NECESSARY AND SUFFICIENT CONDITIONS FOR THE ASYMPTOTIC DISTRIBUTIONS OF COHERENCE OF ULTRA-HIGH DIMENSIONAL RANDOM MATRICES

BY QI-MAN SHAO[1] AND WEN-XIN ZHOU

Chinese University of Hong Kong and Hong Kong University of Science and Technology

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a random sample from a p-dimensional population distribution, where $p = p_n \to \infty$ and $\log p = o(n^\beta)$ for some $0 < \beta \leq 1$, and let L_n be the *coherence* of the sample correlation matrix. In this paper it is proved that $\sqrt{n/\log p}\, L_n \to 2$ in probability if and only if $Ee^{t_0|x_{11}|^\alpha} < \infty$ for some $t_0 > 0$, where α satisfies $\beta = \alpha/(4 - \alpha)$. Asymptotic distributions of L_n are also proved under the same sufficient condition. Similar results remain valid for *m-coherence* when the variables of the population are m dependent. The proofs are based on self-normalized moderate deviations, the Stein–Chen method and a newly developed randomized concentration inequality.

1. Introduction. This paper is motivated by the recent results of Cai and Jiang (2011, 2012) on asymptotic behaviors of the largest magnitude of off-diagonal entries of the sample correlation matrix. Consider a p-variable population represented by a random vector $\mathbf{x} = (x_1, \ldots, x_p)^T$ with the covariance matrix Σ, and let $X_n = (x_{ij})$ be an $n \times p$ random matrix where the n rows consist a random sample of size n from the population. The Pearson correlation coefficient ρ_{ij} between the ith and jth columns of X_n is given by

$$(1.1) \qquad \rho_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \cdot \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}, \qquad 1 \leq i, j \leq p,$$

where $\bar{x}_i = (1/n)\sum_{k=1}^n x_{ki}$. Then the sample correlation matrix Γ_n is defined by $\Gamma_n \equiv (\rho_{ij})$.

The main object of interest in this paper is the largest magnitude of off-diagonal entries of the sample correlation matrix, that is,

$$(1.2) \qquad L_n = \max_{1 \leq i < j \leq p} |\rho_{ij}|.$$

As in Cai and Jiang (2011), L_n is called the *coherence* of the random matrix X_n.

In the case where p and n are of the same order, that is, $n/p \to \lambda \in (0, \infty)$, asymptotic properties of coherence L_n have been extensively studied recently. Jiang (2004) was the first to establish the strong laws and limiting distributions of L_n. The moment assumption in Jiang (2004) has been substantially improved by Li and Rosalsky (2006), Zhou (2007), Liu, Lin and Shao (2008), Li, Liu and Rosalsky (2010) and Li, Qi and Rosalsky (2012). Liu, Lin and Shao (2008) proved that similar results hold for $p = O(n^{\alpha})$ where α is a constant. We refer to Cai and Jiang (2011) and references therein for recent developments on this topic. In particular, Cai and Jiang (2011) considered the ultra-high dimensional case where p can be as large as $e^{n^{\beta}}$ for some $\beta \in (0, 1)$. Specifically, assuming all the entries of X_n, $\{x_{ij}, i \geq 1, j \geq 1\}$ are i.i.d. real-valued random variables with mean μ and variance $0 < \sigma^2 < \infty$, they proved the following results.

Suppose $\mathbb{E}e^{t_0|x_{11}|^{\alpha}} < \infty$ for some $t_0 > 0$ and $\alpha > 0$. Assume that $p = p_n \to \infty$ and $\log p = o(n^{\beta})$ as $n \to \infty$, where $\beta = \frac{\alpha}{4+\alpha}$. Then

$$(1.3) \qquad \sqrt{n/(\log p)}\, L_n \to 2 \qquad \text{in probability.}$$

If $0 < \alpha \leq 2$, then

$$(1.4) \qquad nL_n^2 - 4\log p + \log_2 p \xrightarrow{d.} Y,$$

where d. denotes convergence in distribution, $\log_2 p \equiv \log\log p$ and the random variable Y has an extreme distribution of type I with distribution function

$$(1.5) \qquad F_Y(y) = e^{-(1/\sqrt{8\pi})e^{-y/2}}, \qquad y \in \mathbb{R}.$$

The main purpose of this paper is to find necessary and sufficient conditions for (1.3) and (1.4). Our result shows that the optimal choice of β is that $\beta = \alpha/(4-\alpha)$, $0 < \alpha \leq 2$ for (1.3), and the same β for (1.4) when $0 < \alpha \leq 1$. It is also shown that, when $1 < \alpha \leq 4/3$ and $\mathbb{E}(x_{11} - \mu)^3 \neq 0$, (1.4) does not hold, but a recentered L_n will do.

The rest of the paper is organized as follows. The main results, Theorems 2.1, 2.2 and 2.3 will be stated in Section 2. A closely related problem of testing for m-dependence of the population is considered and an application to compressed sensing is revisited in this section. The proofs of Theorems 2.1 and 2.2 are given in Sections 3 and 4, respectively, by using the Stein–Chen method, moderate deviations for both standardized and self-normalized sums of independent random variables. The proof of Theorem 2.3 is postponed to Section 5.

2. Main results. In this section, we consider the law of large numbers and asymptotic distributions of the *coherence L_n*. In Section 2.1, we provide necessary and sufficient conditions for the two aforementioned limiting properties and the optimal choice of β in terms of α. In Section 2.2, we consider the *m-coherence*, $L_{n,m}$, of a random matrix with m-dependent structure in each row.

NOTATION. Throughout this paper, $a_n \asymp b_n$ will denote that there exist two positive constants c_1, c_2 such that $c_1 \leq a_n/b_n \leq c_2$, for all $n \geq 1$; $a_n \sim b_n$ will denote $\lim_{n\to\infty} a_n/b_n = 1$.

2.1. *The i.i.d. case.* In this subsection, we assume that the entries x_{ij} of X_n are i.i.d. with mean μ and variance $\sigma^2 > 0$. Let

$$(2.1) \qquad \beta = \beta_\alpha = \alpha/(4 - \alpha), \qquad 0 < \alpha \leq 2.$$

We first state the law of large numbers for L_n.

THEOREM 2.1. (i) *Suppose* $\mathbb{E} \exp\{t_0 |x_{11}|^\alpha\} < \infty$ *for some* $0 < \alpha \leq 2$ *and* $t_0 > 0$. *Assume* $p = p_n \to \infty$ *and* $\log p = o(n^{\beta_\alpha})$ *as* $n \to \infty$. *Then*

$$(2.2) \qquad \sqrt{n/(\log p)} L_n \to 2$$

in probability as $n \to \infty$.
(ii) *Let* $0 < \beta \leq 1$. *If* (2.2) *holds for any* $p \to \infty$ *satisfying* $\log p = o(n^\beta)$, *then* $\mathbb{E} \exp\{t_0 |x_{11}|^\alpha\} < \infty$ *for some* $t_0 > 0$, *where* $\alpha = \alpha_\beta = 4\beta/(1 + \beta)$; *that is, α and β satisfy* (2.1).

REMARK 2.1. Clearly, when $\alpha = 2$, β equals to 1, so the range for dimension p reduces to $\log p = o(n)$. On the other hand, as proved by Cai and Jiang (2012), if $x_{11} \sim \mathcal{N}(0, 1)$ and $(\log p)/n \to \gamma \in (0, \infty)$, then

$$L_n \to \sqrt{1 - e^{-4\gamma}} > 0 \qquad \text{in probability as } n \to \infty.$$

Hence, result (2.2) no longer holds for $\log p \asymp n$. We believe that the limit of L_n will also depend on the distribution of x_{11} in this case, which still remains an open question.

The next theorem gives the asymptotic distribution of L_n after proper normalization. Let $\kappa = \mathbb{E}(x_{11} - \mu)^3/\sigma^3$ and

$$(2.3) \quad W_n = \begin{cases} nL_n^2 - 4\log p + \log_2 p, & 0 < \alpha \leq 1, \\ nL_n^2 - 4\log p - (8\kappa^2/3)n^{-1/2}(\log p)^{3/2} \\ \qquad + \log_2 p, & 1 < \alpha \leq 4/3. \end{cases}$$

THEOREM 2.2. *Suppose* $\mathbb{E} \exp\{t_0 |x_{11}|^\alpha\} < \infty$ *for some* $0 < \alpha \leq 4/3$ *and* $t_0 > 0$. *Assume* $p = p(n) \to \infty$, $\log p = o(n^{\beta_\alpha})$ *as* $n \to \infty$. *Then*

$$(2.4) \qquad W_n \xrightarrow{d} Y,$$

where Y has the distribution function given in (1.5).

Clearly, when $\alpha = 4/3$, $\beta_\alpha = 1/2$, (2.4) converges weakly to the distribution function (1.5) provided that $\log p = o(n^{1/2})$. However, (2.4) is not valid when $\log p \asymp n^{1/2}$ as shown in Cai and Jiang (2012); that is, if $x_{11} \sim \mathcal{N}(0,1)$ and $(\log p)/n^{1/2} \to \gamma \in [0, \infty)$, the limiting distribution of (1.4) is shifted to the left by $8\gamma^2$, that is, $\exp\{-(1/\sqrt{8\pi})e^{-(y+8\gamma^2)/2}\}$, $y \in \mathbb{R}$. For $4/3 < \alpha \leq 2$, derivation of the limiting distribution of L_n needs more delicate arguments.

Theorems 2.1 and 2.2 together fully exhibit the dependence between ranges of dimension p and the optimal moment conditions for asymptotic properties (1.3) and (1.4) of the coherence L_n.

