the proposition by following the outline of the proof of Proposition 6.1 step by step. It suffices to show

$$\lim_{n \rightarrow \infty} P\left(\frac{W_n}{\sqrt{n \log p}} \geq 2 + 2\epsilon\right) = 0 \quad \text{and} \quad (79)$$

$$\lim_{n \rightarrow \infty} P\left(\frac{W_n}{\sqrt{n \log p}} \leq 2 - \epsilon\right) = 0 \quad (80)$$

for any $\epsilon > 0$ small enough. Note that $|x_{11}x_{12}|^\rho = |x_{11}|^\rho \cdot |x_{12}|^\rho \leq |x_{11}|^{2\rho} + |x_{12}|^{2\rho}$ for any $\rho > 0$. The given moment condition implies that $E \exp(t_0|x_{11}|^{4\beta/(1-\beta)}) < \infty$. Hence $E \exp(|x_{11}|^{\frac{4\beta}{1+\beta}}) < \infty$ and $E \exp(|x_{11}x_{12}|^{\frac{2\beta}{1+\beta}}) < \infty$. By (i) of Lemma 6.4, (30) holds for $\{p_n\}$ such that $p_n \rightarrow \infty$ and $\log p_n = o(n^\beta)$. By using (29) and (31), we obtain (79).

By using condition $E \exp\{t_0|x_{11}|^{\frac{4\beta}{1+\beta}}\} < \infty$ again, we know (35) also holds for $\{p_n\}$ such that $p_n \rightarrow \infty$ and $\log p_n = o(n^\beta)$. Then all statements after (32) and before (38) hold. Now, by Lemma 6.7, (39) holds for $\{p_n\}$ such that $p_n \rightarrow \infty$ and $\log p_n = o(n^\beta)$, we then have (40). This implies (32), which is the same as (80). \blacksquare

Verifications of (23), (24) and (25). We consider the three one by one.

(i) If $x_{11} \sim N(0, n^{-1})$ as in (19), then ξ and η are i.i.d. with distribution $N(0, 1)$. By Lemma 3.2 from Jiang (2005), $I_2(x) = (x - 1 - \log x)/2$ for $x > 0$. So $I_2(1/2) > 1/12$. Also, since $Ee^{\theta\xi\eta} = Ee^{\theta^2\xi^2/2} = (1 - \theta^2)^{-1/2}$ for $|\theta| < 1$. It is straightforward to get

$$I_1(x) = \frac{\sqrt{4x^2 + 1} - 1}{2} - \frac{1}{2} \log \frac{\sqrt{4x^2 + 1} + 1}{2}, \quad x > 0.$$

Let $y = \frac{\sqrt{4x^2 + 1} - 1}{2}$. Then $y > 2x^2/3$ for all $|x| \leq 4/5$. Thus, $I_1(x) = y - \frac{1}{2} \log(1+y) > \frac{y}{2} > \frac{x^2}{3}$ for $|x| \leq 4/5$. Therefore, $g(t) \geq \min\{I_1(\frac{t}{2}), \frac{1}{12}\} \geq \min\{\frac{t^2}{12}, \frac{1}{12}\} = \frac{t^2}{12}$ for $|t| \leq 1$. Since $1/(2k - 1) \leq 1$ if $k \geq 1$. By Proposition 4.1, we have

$$P\left((2k - 1)\tilde{L}_n < 1\right) \geq 1 - 3p^2 \exp\left\{-\frac{n}{12(2k - 1)^2}\right\} \quad (81)$$

for all $n \geq 2$ and $k \geq 1$, which is (23).

(ii) Let x_{11} be such that $P(x_{11} = \pm 1/\sqrt{n}) = 1/2$ as in (20). Then ξ and η in Proposition 4.1 are i.i.d. with $P(\xi = \pm 1) = 1/2$. Hence, $P(\xi\eta = \pm 1) = 1/2$ and $\xi^2 = 1$. Immediately, $I_2(1) = 0$ and $I_2(x) = +\infty$ for all $x \neq 1$. If $\alpha = \log \frac{3}{2} \sim 0.405$, then $E(Z^2 e^{\alpha|Z|}) = e^\alpha \leq \frac{3}{2}$

with $Z = \xi\eta$. Thus, by Lemma 6.8, $I_1(x) \geq x^2/3$ for all $0 \leq x \leq \frac{3}{5} \leq \frac{3\alpha}{2}$. Therefore, $g(t) \geq \frac{t^2}{12}$ for $0 \leq t \leq \frac{6}{5}$. This gives that

$$P\left((2k-1)\tilde{L}_n < 1\right) \geq 1 - 3p^2 \exp\left\{-\frac{n}{12(2k-1)^2}\right\} \quad (82)$$

provided $\frac{1}{2k-1} \leq \frac{6}{5}$, that is, $k \geq \frac{11}{12}$. We then obtain (24) since k is an integer.

(iii) 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 ξ and η in Proposition 4.1 are i.i.d. with $P(\xi = \pm\sqrt{3}) = 1/6$ and $P(\xi = 0) = 2/3$. It follows that $P(Z = \pm 3) = 1/18$ and $P(Z = 0) = 8/9$ with $Z = \xi\eta$. Take $\alpha = \frac{1}{3} \log \frac{3}{2} > 0.13$. Then $E(Z^2 e^{\alpha|Z|}) = \frac{2 \times 9}{18} e^{3\alpha} = \frac{3}{2}$. Thus, by Lemma 6.8, $I_1(x) \geq x^2/3$ for all $0 \leq x \leq \frac{3\alpha}{2} = \frac{1}{2} \log \frac{3}{2} \sim 0.2027$. Now, $P(\xi^2 = 3) = \frac{1}{3} = 1 - P(\xi^2 = 0)$. Hence, $\xi^2/3 \sim \text{Ber}(p)$ with $p = \frac{1}{3}$. It follows that

$$\begin{aligned} I_2(x) &= \sup_{\theta \in \mathbb{R}} \left\{ (3\theta) \frac{x}{3} - \log E e^{3\theta(\xi^2/3)} \right\} \\ &= \Lambda^*\left(\frac{x}{3}\right) = \frac{x}{3} \log x + \left(1 - \frac{x}{3}\right) \log \frac{3-x}{2} \end{aligned}$$

for $0 \leq x \leq 3$ by (b) of Exercise 2.2.23 from [21]. Thus, $I_2(\frac{1}{2}) = \frac{1}{6} \log \frac{1}{2} + \frac{5}{6} \log \frac{5}{4} \sim 0.0704 > \frac{1}{15}$. Now, for $0 \leq t \leq \frac{2}{5}$, we have

$$g(t) = \min \left\{ I_1\left(\frac{t}{2}\right), I_2\left(\frac{1}{2}\right) \right\} \geq \min \left\{ \frac{t^2}{12}, \frac{1}{15} \right\} = \frac{t^2}{12}.$$

Easily, $t := \frac{1}{2k-1} \leq \frac{2}{5}$ if and only if $k \geq \frac{7}{4}$. Thus, by Proposition 4.1,

$$P\left((2k-1)\tilde{L}_n < 1\right) \geq 1 - 3p^2 \exp\left\{-\frac{n}{12(2k-1)^2}\right\} \quad (83)$$

for all $n \geq 2$ and $k \geq \frac{7}{4}$. We finally conclude (25) since k is an integer. \blacksquare

Proof of Lemma 6.5. (i) First, since x_{ij} 's are i.i.d. bounded random variables with mean zero and variance one, by (i) of Lemma 6.4,

$$P(\sqrt{n/\log p} b_{n,4} \geq K) = P\left(\max_{1 \leq i \leq p} \left| \frac{1}{\sqrt{n \log p}} \sum_{k=1}^n x_{ki} \right| \geq K\right) \quad (84)$$

$$\begin{aligned} &\leq p \cdot P\left(\left| \frac{1}{\sqrt{n \log p}} \sum_{k=1}^n x_{k1} \right| \geq K\right) \\ &\leq p \cdot e^{-(K^2/3) \log p} = \frac{1}{p^{K^2/3-1}} \rightarrow 0 \end{aligned} \quad (85)$$

as $n \rightarrow \infty$ for any $K > \sqrt{3}$. This says that $\{\sqrt{n/\log p} b_{n,4}\}$ are tight.

Second, noticing that $|t-1| \leq |t^2-1|$ for any $t > 0$ and $nh_i^2 = \|x_i - \bar{x}_i\|^2 = x_i^T x_i - n|\bar{x}_i|^2$, we get that

$$\begin{aligned} b_{n,1} &\leq \max_{1 \leq i \leq p} |h_i^2 - 1| \leq \max_{1 \leq i \leq p} \left| \frac{1}{n} \sum_{k=1}^n (x_{ki}^2 - 1) \right| + \max_{1 \leq i \leq p} \left| \frac{1}{n} \sum_{k=1}^n x_{ki} \right|^2 \\ &= Z_n + b_{n,4}^2 \end{aligned} \quad (86)$$

where $Z_n = \max_{1 \leq i \leq p} \left| \frac{1}{n} \sum_{k=1}^n (x_{ki}^2 - 1) \right|$. Therefore,

$$\sqrt{\frac{n}{\log p}} b_{n,1} \leq \sqrt{\frac{n}{\log p}} Z_n + \sqrt{\frac{\log p}{n}} \cdot \left(\sqrt{\frac{n}{\log p}} b_{n,4} \right)^2. \quad (87)$$

Replacing “ x_{ki} ” in (84) with “ $x_{ki}^2 - 1$ ” and using the same argument, we obtain that $\{\sqrt{n/\log p} Z_n\}$ are tight. Since $\log p = o(n)$ and $\{\sqrt{n/\log p} b_{n,4}\}$ are tight, using (27) we know the second term on the right hand side of (87) goes to zero in probability as $n \rightarrow \infty$. Hence, we conclude from (87) that $\{\sqrt{n/\log p} b_{n,1}\}$ are tight.

Finally, since $\log p = o(n)$ and $\{\sqrt{n/\log p} b_{n,1}\}$ are tight, use (27) to have $b_{n,1} \rightarrow 0$ in probability as $n \rightarrow \infty$. This implies that $b_{n,3} \rightarrow 1$ in probability as $n \rightarrow \infty$.

(ii) By (85) and (87), to prove the conclusion, it is enough to show, for some constant $K > 0$,

$$p \cdot P\left(\left|\frac{1}{\sqrt{n \log p}} \sum_{k=1}^n x_{k1}\right| \geq K\right) \rightarrow 0 \quad \text{and} \quad (88)$$

$$p \cdot P\left(\left|\frac{1}{\sqrt{n \log p}} \sum_{k=1}^n (x_{k1}^2 - 1)\right| \geq K\right) \rightarrow 0 \quad (89)$$

as $n \rightarrow \infty$. Using $a_n := \sqrt{\log p_n} = o(n^{\beta/2})$ and (i) of Lemma 6.4, we have

$$\begin{aligned} P\left(\left|\frac{1}{\sqrt{n \log p}} \sum_{k=1}^n x_{k1}\right| \geq K\right) &\leq \frac{1}{p^{K^2/3}} \quad \text{and} \\ P\left(\left|\frac{1}{\sqrt{n \log p}} \sum_{k=1}^n (x_{k1}^2 - 1)\right| \geq K\right) &\leq \frac{1}{p^{K^2/3}} \end{aligned}$$

as n is sufficiently large, where the first inequality holds provided $E \exp(t_0 |x_{11}|^{2\beta/(1+\beta)}) = E \exp(t_0 |x_{11}|^{\alpha/2}) < \infty$; the second holds since $E \exp(t_0 |x_{11}^2 - 1|^{2\beta/(1+\beta)}) = E \exp(t_0 |x_{11}^2 - 1|^{\alpha/2}) < \infty$ for some $t_0 > 0$, which is equivalent to $E e^{t'_0 |x_{11}|^\alpha} < \infty$ for some $t'_0 > 0$. We then get (88) and (89) by taking $K = 2$. \blacksquare