REMARK 2.2.   It is known that the convergence rate to type I extreme distribution is typically slow. When $p \asymp n$, Liu, Lin and Shao (2008) proved that the rate of convergence can be improved to $O((\log n)^{5/2}n^{-1/2})$ if an "intermediate" approximation is used, that is,

$$\sup_{y \in \mathbb{R}} \left| P(nL_n^2 \leq y) - \exp\left\{ -\frac{p(p-1)}{2} P(\chi_1^2 \geq y) \right\} \right|$$

(2.5)

$$= O\left(\frac{(\log n)^{5/2}}{n^{1/2}} \right),$$

where χ_1^2 has a chi-square distribution with one degree of freedom. In the ultra-high dimensional case, Theorem 2.2 implies

$$\sup_{y \in \mathbb{R}} \left| P(W_n \leq y) - \exp\left\{ -\frac{p(p-1)}{2} P(\chi_1^2 \geq 4\log p - \log\log p + y) \right\} \right|$$

(2.6)

$$\to 0.$$

It is possible to prove that the rate of convergence of (2.6) is of order $O(n^{-1/2})$. To test the independence of the p-variate population, it may be better to choose the critical value based on the "intermediate" approximation. That is, reject the null hypothesis if $L_n^2 \geq z_\alpha/n$, where z_α satisfies $P(\chi_1^2 \geq z_\alpha) = -2\log(1-\alpha)/\{p(p-1)\}$.

REMARK 2.3.   Both Theorems 2.1 and 2.2 are still valid if L_n is replaced by

(2.7)
$$\tilde{L}_n = \max_{1 \leq i < j \leq p} |\tilde{\rho}_{ij}|,$$

where

(2.8)
$$\tilde{\rho}_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \mu)(x_{kj} - \mu)}{\sqrt{\sum_{k=1}^{n}(x_{ki} - \mu)^2 \sum_{k=1}^{n}(x_{kj} - \mu)^2}}.$$

The quantity $\tilde{L}_n$ arises from compress sensing literature. See, for example, Donoho, Elad and Temlyakov (2006).

2.2. *m-dependent case.* As discussed in Cai and Jiang (2011), a variant of coherence L_n can be used to construct a test for bandedness of the covariance matrix in the Gaussian case. In this paper, we drop the normality assumption and consider a more general problem of testing whether the population is m-dependent, where m can depend on n. More specifically, let $X_n = (x_{ij})_{n \times p}$, where the n rows are i.i.d. random vectors drawn from a p-variate population represented by $\mathbf{x} = (x_1, \ldots, x_p)^T$ with the covariance matrix Σ. Assume all p components of $\mathbf{x}$ are identically distributed with mean μ and variance $\sigma^2 > 0$. Then, we wish to test the hypothesis

(2.9) $$H_0 : x_i \text{ and } x_j \text{ are independent for all } |i - j| \geq m.$$

Analogous to the definition of L_n, we introduce the *m-coherence* of the matrix X_n as follows:

(2.10) $$L_{n,m} = \max_{|i-j| \geq m} |\rho_{ij}|.$$

In addition, let $(r_{ij})_{p \times p}$ be the correlation matrix of $\mathbf{x}$. For any given $0 < \delta < 1$, set

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

The following theorem establishes the limiting distribution of $L_{n,m}$ under the null hypothesis.

THEOREM 2.3. *Let* $\kappa = \mathbb{E}(x_{11} - \mu)^3/\sigma^3$ *and define*

$$W_{n,m} = \begin{cases} nL_{n,m}^2 - 4\log p + \log_2 p, & 0 < \alpha \leq 1, \\ nL_{n,m}^2 - 4\log p - (8\kappa^2/3)n^{-1/2}(\log p)^{3/2} + \log_2 p, & 1 < \alpha \leq 4/3. \end{cases}$$

Suppose $\mathbb{E}\exp\{t_0|x_{11}|^\alpha\} < \infty$ *for some* $0 < \alpha \leq 4/3$ *and* $t_0 > 0$*. Moreover, assume that, as* $n \to \infty$:

(i) $p = p_n \to \infty$, $\log p = o(n^{\beta_\alpha})$, *where* β_α *is given in* (2.1);
(ii) *there exists some* $\delta \in (0, 1)$ *such that* $|\Gamma_{p,\delta}| = o(p)$ *and* $m = o(p^{\varepsilon_\delta})$*, where* $\varepsilon_\delta = (2\delta - \delta^2)/(4 - 2\delta + \delta^2)$.

Then, under H_0*,* $W_{n,m}$ *converges weakly to the extreme distribution* (1.5).

Theorem 2.3 was proved in Cai and Jiang (2011) when $\mathbf{x}$ is multivariate normal, $\log p = o(n^{1/3})$, $m = o(p^t)$ for any $t > 0$ and $|\Gamma_{p,\delta}| = o(p)$ for some $\delta \in (0, 1)$. It was also pointed out therein that the assumption $|\Gamma_{p,\delta}| = o(p)$ is essential in the sense that there exists a covariance matrix Σ such that the conclusion of Theorem 2.3 for Gaussian entries no longer holds when $p \sim ne^{n^{1/4}}$, $m = n$ and $|\Gamma_{p,\delta}| = p$ for any $\delta > 0$. In Theorem 2.3 here, the assumption on m is weakened, and condition (i) provides the optimal choice of β in terms of α, and more importantly, Gaussian entries are not required.

REMARK 2.4. Similar to Remark 2.2, an "intermediate" approximation can also be applied here based on

$$
(2.12) \quad \begin{aligned} \sup_{y \in \mathbb{R}} \big| P(W_{n,m} \leq y) \\ - \exp\{-(p^2/2) P(\chi_1^2 \geq 4 \log p - \log \log p + y)\} \big| \to 0 \end{aligned}
$$

as $n \to \infty$.

REMARK 2.5. In compressed sensing, the quantity $\tilde{L}_n$, defined in (2.7), is useful because it is closely related to the so-called *mutual incoherence property* (MIP), which requires the pairwise correlations among column vectors of $X = X_{n \times p}$ to be small. More specifically, under certain assumptions on X, the condition

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

guarantees the exact recovery of $\beta \in \mathbb{R}^p$ from linear measurements $y = X\beta$, when β has at most k nonzero entries. This condition is also sharp in the sense that there exists matrices X_0 such that recovering some k-sparse signals β based on $y = X_0 \beta$ when $(2k-1)\tilde{L}_n = 1$ is impossible. See, Donoho and Huo (2001), Fuchs (2004) and Cai, Wang and Xu (2010).

It was shown in Cai and Jiang (2011) that the limiting properties of $\tilde{L}_n$ can be directly applied to compute the probability that random measurement matrices satisfy the MIP conditions (2.13). In particular, Theorem 2.1 with L_n replaced with $\tilde{L}_n$ provides necessary and sufficient conditions for $\tilde{L}_n \sim 2\sqrt{(\log p)/n}$. This suggests that the sparsity k should satisfy $k < \sqrt{n/(\log p)}/4$ approximately in order for the MIP condition (2.13) to hold.

3. Proof of Theorem 2.1. We start with collecting some technical lemmas that will be used to prove our main results. Without loss of generality, assume $\{x_{ij}; 1 \leq i \leq n, 1 \leq j \leq p\}$ are i.i.d. random variables with mean zero and variance one. Both letters C and c denote constants that do not depend on n or p, but may depend on the distribution of x_{11} and vary from line to line.

3.1. *Technical lemmas.* As in many previous works on the extreme distribution approximation, the following lemma is a special case of Theorem 1 of Arratia, Goldstein and Gordon (1989), based on the Stein–Chen method.

LEMMA 3.1. *Let $\{\eta_\alpha, \alpha \in I\}$ be random variables on an index set I. For each $\alpha \in I$, let B_α be a subset of I with $\alpha \in B_\alpha$. For any given $t \in \mathbb{R}$, set $\lambda = \sum_{\alpha \in I} P(\eta_\alpha > t)$. Then*

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

where

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

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

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

For a sequence of random variables $X_1, X_2, \ldots$, we use S_n and V_n^2 to denote the partial sum and the partial quadratic sum, respectively, that is,

$$S_n = \sum_{i=1}^{n} X_i, \qquad V_n^2 = \sum_{i=1}^{n} X_i^2.$$

The following lemma is due to Linnik (1961) on the moderate deviation under i.i.d. assumption.

LEMMA 3.2. *Suppose* $X_1, X_2, \ldots$ *are i.i.d. random variables with* $\mathbb{E}X_1 = 0$ *and* $\mathbb{E}X_1^2 = 1$:

(i) *If* $\mathbb{E}e^{t_0|X_1|^\alpha} < \infty$ *for some* $0 < \alpha \leq 1$ *and* $t_0 > 0$, *then*

$$(3.2) \qquad \lim_{n \to \infty} \frac{1}{x_n^2} \log P(S_n/\sqrt{n} \geq x_n) = -1/2$$

for any $x_n \to \infty$, $x_n = o(n^{\alpha/(2(2-\alpha))})$.

(ii) *If* $\mathbb{E}e^{t_0|X_1|^\alpha} < \infty$ *for some* $0 < \alpha \leq 1/2$ *and* $t_0 > 0$, *then*

$$(3.3) \qquad \frac{P(S_n/\sqrt{n} \geq x)}{1 - \Phi(x)} \to 1$$

holds uniformly for $0 \leq x \leq o(n^{\alpha/(2(2-\alpha))})$.

(iii) *Assume* $\mathbb{E}e^{t_0 X_1} < \infty$ *for some* $t_0 > 0$. *If* $x \geq 0$, $x = o(n^{1/4})$, *then*

$$(3.4) \qquad \frac{P(S_n/\sqrt{n} \geq x)}{1 - \Phi(x)} = \exp\left\{\frac{x^3 \mathbb{E}X_1^3}{6n^{1/2}}\right\}\left[1 + O\left(\frac{1+x}{n^{1/2}}\right)\right].$$

We also need the following self-normalized moderate deviations:

LEMMA 3.3 [Shao (1997)]. *Assume that* $X_1, X_2, \ldots$ *are i.i.d. random variables with* $\mathbb{E}X_1 = 0$ *and* $0 < \sigma^2 = \mathbb{E}X_1^2 < \infty$. *Then, for any sequence of real numbers* x_n *satisfying* $x_n \to \infty$ *and* $x_n = o(\sqrt{n})$,

$$(3.5) \qquad \log P(S_n/V_n \geq x_n) \sim -x_n^2/2.$$

3.2. *Proof of Theorem* 2.1.