Proof of Lemma 6.6. Let $G_n = \{|\sum_{k=1}^n x_{k1}^2/n - 1| < \delta\}$. Then, by the Chernoff bound (see, e.g., p. 27 from Dembo and Zeitouni (1998)), for any $\delta \in (0, 1)$, there exists a constant $C_\delta > 0$ such that $P(G_n^c) \leq 2e^{-nC_\delta}$ for all $n \geq 1$. Set $a_n = t_n \sqrt{n \log p}$. Then

$$\Psi_n \leq E\left\{P^1\left(\left|\sum_{k=1}^n x_{k1} x_{k2}\right| > a_n\right)^2 I_{G_n}\right\} + 2e^{-nC_\delta} \quad (90)$$

for all $n \geq 1$. Evidently, $|x_{k1}x_{k2}| \leq C^2$, $E^1(x_{k1}x_{k2}) = 0$ and $E^1(x_{k1}x_{k2})^2 = x_{k1}^2$, where E^1 stands for the conditional expectation given $\{x_{k1}, 1 \leq k \leq n\}$. By the Bernstein inequality (see, e.g., p.111 from Chow and Teicher (1997)),

$$\begin{aligned} P^1\left(\left|\sum_{k=1}^n x_{k1}x_{k2}\right| > a_n\right)^2 I_{G_n} &\leq 4 \cdot \exp\left\{-\frac{a_n^2}{(\sum_{k=1}^n x_{k1}^2 + C^2 a_n)}\right\} I_{G_n} \\ &\leq 4 \cdot \exp\left\{-\frac{a_n^2}{((1+\delta)n + C^2 a_n)}\right\} \\ &\leq \frac{1}{p^{t^2/(1+2\delta)}} \end{aligned} \quad (91)$$

as n is sufficiently large, since $a_n^2/(n(1+\delta) + C^2 a_n) \sim t^2(\log p)/(1+\delta)$ as $n \rightarrow \infty$. Recalling (90), the conclusion then follows by taking δ small enough. \blacksquare

Proof of Lemma 6.7. Let P^2 stand for the conditional probability given $\{x_{k2}, 1 \leq k \leq n\}$. Since $\{x_{ij}; i \geq 1, j \geq 1\}$ are i.i.d., to prove the lemma, it is enough to prove

$$\Psi_n := E\left\{P^2\left(\left|\sum_{k=1}^n x_{k1}x_{k2}\right| > t_n \sqrt{n \log p}\right)^2\right\} = O\left(\frac{1}{p^{t^2-\epsilon}}\right) \quad (92)$$

as $n \rightarrow \infty$. We do this only for convenience of notation.

Step 1. For any $x > 0$, by the Markov inequality

$$P\left(\max_{1 \leq k \leq n} |x_{k2}| \geq x\right) \leq nP(|x_{12}| \geq x) \leq Cne^{-t_0 x^\alpha} \quad (93)$$

where $C = Ee^{t_0|x_{11}|^\alpha} < \infty$. Second, the given condition implies that $Ee^{t|x_{11}|^{4\beta/(1+\beta)}} < \infty$ for any $t > 0$. For any $\epsilon > 0$, by (ii) of Lemma 6.4, there exists a constant $C = C_\epsilon > 0$ such that

$$P\left(\left|\frac{\sum_{k=1}^n x_{k2}^2 - n}{n^{(\beta+1)/2}}\right| \geq \epsilon\right) \leq e^{-C_\epsilon n^\beta} \quad (94)$$

for each $n \geq 1$.

Set $h_n = n^{(1-\beta)/4}$, $\mu_n = Ex_{ij}I(|x_{ij}| \leq h_n)$,

$$\begin{aligned} y_{ij} &= x_{ij}I(|x_{ij}| \leq h_n) - Ex_{ij}I(|x_{ij}| \leq h_n) \\ z_{ij} &= x_{ij}I(|x_{ij}| > h_n) - Ex_{ij}I(|x_{ij}| > h_n) \end{aligned} \quad (95)$$

for all $i \geq 1$ and $j \geq 1$. Then, $x_{ij} = y_{ij} + z_{ij}$ for all $i, j \geq 1$. Use the inequality $P(U + V \geq u + v) \leq P(U \geq u) + P(V \geq v)$ to obtain

$$\begin{aligned} &P^2\left(\left|\sum_{k=1}^n x_{k1}x_{k2}\right| > t_n \sqrt{n \log p}\right)^2 \\ &\leq 2P^2\left(\left|\sum_{k=1}^n y_{k1}x_{k2}\right| > (t_n - \delta)\sqrt{n \log p}\right)^2 + 2P^2\left(\left|\sum_{k=1}^n z_{k1}x_{k2}\right| > \delta\sqrt{n \log p}\right)^2 \\ &:= 2A_n + 2B_n \end{aligned} \quad (96)$$

for any $\delta > 0$ small enough. Hence,

$$\Psi_n \leq 2EA_n + 2EB_n \quad (97)$$

for all $n \geq 2$.

Step 2: the bound of A_n . Now, if $\max_{1 \leq k \leq n} |x_{k2}| \leq h_n$, then $|y_{k1}x_{k2}| \leq 2h_n^2$ for all $k \geq 1$. It then follows from the Bernstein inequality (see, e.g., p. 111 from Chow and Teicher (1997)) that

$$\begin{aligned} A_n &= P^2\left(\left|\sum_{k=1}^n y_{k1}x_{k2}\right| > (t_n - \delta)\sqrt{n \log p}\right)^2 \\ &\leq 4 \cdot \exp\left\{-\frac{(t_n - \delta)^2 n \log p}{E(y_{11}^2) \sum_{k=1}^n x_{k2}^2 + 2h_n^2(t_n - \delta)\sqrt{n \log p}}\right\} \\ &\leq 4 \cdot \exp\left\{-\frac{(t_n - \delta)^2 n \log p}{E(y_{11}^2)(n + \epsilon n^{(\beta+1)/2}) + 2h_n^2(t_n - \delta)\sqrt{n \log p}}\right\} \end{aligned}$$

for $0 < \delta < t_n$ and $\frac{|\sum_{k=1}^n x_{k2}^2 - n|}{n^{(\beta+1)/2}} < \epsilon$. Notice $E(y_{11}^2) \rightarrow 1$ and $2h_n^2(t_n - \delta)\sqrt{n \log p}/3 = o(n)$ as $n \rightarrow \infty$. Thus,

$$\frac{(t_n - \delta)^2 n \log p}{E(y_{11}^2)(n + \epsilon n^{(\beta+1)/2}) + 2h_n^2(t_n - \delta)\sqrt{n \log p}} \sim (t - \delta)^2 \log p$$

as $n \rightarrow \infty$. In summary, if $\max_{1 \leq k \leq n} |x_{k2}| \leq h_n$ and $\frac{|\sum_{k=1}^n x_{k2}^2 - n|}{n^{(\beta+1)/2}} \leq \epsilon$, then for any $\delta \in (0, t/2)$,

$$A_n \leq \frac{1}{p^{t^2 - 2t\delta}} \quad (98)$$

as n is sufficiently large. Therefore, for any $\epsilon > 0$ small enough, take δ sufficiently small to obtain

$$\begin{aligned} EA_n &= E\left\{P^2\left(\left|\sum_{k=1}^n y_{k1}x_{k2}\right| > (t_n - \delta)\sqrt{n \log p}\right)^2\right\} \\ &\leq \frac{1}{p^{t^2 - \epsilon}} + P(\max_{1 \leq k \leq n} |x_{k2}| \geq h_n) + P\left(\frac{|\sum_{k=1}^n x_{k2}^2 - n|}{n^{(\beta+1)/2}} \geq \epsilon\right) \\ &\leq \frac{1}{p^{t^2 - \epsilon}} + Cne^{-h_n^\alpha} + e^{-C_\epsilon n^\beta} = O\left(\frac{1}{p^{t^2 - \epsilon}}\right) \end{aligned} \quad (99)$$

as $n \rightarrow \infty$, where the second inequality follows from (93) and (94), and the last identity follows from the fact that $h_n^\alpha = n^\beta$ and the assumption $\log p = o(n^\beta)$.

Step 3: the bound of B_n . Recalling the definition of z_{ij} and μ_n in (95), we have

$$\begin{aligned} \sqrt{B_n} &= P^2\left(\left|\sum_{k=1}^n z_{k1}x_{k2}\right| > \delta\sqrt{n \log p}\right) \\ &\leq P^2\left(\left|\sum_{k=1}^n x_{k1}x_{k2}I\{|x_{k1}| > h_n\}\right| > \delta\sqrt{n \log p}/2\right) + I\left(\left|\sum_{k=1}^n x_{k2}\right| > \frac{\delta\sqrt{n \log p}}{2(e^{-n} + |\mu_n|)}\right) \\ &:= C_n + D_n. \end{aligned} \quad (100)$$

Now, by (93),

$$C_n \leq P(\max_{1 \leq k \leq n} |x_{k1}| > h_n) \leq Cn e^{-t_0 h_n^\alpha} = Cn e^{-t_0 n^\beta}. \quad (101)$$

Easily, $|\mu_n| \leq E|x_{11}|I(|x_{11}| > h_n) \leq e^{-t_0 h_n^\alpha/2} E(|x_{11}|e^{t_0 |x_{11}|^\alpha/2}) = C e^{-t_0 n^\beta/2}$. Also, $P(|\sum_{k=1}^n \eta_k| \geq x) \leq \sum_{k=1}^n P(|\eta_k| \geq x/n)$ for any random variables $\{\eta_i\}$ and $x > 0$. We then have

$$\begin{aligned} ED_n &= P\left(\left|\sum_{k=1}^n x_{k2}\right| > \frac{\delta\sqrt{n \log p}}{2(e^{-n} + |\mu_n|)}\right) \\ &\leq nP\left(|x_{11}| > \frac{\delta\sqrt{n \log p}}{2n(e^{-n} + |\mu_n|)}\right) \\ &\leq nP\left(|x_{11}| > e^{t_0 n^\beta/3}\right) \leq e^{-n} \end{aligned} \quad (102)$$

as n is sufficiently large, where the last inequality is from condition $Ee^{t_0 |x_{11}|^\alpha} < \infty$. Consequently,

$$EB_n \leq 2E(C_n^2) + 2E(D_n^2) = 2E(C_n^2) + 2E(D_n) \leq e^{-Cn^\beta} \quad (103)$$

as n is sufficiently large. This joint with (97) and (99) yields (92). \blacksquare

Proof of Lemma 6.9. Take $\gamma = (1 - \beta)/2 \in [1/3, 1/2)$. Set

$$\eta_i = \xi_i I(|\xi_i| \leq n^\gamma), \quad \mu_n = E\eta_1 \quad \text{and} \quad \sigma_n^2 = \text{Var}(\eta_1), \quad 1 \leq i \leq n. \quad (104)$$

Since the desired result is a conclusion about $n \rightarrow \infty$, without loss of generality, assume $\sigma_n > 0$ for all $n \geq 1$. We first claim that there exists a constant $C > 0$ such that

$$\max\left\{|\mu_n|, |\sigma_n - 1|, P(|\xi_1| > n^\gamma)\right\} \leq C e^{-n^\beta/C} \quad (105)$$

for all $n \geq 1$. In fact, since $E\xi_1 = 0$ and $\alpha\gamma = \beta$,

$$|\mu_n| = |E\xi_1 I(|\xi_1| > n^\gamma)| \leq E|\xi_1| I(|\xi_1| > n^\gamma) \leq E\left(|\xi_1| e^{t_0 |\xi_1|^\alpha/2}\right) \cdot e^{-t_0 n^\beta/2} \quad (106)$$

for all $n \geq 1$. Note that $|\sigma_n - 1| \leq |\sigma_n^2 - 1| = \mu_n^2 + E\xi_1^2 I(|\xi_1| > n^\gamma)$, by the same argument as in (106), we know both $|\sigma_n - 1|$ and $P(|\xi_1| > n^\gamma)$ are bounded by $C e^{-n^\beta/C}$ for some $C > 0$. Then (105) follows.