PROOF OF (i).   The main idea of the proof is to show that L_n can be reduced to $L_{n,0} = \max_{1 \leq i < j \leq p} |\rho_{ij,0}|$, where

$$(3.6) \qquad \rho_{ij,0} = \frac{1}{n\sigma^2} \sum_{k=1}^{n} (x_{ki} - \mu)(x_{kj} - \mu), \qquad 1 \leq i, j \leq p.$$

Let

$$S_{n,i} = \sum_{k=1}^{n} x_{ki}, \qquad V_{n,i}^2 = \sum_{k=1}^{n} x_{ki}^2,$$

$$(3.7)$$

$$\Delta_{n,i} = \frac{S_{n,i}}{\sqrt{n} V_{n,i}}, \qquad 1 \leq i \leq p, n \geq 1.$$

Decompose the sample correlation coefficient as

$$(3.8) \qquad\qquad \rho_{ij} = \rho_{ij,1} - \rho_{ij,2}, \qquad 1 \leq i, j \leq p$$

and accordingly, define

$$L_{n,k} = \max_{1 \leq i < j \leq p} |\rho_{ij,k}|, \qquad k = 1, 2,$$

where

$$\rho_{ij,1} = \frac{\sum_{k=1}^{n} x_{ki} x_{kj} / (V_{n,i} V_{n,j})}{\{(1 - \Delta_{n,i}^2)(1 - \Delta_{n,j}^2)\}^{1/2}},$$

$$(3.9)$$

$$\rho_{ij,2} = \frac{\Delta_{n,i} \Delta_{n,j}}{\{(1 - \Delta_{n,i}^2)(1 - \Delta_{n,j}^2)\}^{1/2}}.$$

Intuitively, Lemma 3.3 suggests that $\Delta_{n,i}$ can be negligible and Lemma 3.2 indicates that $V_{n,i}^2/n$ is close to 1. Let

$$(3.10) \qquad \varepsilon_{n1} = c_1 (\log p)^{1/2} / n^{\beta/2} \quad \text{and} \quad \varepsilon_{n2} = c_2 (\log p)^{1/2} / n^{1/2},$$

where c_1 and c_2 are positive constants only depending on the distribution of x_{11} and will be specified later in different cases. Since $\mathbb{E} \exp\{t_0 |x_{11}^2 - 1|^{\alpha/2}\} < \infty$, it follows from (3.2) and (3.5) that

$$(3.11) \qquad P(|V_{n,1}^2 - n|/n^{1/2} > \varepsilon_{n1} n^{\beta/2}) \leq 2 \exp\{-c \varepsilon_{n1}^2 n^{\beta}\}$$

and

$$(3.12) \qquad\qquad P(|\Delta_{n,1}| > \varepsilon_{n2}) \leq 2 \exp\{-c \varepsilon_{n2}^2 n\}$$

for all sufficiently large n. Now define the subset

$$(3.13) \qquad \mathcal{E}_n = \left\{ \max_{1 \leq i \leq p} |V_{n,i}^2/n - 1| \leq \varepsilon_{n1} n^{(\beta-1)/2}, \max_{1 \leq i \leq p} |\Delta_{n,i}| \leq \varepsilon_{n2} \right\}.$$

Then, for properly chosen c_1 and c_2 in (3.10), we have

$$(3.14) \qquad P(\mathcal{E}_n^c) \leq 2p(\exp\{-c\varepsilon_{n1}^2 n^\beta\} + \exp\{-c\varepsilon_{n2}^2 n\}) = o(p^{-4}).$$

Recall $L_{n,0}$ defined through (3.6). Clearly, on $\mathcal{E}_n$

$$\frac{L_{n,0}}{1 + \varepsilon_{n1}n^{(\beta-1)/2}} \leq L_{n,1} \leq \frac{L_{n,0}}{(1 - \varepsilon_{n2}^2)(1 - \varepsilon_{n1}n^{(\beta-1)/2})}$$

and

$$L_{n,2} \leq \varepsilon_{n2}^2/(1 - \varepsilon_{n2}^2).$$

Noting that $\varepsilon_{n1}n^{(\beta-1)/2} = c_1(\log p)^{1/2}/n^{1/2} = o(1)$ and $\sqrt{n/\log p}\,\varepsilon_{n2}^2 = c_2^2(\log p)^{1/2}/n^{1/2} = o(1)$, we have on $\mathcal{E}_n$

$$(3.15) \qquad L_{n,1}/L_{n,0} \to 1, \qquad \sqrt{n/\log p}|L_n - L_{n,1}| \to 0,$$

which together with (3.14) shows that conclusion (2.2) will be a direct consequence of the next proposition. The proof is postponed to the end of this section.

$\square$

PROPOSITION 3.1. *Under the conditions of* (i) *in Theorem* 2.1, *we have* $\sqrt{n/(\log p)}L_{n,0} \to 2$ *in probability as* $n \to \infty$.

PROOF OF (ii). We shall prove the necessity of moment conditions under a weaker assumption than (2.2). Assume that there exists a constant $C_0 \geq 4$, such that

$$(3.16) \qquad P\left(\sqrt{n/(\log p)} \max_{1 \leq i < j \leq p} |\rho_{ij}| \geq C_0\right) \to 0.$$

Note that $\max_{1 \leq i < j \leq p} |\rho_{ij}| \geq \max_{1 \leq i \leq p/2} |\rho_{i,[p/2]+i}|$, then (3.16) implies

$$(3.17) \qquad P\left(\max_{1 \leq i \leq p/2} |\rho_{i,[p/2]+i}| > C_0\sqrt{(\log p)/n}\right) \to 0.$$

Observe that $\{\rho_{i,[p/2]+i}, \ 1 \leq i \leq [p/2]\}$ are i.i.d. random variables and that $\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \leq \sum_{k=1}^n x_{ki}^2$, (3.17) thus yields

$$(3.18) \qquad p \cdot P\left(\frac{|\sum_{k=1}^n x_{k1}x_{k2} - n\bar{x}_1\bar{x}_2|}{(\sum_{k=1}^n x_{k1}^2)^{1/2}(\sum_{k=1}^n x_{k2}^2)^{1/2}} > C_0\sqrt{(\log p)/n}\right) \to 0.$$

For $n \geq 16$, define the subset

$$\mathcal{D}_n = \left\{\frac{\sum_{k=2}^n x_{ki}^2}{n} \leq 2, \frac{|\sum_{k=2}^n x_{ki}|}{\sqrt{n}} \leq n^{1/4}, i = 1, 2; \frac{|\sum_{k=2}^n x_{k1}x_{k2}|}{\sqrt{n}} \leq 1\right\}.$$

By the central limit theorem and the strong law of large numbers, $P(\mathcal{D}_n) \to 2\Phi(1) - 1$, so that $P(\mathcal{D}_n) \geq 1/2$ for sufficiently large n. Furthermore, since $\log p = o(n)$, we have on $\mathcal{D}_n$,

$$\left\{ \frac{|\sum_{k=1}^{n} x_{k1} x_{k2}|}{(\sum_{k=1}^{n} x_{k1}^2)^{1/2}(\sum_{k=1}^{n} x_{k2}^2)^{1/2}} > C_0 \sqrt{\frac{\log p}{n}} \right\}$$

$$\supseteq \left\{ \frac{|x_{11} x_{12}| - 2\sqrt{n} - |x_{11}| - |x_{12}|}{(x_{11}^2 + 2n)^{1/2}(x_{12}^2 + 2n)^{1/2}} > C_0 \sqrt{\frac{\log p}{n}} \right\}$$

$$\supseteq \left\{ (|x_{11}| - c\sqrt{\log p})(|x_{12}| - c\sqrt{\log p}) > 3C_0 \sqrt{n \log p} \right\}$$

for some $c > 0$, which along with the independence of $\mathcal{D}_n$ and $\{x_{11}, x_{12}\}$ yields

$$P\left(\frac{|\sum_{k=1}^{n} x_{k1} x_{k2} - n\bar{x}_1 \bar{x}_2|}{(\sum_{k=1}^{n} x_{k1}^2)^{1/2}(\sum_{k=1}^{n} x_{k2}^2)^{1/2}} > C_0 \sqrt{(\log p)/n} \right)$$

$$(3.19) \qquad \geq P(\mathcal{D}_n) \cdot P\left((|x_{11}| - c\sqrt{\log p})(|x_{12}| - c\sqrt{\log p}) > 3C_0 \sqrt{n \log p} \right)$$

$$\geq (1/2) \cdot \left\{ P(|x_{11}| > 2C_0^{1/2}(n \log p)^{1/4}) \right\}^2.$$

If follows from (3.18) and (3.19) that

$$(3.20) \qquad p^{1/2} P(|x_{11}| > C_0 (n \log p)^{1/4}) = o(1)$$

for any p satisfying $\log p = o(n^\beta)$. By a contradiction argument, it is easy to see that (3.20) implies that $\mathbb{E} \exp\{t_0 |x_{11}|^{4\beta/(1+\beta)}\} < \infty$, for some $t_0 > 0$. This proves part (ii). $\square$

We end this section with the proof of Proposition 3.1.

3.3. *Proof of Proposition* 3.1. It suffices to show, for any $0 < \varepsilon < 1/8$, as $n \to \infty$,

$$(3.21) \qquad P\left(\sqrt{n/(\log p)} L_{n,0} \leq 2 - \varepsilon \right) \to 0$$

and

$$(3.22) \qquad P\left(\sqrt{n/(\log p)} L_{n,0} > 2 + \varepsilon \right) \to 0.$$

We apply Lemma 3.1 to prove (3.21) by using (3.1) to deal with the maximum. The proof of (3.22) is similar, and so the details are omitted here.

Put $y_n = (2 - \varepsilon)\sqrt{(\log p)/n}$, $n \geq 1$. Define

$$I = \{(i, j); 1 \leq i < j \leq p\}, \qquad A_{ij} = \{|\rho_{ij,0}| > y_n\}, \qquad 1 \leq i < j \leq p,$$

and

$$B_{i,j} = \{(k, l) \in I \setminus \{(i, j)\}; \text{ either } k \in \{i, j\} \text{ or } l \in \{i, j\}\}.$$

Since $\{x_{ij}; (i, j) \in I\}$ are identically distributed, by Lemma 3.1,

$$(3.23) \quad \left| P\left(\max_{1 \le i < j \le p} |\rho_{ij,0}| \le (2 - \varepsilon)\sqrt{(\log p)/n} \right) - e^{-\lambda_n} \right| \le b_{n,1} + b_{n,2},$$

where

$$(3.24) \quad \begin{aligned} \lambda_n &= \frac{p(p - 1)}{2} P(A_{12}), \qquad b_{n,1} \le p^3 P^2(A_{12}), \\ b_{n,2} &\le p^3 P(A_{12} A_{13}). \end{aligned}$$

Because $0 < \alpha/2 \le 1$ and $\mathbb{E} \exp\{t_0 |x_{11} x_{12}|^{\alpha/2}\} < \infty$, it follows from (3.2) that, for all sufficiently large n,

$$(3.25) \quad \begin{aligned} P(A_{12}) &= P\left(\frac{|\sum_{k=1}^n x_{k1} x_{k2}|}{n^{1/2}} > \sqrt{n} y_n \right) \\ &\le 2 \exp\{-(1 - \varepsilon) n y_n^2/2\} = 2 p^{-(1-\varepsilon)(2-\varepsilon)^2/2}, \end{aligned}$$

which, in turn implies

$$(3.26) \quad \lambda_n \to \infty \quad \text{and} \quad b_{n,1} = o(1) \qquad \text{as } n \to \infty.$$

As for $b_{n,2}$, we have

$$(3.27) \quad \begin{aligned} P(A_{12} A_{13}) &= P\left(\frac{|\sum_{k=1}^n x_{k1} x_{k2}|}{n} > y_n, \frac{|\sum_{k=1}^n x_{k1} x_{k3}|}{n} > y_n \right) \\ &\le P\left(\frac{|\sum_{k=1}^n x_{k1}(x_{k2} + x_{k3})|}{n} > 2 y_n \right) \\ &\quad + P\left(\frac{|\sum_{k=1}^n x_{k1}(x_{k2} - x_{k3})|}{n} > 2 y_n \right). \end{aligned}$$

Since $\mathbb{E}[x_{k1}(x_{k2} + x_{k3})] = 0$ and $\mathbb{E}[x_{k1}(x_{k2} + x_{k3})]^2 = 2$, applying (3.2) again, we get

$$P\left(\frac{|\sum_{k=1}^n x_{k1}(x_{k2} + x_{k3})|}{n} > 2 y_n \right) \le 2 \exp\{-(1 - \varepsilon) n y_n^2\} = 2 p^{-(1-\varepsilon)(2-\varepsilon)^2}.$$

Similarly, the same result holds for $P(|\sum_{k=1}^n x_{k1}(x_{k2} - x_{k3})| > 2 y_n n)$. Therefore,

$$(3.28) \quad b_{n,2} \le p^3 P(A_{12} A_{13}) = O\left(p^{3-(1-\varepsilon)(2-\varepsilon)^2} \right) = o(1).$$

This completes the proof of (3.21) by (3.23), (3.24), (3.26) and (3.28).

4. Proof of Theorem 2.2. The main idea is to use Lemma 3.1 again. The proof of part (i) is standard while that of part (ii) requires a more delicate estimate of λ_n given in (3.24). In particular, we need a randomized concentration inequality in Lemma 4.2.

We formulate the proof into two cases.

Case 1. $0 < \alpha \leq 1$.

For arbitrary fixed $y \in \mathbb{R}$, let

$$(4.1) \qquad y_n = \sqrt{(y + 4\log p - \log_2 p)/n}, \qquad \log_2 p \equiv \log\log p$$

for large n so that $y + 4\log p - \log_2 p > 0$. We need to prove that

$$(4.2) \qquad P\Big(\max_{1 \leq i < j \leq p} |\rho_{ij}| \leq y_n \Big) \to \exp(-(1/\sqrt{8\pi})e^{-z/2}).$$

Similar to (3.23), we have

$$(4.3) \qquad \Big| P\Big(\max_{1 \leq i < j \leq p} |\rho_{ij}| \leq y_n \Big) - e^{-\lambda_n} \Big| \leq b_{n,1} + b_{n,2},$$

where λ_n, $b_{n,1}$, $b_{n,2}$ and A_{ij} are defined as in (3.24) with $\rho_{ij,0}$ replaced by ρ_{ij}. It suffices to show

$$(4.4) \qquad P(A_{12}) \sim 2\big(1 - \Phi(\sqrt{n}\,y_n)\big) + o\big(p^{-2}\big) \sim \frac{e^{-y/2}}{\sqrt{2\pi}} p^{-2}$$

and

$$(4.5) \qquad P(A_{12}A_{13}) = o\big(p^{-3}\big).$$

Analogously to (3.13), let

$$(4.6) \qquad \mathcal{E}_{n\cdot 3} = \Big\{ \max_{i=1,2,3} |V_{n,i}^2/n - 1| \leq \varepsilon_{n1} n^{(\beta-1)/2}, \ \max_{i=1,2,3} |\Delta_{n,i}| \leq \varepsilon_{n2} \Big\},$$

where $V_{n,i}$ and $\Delta_{n,i}$ are given in (3.7). In view of (3.14), we can choose c_1 and c_2 in (3.10) properly such that

$$(4.7) \qquad P\big(\mathcal{E}_{n\cdot 3}^c\big) = o\big(p^{-3}\big).$$

On $\mathcal{E}_{n\cdot 3}$, we have

$$(4.8) \qquad |\rho_{1i}| \leq \frac{|\rho_{1i,0}|}{(1 - \varepsilon_{n2}^2)(1 - \varepsilon_{n1} n^{(\beta-1)/2})} + \frac{\varepsilon_{n2}^2}{1 - \varepsilon_{n2}^2}, \qquad i = 2, 3,$$

and [recall $y_n \sim 2n^{-1/2}(\log p)^{1/2}$]

$$(4.9) \qquad |\rho_{12}| = \big\{1 + o\big(\sqrt{(\log p)/n}\big)\big\} \cdot |\rho_{12,0}| + O\big((\log p)/n\big).$$

We are now ready to prove (4.4) and (4.5).

PROOF OF (4.4).   By (4.9), it follows that, on $\mathcal{E}_{n\cdot 3}$,

$$\{|\rho_{12}| > y_n\} = \{|\rho_{12,0}| > \hat{y}_n\} \qquad \text{with } \hat{y}_n = y_n\big(1 + o\big(n^{-1/2}(\log p)^{1/2}\big)\big).$$

Recalling the definition of $\rho_{12,0}$ in (3.6) and

$$\mathbb{E}x_{k1}x_{k2} = 0, \qquad \mathbb{E}(x_{k1}x_{k2})^2 = 1, \qquad \mathbb{E}e^{t_0|x_{11}x_{12}|^{\alpha/2}} < \infty \qquad \text{with } 0 < \alpha/2 \le 1,$$

it follows directly from (3.3) that, as $n \to \infty$,

$$(4.10) \qquad \frac{P(\rho_{12,0} > \hat{y}_n)}{1 - \Phi(\sqrt{n}\hat{y}_n)} \to 1.$$

Noticing that $\log p = o(n^{1/3})$, it is easy to check that

$$\frac{1 - \Phi(\sqrt{n}y_n)}{1 - \Phi(\sqrt{n}\hat{y}_n)} \to 1,$$

which, together with (4.10) yields (4.4). $\quad\square$

PROOF OF (4.5). By (4.8), following the same argument as in (3.27) and (3.28), we have for any $0 < \varepsilon < 1/8$,

$$
\begin{aligned}
P(A_{12}A_{13}) \\
\le P\big(|\rho_{12,0}| \ge \{1 - o(1)\}y_n, |\rho_{13,0}| \ge \{1 - o(1)\}y_n\big) + P(\mathcal{E}_{n\cdot3}^c) \\
\le C \exp\{-(1-\varepsilon)ny_n^2\} + o(p^{-3}) \\
\le C(\log p)p^{-4(1-\varepsilon)} + o(p^{-3}) = o(p^{-3}).
\end{aligned}
$$

This gives (4.5).

Case 2. $1 < \alpha \le 4/3$.