Step 1. We prove that, for some constant $C > 0$,

$$\left|P\left(\frac{S_n}{\sqrt{n \log p_n}} \geq y_n\right) - P\left(\frac{\sum_{i=1}^n \eta_i}{\sqrt{n \log p_n}} \geq y_n\right)\right| \leq 2e^{-n^\beta/C} \quad (107)$$

for all $n \geq 1$. Observe

$$\xi_i \equiv \eta_i \quad \text{for } 1 \leq i \leq n \quad \text{if} \quad \max_{1 \leq i \leq n} |\xi_i| \leq n^\gamma. \quad (108)$$

Then, by (105),

$$\begin{aligned} P\left(\frac{S_n}{\sqrt{n \log p_n}} \geq y_n\right) &\leq P\left(\frac{S_n}{\sqrt{n \log p_n}} \geq y_n, \max_{1 \leq i \leq n} |\xi_i| \leq n^\gamma\right) + P\left(\bigcup_{i=1}^n \{|\xi_i| > n^\gamma\}\right) \\ &\leq P\left(\frac{\sum_{i=1}^n \eta_i}{\sqrt{n \log p_n}} \geq y_n\right) + Cne^{-n^\beta/C} \end{aligned} \quad (109)$$

for all $n \geq 1$. Use inequality that $P(AB) \geq P(A) - P(B^c)$ for any events A and B to have

$$\begin{aligned} P\left(\frac{S_n}{\sqrt{n \log p_n}} \geq y_n\right) &\geq P\left(\frac{S_n}{\sqrt{n \log p_n}} \geq y_n, \max_{1 \leq i \leq n} |\xi_i| \leq n^\gamma\right) \\ &= P\left(\frac{\sum_{i=1}^n \eta_i}{\sqrt{n \log p_n}} \geq y_n, \max_{1 \leq i \leq n} |\xi_i| \leq n^\gamma\right) \\ &\geq P\left(\frac{\sum_{i=1}^n \eta_i}{\sqrt{n \log p_n}} \geq y_n\right) - Cne^{-n^\beta/C} \end{aligned}$$

where in the last step the inequality $P(\max_{1 \leq i \leq n} |\xi_i| > n^\gamma) \leq Cne^{-n^\beta/C}$ is used as in (109).

This and (109) concludes (107).

Step 2. Now we prove

$$P\left(\frac{\sum_{i=1}^n \eta_i}{\sqrt{n \log p_n}} \geq y_n\right) \sim \frac{e^{-x_n^2/2}}{\sqrt{2\pi x_n}} \quad (110)$$

as $n \rightarrow \infty$, where

$$x_n = y'_n \sqrt{\log p_n} \quad \text{and} \quad y'_n = \frac{1}{\sigma_n} \left(y_n - \sqrt{\frac{n}{\log p_n}} \mu_n \right). \quad (111)$$

First, by (105),

$$|y'_n - y_n| \leq \frac{|1 - \sigma_n|}{\sigma_n} y_n + \frac{1}{\sigma_n} \cdot \sqrt{\frac{n}{\log p_n}} |\mu_n| \leq Ce^{-n^\beta/C} \quad (112)$$

for all $n \geq 1$ since both σ_n and y_n have limits and $p_n \rightarrow \infty$. In particular, since $\log p_n = o(n^\beta)$,

$$x_n = o(n^{\beta/2}) \quad (113)$$

as $n \rightarrow \infty$. Now, set

$$\eta'_i = \frac{\eta_i - \mu_n}{\sigma_n}$$

for $1 \leq i \leq n$. Easily

$$P\left(\frac{\sum_{i=1}^n \eta_i}{\sqrt{n \log p_n}} \geq y_n\right) = P\left(\frac{\sum_{i=1}^n \eta'_i}{\sqrt{n \log p_n}} \geq y'_n\right) \quad (114)$$

for all $n \geq 1$. Reviewing (104), for some constant $K > 0$, we have $|\eta'_i| \leq Kn^\gamma$ for $1 \leq i \leq n$. Take $c_n = Kn^{\gamma-1/2}$. Recalling x_n in (111). It is easy to check that

$$s_n := \left(\sum_{i=1}^n E\eta_i'^2 \right)^{1/2} = \sqrt{n}, \quad \varrho_n := \sum_{i=1}^n E|\eta_i'|^3 \sim nC, \quad |\eta_i'| \leq c_n s_n \text{ and } 0 < c_n \leq 1$$

as n is sufficiently large. Recall $\gamma = (1 - \beta)/2$, it is easy to see from (113) that

$$0 < x_n < \frac{1}{18c_n}$$

for n large enough. Now, let $\gamma(x)$ be as in Lemma 6.3, since $\beta \leq 1/3$, by the lemma and (113),

$$\left| \gamma\left(\frac{x_n}{s_n}\right) \right| \leq \frac{2x_n^3 \varrho_n}{s_n^3} = o\left(n^{\frac{3\beta}{2} - \frac{1}{2}}\right) \rightarrow 0 \quad \text{and} \quad \frac{(1+x_n)\varrho_n}{s_n^3} = O(n^{(\beta-1)/2}) \rightarrow 0$$

as $n \rightarrow \infty$. By (111) and (112), $x_n s_n = y'_n \sqrt{n \log p_n}$ and $x_n \rightarrow \infty$ as $n \rightarrow \infty$. Use Lemma 6.3 and the fact $1 - \Phi(t) = \frac{1}{\sqrt{2\pi t}} e^{-t^2/2}$ as $t \rightarrow +\infty$ to obtain

$$P\left(\frac{\sum_{i=1}^n \eta_i'}{\sqrt{n \log p_n}} \geq y'_n\right) = P\left(\sum_{i=1}^n \eta_i' \geq x_n s_n\right) \sim 1 - \Phi(x_n) \sim \frac{e^{-x_n^2/2}}{\sqrt{2\pi x_n}} \quad (115)$$

as $n \rightarrow \infty$. This and (114) conclude (110).

Step 3. Now we show

$$\frac{e^{-x_n^2/2}}{\sqrt{2\pi x_n}} \sim \frac{p_n^{-y_n^2/2} (\log p_n)^{-1/2}}{\sqrt{2\pi y}} := \omega_n \quad (116)$$

as $n \rightarrow \infty$. Since $y_n \rightarrow y$ and $\sigma_n \rightarrow 1$, we know from (112) that

$$\sqrt{2\pi} x_n = \sqrt{2\pi} y'_n (\log p_n)^{1/2} \sim \sqrt{2\pi} y (\log p_n)^{1/2} \quad (117)$$

as $n \rightarrow \infty$. Further, by (111),

$$\frac{e^{-x_n^2/2}}{p_n^{-y_n^2/2}} = \exp\left\{-\frac{x_n^2}{2} + \frac{y_n^2}{2} \log p_n\right\} = \exp\left\{\frac{1}{2}(y_n^2 - y_n'^2) \log p_n\right\}. \quad (118)$$

Since $y_n \rightarrow y$, by (112), both $\{y_n\}$ and $\{y_n'\}$ are bounded. It follows from (112) again that $|y_n^2 - y_n'^2| \leq C|y_n - y_n'| = O(e^{-n^{\beta/C}})$ as $n \rightarrow \infty$. With assumption $\log p_n = o(n^\beta)$ we get $e^{-x_n^2/2} \sim p_n^{-y_n^2/2}$ as $n \rightarrow \infty$, which combining with (117) yields (116).