Similar to y_n in (4.1), for $y \in \mathbb{R}$ we now define

$$(4.11) \qquad y_n = \sqrt{(y + 4\log p + c_{n,p} - \log_2 p)/n},$$

where $c_{n,p} = (8\kappa^2/3)n^{-1/2}(\log p)^{3/2}$. Following the same argument as in the proof of case 1, (4.5) remains valid. It thus remains to show that

$$(4.12) \qquad P(A_{12}) \sim 2\mathcal{L}_{n,y} + o(p^{-2}),$$

where

$$\mathcal{L}_{n,y} = \big(1 - \Phi(\sqrt{n}y_n)\big)\exp(\kappa^2 n y_n^3/6).$$

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ni})^T$, $i = 1, \ldots, p$ be the p columns of X_n, and $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$. Rewrite ρ_{12} as

$$(4.13) \qquad \rho_{12} = \hat{\rho}_{12}/\{(1 - \Delta_{n,1}^2)(1 - \Delta_{n,2}^2)\}^{1/2}$$

$$\text{with } \hat{\rho}_{12} \equiv \frac{\mathbf{x}_1^T\mathbf{x}_2 - n^{-1}S_{n,1}S_{n,2}}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|}.$$

Define the subset

$$(4.14) \qquad \mathcal{E}_{n\cdot 2} = \{\max(|\Delta_{n,1}|, |\Delta_{n,2}|) \leq \varepsilon_{n2}\},$$

where $\varepsilon_{n2} = c_2(\log p)^{1/2}/n^{1/2}$ is given in (3.10) with $c_2 > 0$ chosen appropriately such that $P(\mathcal{E}_{n\cdot 2}^c) = o(p^{-4})$. Hence, with probability at least $1 - o(p^{-4})$,

$$(4.15) \qquad |\rho_{12}|/|\hat{\rho}_{12}| = 1 + o(n^{-1/2}).$$

For $\hat{\rho}_{12}$, using the elementary inequalities

$$2ab \leq a^2 + b^2 \quad \text{and} \quad (1+s)^{1/2} \geq 1 + s/2 - s^2/2 \qquad \text{for any } s > -1$$

to give lower and upper bounds as follows:

$$(4.16) \qquad \{\hat{\rho}_{12} > y_n\} \supseteq \{\mathbf{x}_1^T \mathbf{x}_2 - y_n(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2)/2 > n^{-1} S_{n,1} S_{n,2}\}$$

and

$$\{\hat{\rho}_{12} > y_n\}$$
$$(4.17) \qquad \subseteq \{\mathbf{x}_1^T \mathbf{x}_2 - y_n(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2)/2$$
$$> n^{-1} S_{n,1} S_{n,2} - n y_n^2[(\|\mathbf{x}_1\|^2/n - 1)^2 + (\|\mathbf{x}_2\|^2/n - 1)^2]\}.$$

Therefore, in order to prove (4.12), we need to show the following two claims:

$$(4.18) \qquad P(\mathbf{x}_1^T \mathbf{x}_2 - y_n(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2)/2 > 0) \sim \mathcal{L}_{n,y} + o(p^{-2})$$

and

$$(4.19) \qquad P(\Delta_n < \mathbf{x}_1^T \mathbf{x}_2 - y_n(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2)/2 \leq 0) = o(1)\{\mathcal{L}_{n,y} + p^{-2}\},$$

where $\Delta_n = \Delta(S_{n,1}, S_{n,2}, V_{n,1}^2, V_{n,2}^2)$ is given by

$$(4.20) \qquad \Delta_n = n^{-1} S_{n,1} S_{n,2} - n y_n^2[(\|\mathbf{x}_1\|^2/n - 1)^2 + (\|\mathbf{x}_2\|^2/n - 1)^2]. \qquad \square$$

PROOF OF (4.18). Given two random vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$, truncate one of which as follows:

$$(4.21) \quad x_{k2}^\tau = x_{k2} I_{\{|x_{k2}| \leq \tau\}}, \qquad k = 1, \ldots, n, \qquad \text{with } \tau = \tau_n = t_0^{-1/\alpha} n^{\beta/\alpha}$$

and write

$$(4.22) \qquad \xi_k = \xi_{n,k} = y_n x_{k1} x_{k2}^\tau - y_n^2(x_{k1}^2 + x_{k2}^{\tau 2})/2, \qquad k = 1, \ldots, n.$$

By the union bound and Markov inequality,

$$(4.23) \qquad P\Big(\max_{1 \leq k \leq n} |x_{k2}| > \tau\Big) \leq \mathbb{E}[e^{t_0|x_{11}|^\alpha}] \cdot n e^{-n^\beta}$$

and it is easy to see that $\mathbf{x}_1^T \mathbf{x}_2 - y_n(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2)/2 = y_n^{-1} \sum_{k=1}^{n} \xi_k$ on $\{\max_k |x_{k2}| \le \tau\}$. We thus aim to estimate the probability $P(\sum_{k=1}^{n} \xi_k > 0)$. Since $\alpha > 1$ and $y_n \tau^{2-\alpha} = O((\log p)^{1/2}/n^{\beta/2}) = o(1)$, it follows that

$$\xi_k \le y_n \tau^{2-\alpha} |x_{k1}| |x_{k2}|^{\alpha-1} \le y_n \tau^{2-\alpha} (|x_{k1}|^\alpha + |x_{k2}|^\alpha)$$
$$= o(1)(|x_{k1}|^\alpha + |x_{k2}|^\alpha),$$

which, in turn, implies $\sup_{1 \le k \le n, n \ge 1} \mathbb{E} e^{\xi_k} < \infty$. Moreover, it is easy to verify that

$$\mathbb{E}\xi_k = -y_n^2 + y_n^2 \mathbb{E} x_{11}^2 I_{\{|x_{11}| > \tau\}}/2 = -y_n^2 \{1 + O(y_n^2)\},$$

$$\text{Var}(\xi_k) = y_n^2 \{1 + O(y_n^2)\} \quad \text{and} \quad \frac{\mathbb{E}(\xi_k - \mathbb{E}\xi_k)^3}{\text{Var}^{3/2}(\xi_k)} = (\mathbb{E} x_{11}^3)^2 + O(y_n).$$

Let $\mu_n = \sum_{k=1}^{n} \mathbb{E}\xi_k$ and $\sigma_n^2 = \sum_{k=1}^{n} \text{Var}(\xi_k)$, then $-\mu_n/\sigma_n = \sqrt{n} y_n \{1 + O(y_n^2)\}$. Moreover, noting that $\sqrt{n} y_n = o(n^{1/4})$ and $\kappa = \mathbb{E} x_{11}^3$ (with $\mu = 0$ and $\sigma^2 = 1$), it follows from (3.4) and the above facts that

$$P\left(\sum_{k=1}^{n} \xi_k > 0\right) = P\left(\frac{\sum_{k=1}^{n}(\xi_k - \mathbb{E}\xi_k)}{\sigma_n} > -\mu_n/\sigma_n\right)$$

$$\sim (1 - \Phi(-\mu_n/\sigma_n)) \exp\left(\frac{(-\mu_n/\sigma_n)^3}{6n^{1/2}} (\kappa^2 + O(y_n))\right)$$

$$\sim (1 - \Phi(\sqrt{n} y_n)) \exp\left\{\frac{\kappa^2 n y_n^3}{6}\right\} = \mathcal{L}_{n,y} \qquad \text{as } n \to \infty.$$

This, along with (4.23), implies (4.18) immediately. □

PROOF OF (4.19). This requires a more delicate analysis. The main idea is to apply a combination of the multivariate conjugate method and a randomized concentration inequality to the truncated variables as defined in (4.22) and (4.21). Further to the notation used in the proof of (4.18), let $\{\mathbf{y}_k = (x_{k1}, x_{k2}^\tau); 1 \le k \le n\}$ be a sequence of independent $\mathbb{R}^2$-valued random variables and let measurable function $g : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$(4.24) \qquad \forall (u, v) \in \mathbb{R}^2 \qquad g(u, v) = (uv, u^2, v^2).$$

Put

$$\mathbf{S}_n = \sum_{k=1}^{n} \mathbf{y}_k = \left(\sum_{k=1}^{n} x_{k1}, \sum_{k=1}^{n} x_{k2}^\tau\right)^T$$

and

$$\mathbf{V}_n = \sum_{k=1}^{n} g(\mathbf{y}_k) = \left(\sum_{k=1}^{n} x_{k1} x_{k2}^\tau, \sum_{k=1}^{n} x_{k1}^2, \sum_{k=1}^{n} x_{k2}^{\tau 2}\right)^T.$$

Let $\lambda_n = (y_n, -y_n^2/2, -y_n^2/2)^T \in \mathbb{R}^3$. Observe that $\xi_k = \xi_{n,k}$ given in (4.22) can be rewritten as $\lambda_n^T g(\mathbf{y}_k)$ that satisfy

$$(4.25) \qquad \max_{1 \le k \le n, n \ge 1} m_{n,k} < \infty,$$

where

$$m_{n,k} = \mathbb{E}e^{\xi_k} = \mathbb{E}[e^{\lambda_n^T g(\mathbf{y}_k)}].$$

Now, let $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_n$ be a sequence of independent $\mathbb{R}^2$-valued random variables such that $\hat{\mathbf{y}}_k$ has the following distribution:

$$(4.26) \qquad \forall B \in \mathcal{B}^2 \qquad P(\hat{\mathbf{y}}_k \in B) = \frac{1}{m_{n,k}} \mathbb{E}[e^{\lambda_n^T g(\mathbf{y}_k)} I_{\{\mathbf{y}_k \in B\}}].$$

Accordingly, put $\hat{\mathbf{S}}_n = \sum_{k=1}^n \hat{\mathbf{y}}_k$, $\hat{\mathbf{V}}_n = \sum_{k=1}^n g(\hat{\mathbf{y}}_k)$. The multivariate conjugate method says that, for any $C \in \mathcal{B}^5$,

$$(4.27) \qquad P\{(\mathbf{S}_n, \mathbf{V}_n) \in C\} = \mathbb{E}[e^{\lambda_n^T \hat{\mathbf{V}}_n} I_{\{(\hat{\mathbf{S}}_n, \hat{\mathbf{V}}_n) \in C\}}] \prod_{k=1}^n m_{n,k}.$$

In particular, define subsets

$$C_n = \{\mathbf{u} \in \mathbb{R}^5 : \Delta(u_1, u_2, u_4, u_5) \le u_3 - y_n(u_4 + u_5)/2 < 0\} \cap E_n,$$

$$E_n = \left\{\mathbf{u} \in \mathbb{R}^3 \times \mathbb{R}_+^2 : \frac{u_1}{\sqrt{u_4}} \le \varepsilon_{n2} n^{1/2}, \left|\frac{u_j}{n} - 1\right| \le \varepsilon_{n1} n^{(\beta-1)/2}, j = 4, 5\right\},$$

where in accordance with (4.20),

$$(4.28) \qquad \Delta(v_1, v_2, v_3, v_4) = n^{-1} v_1 v_2 - n y_n^2 [(v_3/n - 1)^2 + (v_4/n - 1)^2]$$

and $\{\varepsilon_{n1}, \varepsilon_{n2}; n \ge 1\}$ are given as in (3.10), such that

$$(4.29) \qquad P\{(\mathbf{S}_n, \mathbf{V}_n) \in E_n^c\} = o(p^{-4}).$$

By (4.27), we have

$$P\{(\mathbf{S}_n, \mathbf{V}_n) \in C_n\} = \left(\prod_{k=1}^n m_{n,k}\right) \times \mathbb{E}[e^{-\lambda_n^T \hat{\mathbf{V}}_n} I_{\{(\hat{\mathbf{S}}_n, \hat{\mathbf{V}}_n) \in C_n\}}]$$

$$(4.30) \qquad\qquad := \left(\prod_{k=1}^n m_{n,k}\right) \times K_n.$$

Let $\hat{\xi}_k = \lambda_n^T g(\hat{\mathbf{y}}_k)$ be the conjugate version of ξ_k. Then, by (4.26),

$$\mathbb{E}\hat{\xi}_k = \mathbb{E}[\xi_k e^{\xi_k}]/\mathbb{E}[e^{\xi_k}], \qquad \text{Var}(\hat{\xi}_k) = \mathbb{E}[\xi_k^2 e^{\xi_k}]/\mathbb{E}[e^{\xi_k}] - (\mathbb{E}\hat{\xi}_k)^2.$$

Put $\hat{\mu}_n = \sum_{k=1}^{n} \mathbb{E}\hat{\xi}_k$ and $\hat{\sigma}_n^2 = \sum_{k=1}^{n} \text{Var}(\hat{\xi}_k)$. Routine calculations show (recall $\kappa = \mathbb{E}x_{11}^3$)

$$\mathbb{E}[e^{\xi_k}] = 1 - y_n^2/2 + \kappa^2 y_n^3/6 + O(y_n^4),$$

$$\mathbb{E}[\xi_k e^{\xi_k}] = \kappa^2 y_n^3/2 + O(y_n^4),$$

$$\mathbb{E}[\xi_k^2 e^{\xi_k}] = y_n^2 + \kappa^2 y_n^3 + O(y_n^4).$$

Consequently,

$$(4.31) \qquad \hat{\mu}_n = \kappa^2 n y_n^3/2 + O(n y_n^4), \qquad \hat{\sigma}_n^2 = n y_n^2 + \kappa^2 n y_n^3 + O(n y_n^4)$$

and

$$(4.32) \qquad \prod_{k=1}^{n} m_{n,k} = \exp(-n y_n^2/2 + \kappa^2 n y_n^3/6 + O(n y_n^4)).$$

As for K_n in (4.30), we shall show that

$$(4.33) \qquad \sqrt{n} y_n K_n = o(1).$$

Now combining (4.30), (4.32), (4.33) and the well-known result $1 - \Phi(s) \sim (2\pi)^{-1/2} s^{-1} e^{-s^2/2}$ as $s \to \infty$, it follows

$$P\{(\mathbf{S}_n, \mathbf{V}_n) \in C_n\} = o(\mathcal{L}_{n,y}).$$

This, together with (4.23), (4.29) and the definition of C_n, gives (4.19).

PROOF OF (4.33). Observe that on the event $\{(\hat{\mathbf{S}}_n, \hat{\mathbf{V}}_n) \in C_n\}$,

$$(4.34) \qquad \lambda_n^T \hat{\mathbf{V}}_n = \sum_{k=1}^{n} \hat{\xi}_k \geq (y_n/n) \hat{S}_{n,1} \hat{S}_{n,2} - 2n^\beta y_n^3 \varepsilon_{n1}^2,$$

where $\hat{S}_{n,1} = \sum_{k=1}^{n} \hat{x}_{k1}$, $\hat{S}_{n,2} = \sum_{k=1}^{n} \hat{x}_{k2}^\tau$. Using Hölder's inequality gives

$$(4.35) \qquad \begin{aligned} K_n &\leq \left(\mathbb{E}e^{-2\lambda_n^T \hat{\mathbf{V}}_n} I_{\{(\hat{\mathbf{S}}_n, \hat{\mathbf{V}}_n) \in C_n\}}\right)^{1/2} \\ &\quad \times \left(P\left((y_n/n)\hat{S}_{n,1}\hat{S}_{n,2} - 2n^\beta y_n^3 \varepsilon_{n1}^2 \leq \sum_{k=1}^{n} \hat{\xi}_k < 0\right)\right)^{1/2} \\ &:= K_{n,1}^{1/2} \times K_{n,2}^{1/2}. \end{aligned}$$

We first estimate $K_{n,1}$. By (4.26),

$$\mathbb{E}[\hat{x}_{k1}] = m_{n,k}^{-1} \mathbb{E}[x_{k1} e^{\xi_k}] = -\kappa y_n^3/2 + O(y_n^4),$$

$$\mathbb{E}[\hat{x}_{k1}^2] = m_{n,k}^{-1} \mathbb{E}[x_{k1}^2 e^{\xi_k}] = 1 - y_n^2/2 - \kappa^2 y_n^3/2 + O(y_n^4)$$

and same expansions hold for $\mathbb{E}[\hat{x}_{k2}^{\tau}]$ and $\mathbb{E}[\hat{x}_{k2}^{\tau 2}]$ as well. Thus, for all sufficiently large n, $\sum_{k=1}^{n} \mathbb{E}\hat{x}_{k2}^{\tau 2} \leq n$ and on $\{(\hat{\mathbf{S}}_n, \hat{\mathbf{V}}_n) \in C_n\}$,

$$|\hat{S}_{n,1}| \leq \sqrt{2}\varepsilon_{n2}n, \qquad \sum_{k=1}^{n} \hat{x}_{k2}^{\tau 2} \leq 2n.$$

In view of (3.10) and (4.34),

$$
\begin{aligned}
(4.36) \quad -2\lambda_n^T \hat{\mathbf{V}}_n &\leq -2(y_n/n)\hat{S}_{n,1}(\hat{S}_{n,2} - \mathbb{E}\hat{S}_{n,2}) - 2y_n\mathbb{E}[\hat{x}_{12}^{\tau}]\hat{S}_{n,1} + 4n^{\beta}y_n^3\varepsilon_{n1}^2 \\
&\leq Cn^{-1/2}(\log p)Z_n + O\big(n^{-3/2}(\log p)^{5/2}\big),
\end{aligned}
$$

where

$$Z_n \equiv \frac{|\sum_{k=1}^{n}(\hat{x}_{k2}^{\tau} - \mathbb{E}\hat{x}_{k2}^{\tau})|}{4\sqrt{\sum_{k=1}^{n} \operatorname{Var}(\hat{x}_{k2}^{\tau})} + \sqrt{\sum_{k=1}^{n}(\hat{x}_{k2}^{\tau} - \mathbb{E}\hat{x}_{k2}^{\tau})^2}}.$$

Now we can use the following sub-Gaussian property of self-normalized sums [see Lemma 6.4 in Jing, Shao and Wang (2003)]:

LEMMA 4.1. *Let $\{X_i, 1 \leq i \leq n\}$ be a sequence of independent random variables with $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 < \infty$. Then, for $a > 0$,*

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \geq a\left(4D_n + \left(\sum_{i=1}^{n} X_i^2\right)^{1/2}\right)\right) \leq 8e^{-a^2/2},$$

where $D_n^2 = \sum_{i=1}^{n} \mathbb{E}X_i^2$.

Indeed, Lemma 4.1 implies $P(Z_n \geq a) \leq 8e^{-a^2/2}$, $\forall a > 0$. Hence,

$$\forall t > 0 \qquad \mathbb{E}e^{tZ_n} \leq 1 + 8\sqrt{2\pi}te^{t^2/2},$$

which together with (4.36) yields

$$(4.37) \qquad\qquad\qquad K_{n,1} = O(1).$$

Next, we estimate $K_{n,2}$. The key technical tool is the randomized concentration inequality below developed in Shao and Zhou (2012):

LEMMA 4.2. *Let $\eta_1, \ldots, \eta_n$ be independent random variables,*

$$W_n = \sum_{k=1}^{n} \eta_k$$

and let $\Delta_1 = \Delta_1(\eta_1, \ldots, \eta_n)$ and $\Delta_2 = \Delta_2(\eta_1, \ldots, \eta_n)$ be two measurable functions of $\eta_1, \ldots, \eta_n$. Assume that

$$\mathbb{E}\eta_k = 0 \qquad for\ k = 1, 2, \ldots, n \quad and \quad \sum_{k=1}^{n} \mathbb{E}\eta_k^2 = 1.$$

For each $1 \le k \le n$, let $\Delta_1^{(k)}$ and $\Delta_2^{(k)}$ be any random variables such that η_k and $(\Delta_1^{(k)}, \Delta_2^{(k)}, W_n - \eta_k)$ are independent. Then

$$P(\Delta_1 \le W_n \le \Delta_2)$$
$$\le 21\left(\sum_{k=1}^n \mathbb{E}|\eta_k|^3 + \mathbb{E}|\Delta_2 - \Delta_1|\right.$$
$$\left. + \sum_{k=1}^n \{\mathbb{E}|\eta_k(\Delta_1 - \Delta_1^{(k)})| + \mathbb{E}|\eta_k(\Delta - \Delta_2^{(k)})|\}\right).$$

We now let W_n be the standardized $\sum_{k=1}^n \hat{\boldsymbol{\xi}}_k$ given by