Finally, we compare the right hand sides of (107) and (116). Choose $C' > \max\{y_n^2; n \geq 1\}$, since $\log p_n = o(n^\beta)$, recall ω_n in (116),

$$\begin{aligned} \frac{2e^{-n^{\beta/C}}}{\omega_n} &= 2\sqrt{2\pi} y (\log p_n)^{1/2} p_n^{y_n^2/2} e^{-n^{\beta/C}} \\ &= O\left(n^{\beta/2} \cdot \exp\left\{C' \log p_n - \frac{n^\beta}{C}\right\}\right) \\ &= O\left(n^{\beta/2} \cdot \exp\left\{-\frac{n^\beta}{2C}\right\}\right) \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ for any constant $C > 0$. This fact joint with (107), (110) and (116) proves the lemma. \blacksquare

Proof of Lemma 6.10. For any Borel set $A \subset \mathbb{R}$, set $P_2(A) = P(A|u_{k1}, u_{k3}, 1 \leq k \leq n)$, the conditional probability of A with respect to $u_{k1}, u_{k3}, 1 \leq k \leq n$. Observe from the expression of Σ_4 that three sets of random variables $\{u_{k1}, u_{k3}; 1 \leq k \leq n\}$, $\{u_{k2}; 1 \leq k \leq n\}$ and $\{u_{k4}; 1 \leq k \leq n\}$ are independent. Then

$$\begin{aligned} & P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n, \left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) \\ &= E\left\{P_2\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n\right)P_2\left(\left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right)\right\} \\ &\leq \left\{E P_2\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n\right)^2\right\}^{1/2} \cdot \left\{E P_2\left(\left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right)^2\right\}^{1/2} \end{aligned}$$

by the Cauchy-Schwartz inequality. Use the same independence again

$$P_2\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n\right) = P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n \mid u_{k1}, 1 \leq k \leq n\right); \quad (119)$$

$$P_2\left(\left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) = P\left(\left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n \mid u_{k3}, 1 \leq k \leq n\right). \quad (120)$$

These can be also seen from Proposition 27 in Fristedt and Gray (1997). It follows that

$$\begin{aligned} & \sup_{|r| \leq 1} P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n, \left|\sum_{k=1}^n u_{k3}u_{k4}\right| > a_n\right) \\ & \leq E\left\{P\left(\left|\sum_{k=1}^n u_{k1}u_{k2}\right| > a_n \mid u_{11}, \dots, u_{n1}\right)^2\right\}. \end{aligned}$$

Since $\{u_{k1}; 1 \leq k \leq n\}$ and $\{u_{k2}; 1 \leq k \leq n\}$ are independent, and $t_n := a_n/\sqrt{n \log p} \rightarrow t = 2$, taking $\alpha = 2$ in Lemma 6.7, we obtain the desired conclusion from the lemma. \blacksquare

Proof of Lemma 6.11. Since Σ_4 is always non-negative definite, the determinant of the first 3×3 minor of Σ_4 is non-negative: $1 - r_1^2 - r_2^2 \geq 0$. Let $r_3 = \sqrt{1 - r_1^2 - r_2^2}$ and $\{u_{k5}; 1 \leq k \leq n\}$ be i.i.d. standard normals which are independent of $\{u_{ki}; 1 \leq i \leq 4; 1 \leq k \leq n\}$. Then,

$$(u_{11}, u_{12}, u_{13}, u_{14}) \stackrel{d}{=} (u_{11}, u_{12}, r_1 u_{11} + r_2 u_{12} + r_3 u_{15}, u_{14}).$$

Define $Z_{ij} = |\sum_{k=1}^n u_{ki}u_{kj}|$ for $1 \leq i, j \leq 5$ and $r_5 = r_3$. By the Cauchy-Schwartz inequality,

$$\begin{aligned} \left| \sum_{k=1}^n (r_1 u_{k1} + r_2 u_{k2} + r_3 u_{k5}) u_{k4} \right| &\leq \sum_{i \in \{1,2,5\}} |r_i| \cdot \left| \sum_{k=1}^n u_{ki} u_{k4} \right| \\ &\leq \left(r_1^2 + r_2^2 + r_3^2 \right)^{1/2} \left(Z_{14}^2 + Z_{24}^2 + Z_{54}^2 \right)^{1/2} \\ &\leq \sqrt{3} \cdot \max\{Z_{14}, Z_{24}, Z_{54}\}. \end{aligned}$$

It follows from the above two facts that

$$\begin{aligned} &P\left(\left| \sum_{k=1}^n u_{k1} u_{k2} \right| > a_n, \left| \sum_{k=1}^n u_{k3} u_{k4} \right| > a_n \right) \\ &\leq P\left(Z_{12} > a_n, \max\{Z_{14}, Z_{24}, Z_{54}\} > \frac{a_n}{\sqrt{3}} \right) \\ &\leq \sum_{i \in \{1,2,5\}} P\left(Z_{12} > a_n, Z_{i4} > \frac{a_n}{\sqrt{3}} \right) \\ &= 2P\left(Z_{12} > a_n, Z_{14} > \frac{a_n}{\sqrt{3}} \right) + P\left(Z_{12} > a_n \right) \cdot P\left(Z_{54} > \frac{a_n}{\sqrt{3}} \right) \end{aligned} \quad (121)$$

by symmetry and independence. For any Borel set $A \subset \mathbb{R}$, set $P^1(A) = P(A|u_{k1}, 1 \leq k \leq n)$, the conditional probability of A with respect to $u_{k1}, 1 \leq k \leq n$. For any $s > 0$, from the fact that $\{u_{k1}\}, \{u_{k2}\}$ and $\{u_{k4}\}$ are independent, we see that

$$\begin{aligned} P\left(Z_{12} > a_n, Z_{14} > sa_n \right) &= E\left(P^1(Z_{12} > a_n) \cdot P^1(Z_{14} > sa_n) \right) \\ &\leq \left\{ E P^1(Z_{12} > a_n)^2 \right\}^{1/2} \cdot \left\{ E P^1(Z_{14} > sa_n)^2 \right\}^{1/2} \end{aligned}$$

by the Cauchy-Schwartz inequality. Taking $t_n := a_n/\sqrt{n \log p} \rightarrow t = 2$ and $t_n := sa_n/\sqrt{n \log p} \rightarrow t = 2s$ in Lemma 6.7, respectively, we get

$$E P^1(Z_{12} > a_n)^2 = O\left(p^{-4+\epsilon}\right) \quad \text{and} \quad E P^1(Z_{14} > sa_n)^2 = O\left(p^{-4s^2+\epsilon}\right)$$

as $n \rightarrow \infty$ for any $\epsilon > 0$. This implies that, for any $s > 0$ and $\epsilon > 0$,

$$P\left(Z_{12} > a_n, Z_{14} > sa_n \right) \leq O\left(p^{-2-2s^2+\epsilon}\right) \quad (122)$$

as $n \rightarrow \infty$. In particular,

$$P\left(Z_{12} > a_n, Z_{14} > \frac{a_n}{\sqrt{3}} \right) \leq O\left(p^{-\frac{8}{3}+\epsilon}\right) \quad (123)$$

as $n \rightarrow \infty$ for any $\epsilon > 0$.

Now we bound the last term in (121). Note that $|u_{11}u_{12}| \leq (u_{11}^2 + u_{12}^2)/2$, it follows that $E e^{|u_{11}u_{12}|/2} < \infty$ by independence and $E \exp(N(0, 1)^2/4) < \infty$. Since $\{u_{k1}, u_{k2}; 1 \leq k \leq n\}$

are i.i.d. with mean zero and variance one, and $y_n := a_n/\sqrt{n \log p} \rightarrow 2$ as $n \rightarrow \infty$, taking $\alpha = 1$ in Lemma 6.9, we get

$$\begin{aligned} P\left(Z_{12} > a_n\right) &= P\left(\frac{1}{\sqrt{n \log p}} \left| \sum_{k=1}^n u_{k1} u_{k2} \right| > \frac{a_n}{\sqrt{n \log p}}\right) \\ &\sim 2 \cdot \frac{p^{-y_n^2/2} (\log p)^{-1/2}}{2\sqrt{2\pi}} \sim \frac{e^{-y/2}}{\sqrt{2\pi}} \cdot \frac{1}{p^2} \end{aligned} \quad (124)$$

as $n \rightarrow \infty$. Similarly, for any $t > 0$,

$$P\left(Z_{12} > ta_n\right) = O\left(p^{-2t^2+\epsilon}\right) \quad (125)$$

as $n \rightarrow \infty$ (this can also be derived from (i) of Lemma 6.4). In particular,

$$P\left(Z_{54} > \frac{a_n}{\sqrt{3}}\right) = P\left(Z_{12} > \frac{a_n}{\sqrt{3}}\right) = O\left(p^{-\frac{2}{3}+\epsilon}\right) \quad (126)$$

as $n \rightarrow \infty$ for any $\epsilon > 0$. Combining (124) and (126), we know that the last term in (121) is bounded by $O(p^{-\frac{8}{3}+\epsilon})$ as $n \rightarrow \infty$ for any $\epsilon > 0$. This together with (121) and (123) concludes the lemma. \blacksquare

Proof of Lemma 6.12. Fix $\delta \in (0, 1)$. Take independent standard normals $\{u_{k5}, u_{k6}; 1 \leq k \leq n\}$ that are also independent of $\{u_{ki}; 1 \leq i \leq 4; 1 \leq k \leq n\}$. Then, since $\{u_{k1}, u_{k2}, u_{k5}, u_{k6}; 1 \leq k \leq n\}$ are i.i.d. standard normals, by checking covariance matrix Σ_4 , we know

$$(u_{11}, u_{12}, u_{13}, u_{14}) \stackrel{d}{=} (u_{11}, u_{12}, r_1 u_{11} + r'_1 u_{15}, r_2 u_{12} + r'_2 u_{16}) \quad (127)$$

where $r'_1 = \sqrt{1 - r_1^2}$ and $r'_2 = \sqrt{1 - r_2^2}$. Define $Z_{ij} = \left| \sum_{k=1}^n u_{ki} u_{kj} \right|$ for $1 \leq i, j \leq 6$. Then

$$\begin{aligned} &\left| \sum_{k=1}^n (r_1 u_{k1} + r'_1 u_{k5})(r_2 u_{k2} + r'_2 u_{k6}) \right| \\ &\leq |r_1 r_2| Z_{12} + |r_1 r'_2| Z_{16} + |r'_1 r_2| Z_{25} + |r'_1 r'_2| Z_{56} \\ &\leq (1 - \delta)^2 Z_{12} + 3 \max\{Z_{16}, Z_{25}, Z_{56}\} \end{aligned} \quad (128)$$

for all $|r_1|, |r_2| \leq 1 - \delta$. Let $\alpha = (1 + (1 - \delta)^2)/2$, $\beta = \alpha/(1 - \delta)^2$ and $\gamma = (1 - \alpha)/3$. Then

$$\beta > 1 \quad \text{and} \quad \gamma > 0. \quad (129)$$

Easily, if $Z_{12} \leq \beta a_n$, $\max\{Z_{16}, Z_{25}, Z_{56}\} \leq \gamma a_n$, then from (128) we know that the left hand side of (128) is controlled by a_n . Consequently, by (127) and the i.i.d. property,