(4.38) $$W_n = \frac{1}{\hat{\sigma}_n}\left(\sum_{k=1}^n \hat{\boldsymbol{\xi}}_k - \hat{\mu}_n\right),$$

where $\hat{\mu}_n$ and $\hat{\sigma}_n$ are defined in (4.31). As a direct consequence of Lemma 4.2 by letting $\omega_k = (\hat{\boldsymbol{\xi}}_k - \mathbb{E}\hat{\boldsymbol{\xi}}_k)/\hat{\sigma}_n$,

$$\Delta_1 = -\hat{\mu}_n/\hat{\sigma}_n + y_n \hat{S}_{n,1}\hat{S}_{n,2}/(n\hat{\sigma}_n) - 2n^\beta y_n^3 \varepsilon_{n1}^2/\hat{\sigma}_n, \qquad \Delta_2 = -\hat{\mu}_n/\hat{\sigma}_n$$

and

$$\hat{S}_{n,1}^{(k)} = \hat{S}_{n,1} - \hat{x}_{k1}, \qquad \hat{S}_{n,2}^{(k)} = \hat{S}_{n,2} - \hat{x}_{k2}^\tau, \qquad 1 \le k \le n,$$

we have

$$P\left\{(y_n/n)\hat{S}_{n,1}\hat{S}_{n,2} - 2n^\beta y_n^3 \varepsilon_{n1}^2 \le \sum_{k=1}^n \hat{\boldsymbol{\xi}}_k < 0\right\}$$

$$\le 21\left(\hat{\sigma}_n^{-3}\sum_{k=1}^n \mathbb{E}|\hat{\boldsymbol{\xi}}_k|^3 + y_n(n\hat{\sigma}_n)^{-1}\mathbb{E}|\hat{S}_{n,1}\hat{S}_{n,2}|\right.$$

$$+ (\log p)^2 n^{-3/2} + y_n n^{-1}\hat{\sigma}_n^{-2}\sum_{k=1}^n \mathbb{E}|\hat{\boldsymbol{\xi}}_k\hat{x}_{k1}\hat{S}_{n,2}^{(k)} + \hat{\boldsymbol{\xi}}_k\hat{x}_{k2}^\tau\hat{S}_{n,1}^{(k)}|$$

$$\left. + y_n n^{-1}\hat{\sigma}_n^{-2}\sum_{k=1}^n \mathbb{E}|\hat{\boldsymbol{\xi}}_k\hat{x}_{k1}\hat{x}_{k2}^\tau|\right)$$

$$\le C\left(n^{-1/2} + n^{-3/2}(\mathbb{E}\hat{S}_{n,1}^2)^{1/2} \cdot (\mathbb{E}\hat{S}_{n,2}^2)^{1/2}\right.$$

$$\left. + n^{-2}\sum_{k=1}^n \{\mathbb{E}\hat{S}_{n,1}^{(k)2}\}^{1/2} + n^{-2}\sum_{k=1}^n \{\mathbb{E}\hat{S}_{n,2}^{(k)2}\}^{1/2}\right)$$

$$\le Cn^{-1/2}.$$

This, together with expressions (4.35) and (4.37), verify our claim (4.33) and thus complete the proof of case 2. $\quad\square$

5. Proof of Theorem 2.3. The main idea of the proof is similar to that of Theorem 2.2. We start with the following three technical lemmas, and their proofs are postponed to the end of this section.

Let $\{(z_{k1}, z_{k2}, z_{k3}, z_{k4})^T; k \geq 1\}$ be a sequence of i.i.d. random vectors with mean zero and common covariance matrix Σ_4, which will be specified under different settings. Set

$$D_{n,i}^2 = \sum_{k=1}^{n} z_{ki}^2, \qquad i \in \{1, 2, 3, 4\}.$$

Suppose $p = p_n \to \infty$, $\log p = o(n^\beta)$ as $n \to \infty$. For $y \in \mathbb{R}$, let

$$(5.1) \qquad y_n = \begin{cases} \sqrt{(y + 4\log p - \log_2 p)/n}, & 0 < \alpha \leq 1, \\ \sqrt{(y + 4\log p + c_{n,p} - \log_2 p)/n}, & 1 < \alpha \leq 4/3, \end{cases}$$

for large n, where $c_{n,p} = (8\kappa^2/3)n^{-1/2}(\log p)^{3/2}$.

LEMMA 5.1. *Assume*
$$\Sigma_4 = \begin{pmatrix} 1 & 0 & r & 0 \\ 0 & 1 & 0 & 0 \\ r & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad |r| \leq 1.$$

Then, for any $0 < \varepsilon < 1$,
$$\sup_{|r| \leq 1} P\left(\frac{|\sum_{k=1}^{n} z_{k1} z_{k2}|}{D_{n,1} D_{n,2}} > y_n, \frac{|\sum_{k=1}^{n} z_{k3} z_{k4}|}{D_{n,3} D_{n,4}} > y_n \right) = O(p^{-4(1-\varepsilon)}).$$

LEMMA 5.2. *Assume*
$$\Sigma_4 = \begin{pmatrix} 1 & 0 & r_1 & 0 \\ 0 & 1 & r_2 & 0 \\ r_1 & r_2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad |r_1| \leq 1, \qquad |r_2| \leq 1.$$

Then, for any $0 < \varepsilon < 1$,
$$\sup_{|r_1|, |r_2| \leq 1} P\left(\frac{|\sum_{k=1}^{n} z_{k1} z_{k2}|}{D_{n,1} D_{n,2}} > y_n, \frac{|\sum_{k=1}^{n} z_{k3} z_{k4}|}{D_{n,3} D_{n,4}} > y_n \right) = O(p^{-4(1-\varepsilon)}).$$

LEMMA 5.3. *Assume*
$$\Sigma_4 = \begin{pmatrix} 1 & 0 & r_1 & 0 \\ 0 & 1 & 0 & r_2 \\ r_1 & 0 & 1 & 0 \\ 0 & r_2 & 0 & 1 \end{pmatrix}, \qquad |r_1| \leq 1, \qquad |r_2| \leq 1.$$

Then, for any $\delta \in (0, 1)$,

$$\sup_{|r_1|,|r_2|\le 1-\delta} P\left(\frac{|\sum_{k=1}^n z_{k1}z_{k2}|}{D_{n,1}D_{n,2}} > y_n, \frac{|\sum_{k=1}^n z_{k3}z_{k4}|}{D_{n,3}D_{n,4}} > y_n\right) = O\left(p^{-2(1+\varepsilon_\delta)}\right),$$

where

$$\varepsilon_\delta = (2\delta - \delta^2)/(4 - 2\delta + \delta^2).$$

Back to the proof of Theorem 2.3, w.l.o.g., we assume $\mu = 0$ and $\sigma^2 = 1$. Following the arguments for Theorem 2.2, we sketch the proof as follows:

Step 1: We have

$$P\left(\max_{1\le i < j \le p, j-i\ge m} |\rho_{ij}| \le y_n\right) \to e^{-e^{-y/2}/\sqrt{8\pi}} \qquad \text{as } n \to \infty.$$

Set

(5.2) $$\Lambda_p = \left\{(i, j) : 1 \le i < j \le p, j - i \ge m, i, j \notin \Gamma_{p,\delta}\right\}$$

and

(5.3) $$L'_n = \max_{(i,j)\in\Lambda_p} |\rho_{ij}|.$$

Clearly,

(5.4)
$$P(L'_n > y_n) \le P\left(\max_{1\le i < j \le p, j-i\ge m} |\rho_{ij}| > y_n\right)$$
$$\le P(L'_n > y_n) + \sum P(|\rho_{ij}| > y_n),$$

where the last summation is carried out over all pairs (i, j) such that $1 \le i < j \le p$, $j - i \ge m$ and either i or j is in $\Gamma_{p,\delta}$. The total number of such pairs is no more than $2p|\Gamma_{p,\delta}| = o(p^2)$.

Under H_0, $\mathbf{x}_1$ and $\mathbf{x}_{m+1}$ are independent and identically distributed. Then, by (4.4) and (4.12), we have for all $0 < \alpha \le 4/3$,

(5.5) $$P(|\rho_{1,m+1}| > y_n) \sim \frac{e^{-y/2}}{\sqrt{2\pi}} p^{-2},$$

which, in turn, implies that the last summation in (5.4) is $o(1)$.

Step 2: In view of (5.4) and (5.5), it suffices to prove

(5.6) $$P(L'_n \le y_n) \to e^{-e^{-y/2}/\sqrt{8\pi}}.$$

We follow the lines of proof of Proposition 6.4 in Cai and Jiang (2011) with the help of Lemma 3.1 and Lemmas 5.1–5.3. For $(i, j) \in \Lambda_p$, set

$$B_{i,j} = \left\{(k, l) \in \Lambda_p \setminus \{(i, j)\}; \min\{|k - i|, |l - j|, |k - j|, |l - i|\} < m\right\}$$

and $A_{ij} = \{|\rho_{ij}| > y_n\}$ with y_n given in (5.1). Note that $|B_{i,j}| \le 4 \times (2m \times p) = 8mp$ and $(\mathbf{x}_i, \mathbf{x}_j)$ are independent of $\{(\mathbf{x}_k, \mathbf{x}_l); (k, l) \in \Lambda_p \setminus B_{i,j}\}$. By Lemma 3.1,

$$\text{(5.7)} \qquad |P(L_n' \le y_n) - e^{-\lambda_n}| \le b_{n,1} + b_{n,2},$$

where

$$\text{(5.8)} \qquad \begin{aligned} \lambda_n &= |\Lambda_p| P(A_{1,m+1}), \\ b_{n,1} &= \sum_{\substack{(i,j) \in \Lambda_p \\ (k,l) \in B_{i,j}}} P(A_{1,m+1})^2 \le 4mp^3 P(A_{1,m+1})^2 \end{aligned}$$

and

$$\text{(5.9)} \qquad b_{n,2} = \sum_{(i,j) \in \Lambda_p} \sum_{(k,l) \in B_{i,j}} P(A_{ij} A_{kl}).$$

Clearly, $|\{(i, j): j \ge i + m\}| = (p - m)(p - m + 1)/2$ and by definition (5.2),

$$||\Lambda_p| - |\{(i, j): j \ge i + m\}|| \le 2p|\Gamma_{p,\delta}| = o(p^2).$$

This implies $|\Lambda_p| \sim p^2/2$ by assumption on m, which, together with (5.5) gives

$$\text{(5.10)} \qquad \lambda_n \sim e^{-y/2}/\sqrt{8\pi} \quad \text{and} \quad b_{n,1} = o(1) \qquad \text{as } n \to \infty.$$

It remains to estimate $b_{n,2}$. Fix $(i, j) \in \Lambda_p$ and $(k, l) \in B_{i,j}$ with $i < j$ and $k < l$. Without loss of generality, assume $i \le k$ (the case $k < i$ can be identically proved), then by definition of $B_{i,j}$

$$\text{(5.11)} \qquad \min\{k - i, |k - j|, |l - j|\} < m.$$

Consider three different cases for the locations of (i, j) and (k, l) from the above restrictions:

(1) $i < j \le k < l, k - j < m$;
(2) $i \le k < l \le j, \min\{k - i, j - l\} < m$;
(3) $i \le k \le j \le l, \min\{k - i, j - k, l - j\} < m$.