$$\begin{aligned} P(Z_{12} > a_n, Z_{34} > a_n) &= P\left(Z_{12} > a_n, \left| \sum_{k=1}^n (r_1 u_{k1} + r'_1 u_{k5})(r_2 u_{k2} + r'_2 u_{k6}) \right| > a_n\right) \\ &\leq P(Z_{12} > a_n, Z_{12} > \beta a_n) + \sum_{i \in \{1, 2, 5\}} P(Z_{12} > a_n, Z_{i6} > \gamma a_n) \\ &= P(Z_{12} > \beta a_n) + 2P(Z_{12} > a_n, Z_{16} > \gamma a_n) \\ &\quad + P(Z_{12} > a_n) \cdot P(Z_{56} > \gamma a_n) \end{aligned} \quad (130)$$

where “ $2P(Z_{12} > a_n, Z_{16} > \gamma a_n)$ ” comes from the fact $(Z_{12}, Z_{16}) \stackrel{d}{=} (Z_{12}, Z_{26})$. Keep in mind that $(Z_{12}, Z_{16}) \stackrel{d}{=} (Z_{12}, Z_{14})$ and $Z_{56} \stackrel{d}{=} Z_{12}$. Recall (129), applying (122) and (125) to the three terms in the sum on the right hand side of (130), we conclude (55). \blacksquare

Proof of Lemma 6.13. Reviewing notation Ω_3 defined below (68), the current case is that $d_1 \leq d_3 \leq d_2 \leq d_4$ with $d = (d_1, d_2)$ and $d' = (d_3, d_4)$. Of course, by definition, $d_1 < d_2$ and $d_3 < d_4$. To save notation, define the “neighborhood” of d_i as follows:

$$N_i = \left\{ d \in \{1, \dots, p\}; |d - d_i| < \tau \right\} \quad (131)$$

for $i = 1, 2, 3, 4$.

Given $d_1 < d_2$, there are two possibilities for d_4 : (a) $d_4 - d_2 > \tau$ and (b) $0 \leq d_4 - d_2 \leq \tau$. There are four possibilities for d_3 : (A) $d_3 \in N_2 \setminus N_1$; (B) $d_3 \in N_1 \setminus N_2$; (C) $d_3 \in N_1 \cap N_2$; (D) $d_3 \notin N_1 \cup N_2$. There are eight combinations for the locations of (d_3, d_4) in total. However, by (67) the combination (a) & (D) is excluded. Our analysis next will exhaust all of the seven possibilities.

Case (a) & (A). Let $\Omega_{a,A}$ be the subset of $(d, d') \in \Omega_3$ satisfying restrictions (a) and (A), and others such as $\Omega_{b,C}$ are similarly defined. Thus,

$$\sum_{(d,d') \in \Omega_3} P(Z_d > a_n, Z_{d'} > a_n) \leq \sum_{\theta, \Theta} \sum_{(d,d') \in \Omega_{\theta, \Theta}} P(Z_d > a_n, Z_{d'} > a_n) \quad (132)$$

where θ runs over set $\{a, b\}$ and Θ runs over set $\{A, B, C, D\}$ but $(\theta, \Theta) \neq (a, D)$.

Easily, $|\Omega_{a,A}| \leq \tau p^3$ and the covariance matrix of $(w_{d_2}, w_{d_1}, w_{d_3}, w_{d_4})$ (see (70)) is

$$\begin{pmatrix} 1 & 0 & \gamma & 0 \\ 0 & 1 & 0 & 0 \\ \gamma & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad |\gamma| \leq 1.$$

Take $\epsilon = 1/2$ in Lemma 6.10 to have $P(Z_d > a_n, Z_{d'} > a_n) \equiv \rho_n = o(p^{-7/2})$ for all $(d, d') \in \Omega_{a,A}$. Thus

$$\sum_{(d,d') \in R} P(Z_d > a_n, Z_{d'} > a_n) = |R| \cdot \rho_n \rightarrow 0 \quad (133)$$

as $n \rightarrow \infty$ for $R = \Omega_{a,A}$.

Case (a) & (B). Notice $|\Omega_{a,B}| \leq \tau p^3$ and the covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is the same as that in Lemma 6.10. By the lemma we then have (133) for $R = \Omega_{a,B}$.

Case (a) & (C). Notice $|\Omega_{a,C}| \leq \tau^2 p^2$ and the covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is the same as that in Lemma 6.11. By the lemma, we know (133) holds for $R = \Omega_{a,C}$.

Case (b) & (A). In this case, $|\Omega_{b,A}| \leq \tau^2 p^2$ and the covariance matrix of $(w_{d_3}, w_{d_4}, w_{d_2}, w_{d_1})$ is the same as that in Lemma 6.11. By the lemma and using the fact that

$$P(Z_d > a_n, Z_{d'} > a_n) = P(Z_{(d_3, d_4)} > a_n, Z_{(d_2, d_1)} > a_n)$$

we see (133) holds with $R = \Omega_{b,A}$.

Case (b) & (B). In this case, $|\Omega_{b,B}| \leq \tau^2 p^2$ and the covariance matrix of $(w_{d_1}, w_{d_2}, w_{d_3}, w_{d_4})$ is the same as that in Lemma 6.12. By the lemma, we know (133) holds for $R = \Omega_{b,B}$.

Case (b) & (C). We assign positions for d_1, d_3, d_2, d_4 step by step: there are at most p positions for d_1 and at most k positions for each of d_3, d_2 and d_4 . Thus, $|\Omega_{b,C}| \leq \tau^3 p$. By (124),

$$P(Z_d > a_n, Z_{d'} > a_n) \leq P(Z_d > a_n) = P\left(\left|\sum_{i=1}^n \xi_i \eta_i\right| > a_n\right) = O\left(\frac{1}{p^2}\right)$$

as $n \rightarrow \infty$, where $\{\xi_i, \eta_i; i \geq 1\}$ are i.i.d. standard normals. Therefore, (133) holds with $R = \Omega_{b,C}$.

Case (b) & (D). In this case, $|\Omega_{b,D}| \leq \tau p^3$ and the covariance matrix of $(w_{d_4}, w_{d_3}, w_{d_2}, w_{d_1})$ is the same as that in Lemma 6.10. By the lemma and noting the fact that

$$P(Z_d > a_n, Z_{d'} > a_n) = P(Z_{(d_4, d_3)} > a_n, Z_{(d_2, d_1)} > a_n)$$

we see (133) holds with $R = \Omega_{b,D}$.

We obtain (75) by combining (133) for all the cases considered above with (132). ■