Let Ω_ν be the subset of index (i, j, k, l) with restriction (ν) for $\nu = 1, 2, 3$ and formulate the estimation of $P(A_{ij} A_{kl})$ into three different cases accordingly.

Case (1). It is easy to see that $|\Omega_1| \le mp^3 = o(p^{3+\varepsilon_\delta})$. For fixed $(i, j, k, l) \in \Omega_1$, the covariance matrix of $(x_{1j}, x_{1i}, x_{1k}, x_{1l})$ is equal to

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

for some $|r| \le 1$. Now we apply Lemma 5.1 to bound $P(A_{ij} A_{kl})$. Put

$$\hat{\rho}_{st} = \frac{\sum_{k=1}^n x_{ks} x_{kt}}{V_{n,s} V_{n,t}}, \qquad 1 \le s < t \le p,$$

and analogously to (3.13), let

$$(5.12) \qquad \mathcal{E}_{n\cdot 4} = \Big\{ \max_{s \in \{i,j,k,l\}} |\Delta_{n,s}| \leq \varepsilon_{n2} \Big\},$$

where ε_{n2} are chosen of the same type as in (3.10) such that $P(\mathcal{E}_{n\cdot 4}^c) = o(p^{-4})$. On $\mathcal{E}_{n\cdot 4}$, we have

$$|\rho_{st}| \leq (|\hat{\rho}_{st}| + \varepsilon_{n2}^2)/(1 - \varepsilon_{n2}^2) \qquad \text{with } \varepsilon_{n2}^2 \asymp (\log p)/n,$$

which, together with Lemma 5.1 and the fact that $y_n \sim 2n^{-1/2}(\log p)^{1/2}$, implies that, for any $0 < \varepsilon < (1 - \varepsilon_\delta)/4$ and all sufficiently large n,

$$
\begin{aligned}
&P(A_{ij}A_{kl}) \\
(5.13) \qquad &\leq P\big(|\hat{\rho}_{ij}| > (1 + o(1))y_n, |\hat{\rho}_{kl}| > (1 + o(1))y_n\big) + o(p^{-4}) \\
&\leq Cp^{-4(1-\varepsilon)}
\end{aligned}
$$

and hence

$$(5.14) \qquad \sum_{\Omega_1} P(A_{ij}A_{kl}) = o(1).$$

We remark that the $o(1)$'s appeared in (5.13) are of order $n^{-1/2}(\log p)^{1/2}$.

Case (2). Decompose Ω_2 as

$$
\begin{aligned}
\Omega_2 = \ &\{(i, j, k, l) \in \Omega_2; k - i < m, j - l < m\} \\
&+ \{(i, j, k, l) \in \Omega_2; k - i < m, j - l \geq m\} \\
&+ \{(i, j, k, l) \in \Omega_2; k - i \geq m, j - l < m\} \\
:=\ &\Omega_{2,a} + \Omega_{2,b} + \Omega_{2,c}.
\end{aligned}
$$

Observe that $|\Omega_{2,a}| \leq m^2 p^2 = o(p^{2(1+\varepsilon_\delta)})$. For $(i, j, k, l) \in \Omega_{2,a}$, the covariance matrix of $(x_{1i}, x_{1j}, x_{1k}, x_{1l})$ is equal to

$$
\begin{pmatrix}
1 & 0 & r_1 & 0 \\
0 & 1 & 0 & r_2 \\
r_1 & 0 & 1 & 0 \\
0 & r_2 & 0 & 1
\end{pmatrix}
$$

for some $|r_1|, |r_2| \leq 1 - \delta$. Using Lemma 5.3, along the lines of the argument in case (1), we get

$$P(A_{ij}A_{kl}) \leq Cp^{-2(1+\varepsilon_\delta)}$$

and therefore

$$(5.15) \qquad \sum_{\Omega_{2,a}} P(A_{ij}A_{kl}) = o(1).$$

Clearly, $|\Omega_{2,b}| \le mp^3$ and $|\Omega_{2,c}| \le mp^3$. For (i, j, k, l) in either $\Omega_{2,b}$ or $\Omega_{2,c}$, the corresponding covariance matrix of $(x_{1i}, x_{1j}, x_{1k}, x_{1l})$ is

$$\text{either } \begin{pmatrix} 1 & 0 & r & 0 \\ 0 & 1 & 0 & 0 \\ r & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & r \\ 0 & 0 & 1 & 0 \\ 0 & r & 0 & 1 \end{pmatrix}, \qquad |r| \le 1.$$

By the same argument as that in the proof of (5.14), we have

$$(5.16) \qquad \sum_{\Omega_{2,b} \cup \Omega_{2,c}} P(A_{ij} A_{kl}) = o(1) \qquad \text{as } n \to \infty.$$

Case (3). We aim to show that

$$(5.17) \qquad \sum_{\Omega_3} P(A_{ij} A_{kl}) = o(1).$$

Essentially, this can be done by following similar arguments as in case (2). However, for $(i, j, k, l) \in \Omega_3$ which satisfies the restriction

$$\min\{k - i, j - k, l - j\} < m,$$

we need to decompose Ω_3 into seven disjoint subsets and estimate all the seven possibilities with the help of Lemmas 5.1–5.3 as before. The details are omitted here.

Finally, combining expressions (5.14), (5.15), (5.16) and (5.17) with (5.9), we get $b_{n,2} \to 0$ as $n \to \infty$. This completes the proof of (5.6). □

PROOF OF LEMMAS 5.1–5.3. We start with a general consideration for estimating joint probabilities, and the results in Lemmas 5.1–5.3 will follow naturally under various dependence structures. Let

$$\varepsilon_{n1} = c_1 (\log p)^{1/2} / n^{\beta/2}$$

for some constant $c_1 > 0$ such that, by (3.2),

$$P(D_{n,1}^2 / n \le 1 - \varepsilon_{n1} n^{(\beta-1)/2}) = o(p^{-4}).$$

Put $\tilde{y}_n = y_n (1 - \varepsilon_{n1} n^{(\beta-1)/2}) \sim 2\sqrt{(\log p)/n}$. Using a similar argument as in the proof of Proposition 3.1 for estimating $P(A_{12} A_{13})$, we have

$$
\begin{aligned}
P\bigg(& \frac{|\sum_{k=1}^{n} z_{k1} z_{k2}|}{D_{n,1} D_{n,2}} > y_n, \frac{|\sum_{k=1}^{n} z_{k3} z_{k4}|}{D_{n,3} D_{n,4}} > y_n \bigg) \\
& \le P\bigg(\frac{|\sum_{k=1}^{n} z_{k1} z_{k2}|}{n} > \tilde{y}_n, \frac{|\sum_{k=1}^{n} z_{k3} z_{k4}|}{n} > \tilde{y}_n \bigg) + o(p^{-4}) \\
& \le P\bigg(\frac{|\sum_{k=1}^{n} (z_{k1} z_{k2} + z_{k3} z_{k4})|}{n^{1/2}} > 2 n^{1/2} \tilde{y}_n \bigg) \\
& \quad + P\bigg(\frac{|\sum_{k=1}^{n} (z_{k1} z_{k2} - z_{k3} z_{k4})|}{n^{1/2}} > 2 n^{1/2} \tilde{y}_n \bigg) + o(p^{-4}).
\end{aligned}
$$

(5.18)

Note that $\{z_{k1}z_{k2} + z_{k3}z_{k4}, 1 \leq k \leq n\}$ is a sequence of i.i.d. random variables with mean zero. $\square$

PROOF OF LEMMAS 5.1 AND 5.2. Under both assumptions on Σ_4, z_{14} is independent of (z_{11}, z_{12}, z_{13}), so that

$$\mathbb{E}(z_{11}z_{12} + z_{13}z_{14})^2 = \mathbb{E}(z_{11}z_{12})^2 + \mathbb{E}(z_{13}z_{14})^2 + 2\mathbb{E}[z_{11}z_{12}z_{13}z_{14}]$$
$$= 2 + 2\mathbb{E}[z_{11}z_{12}z_{13}] \cdot \mathbb{E}z_{14} = 2.$$

It follows from (3.2) that, for any $0 < \varepsilon < 1$,

$$P\left(\frac{|\sum_{k=1}^{n}(z_{k1}z_{k2} + z_{k3}z_{k4})|}{n^{1/2}} > 2n^{1/2}\tilde{y}_n\right) \leq 2\exp\{-(1 - \varepsilon/2)n\tilde{y}_n^2\} \leq 2p^{4(1-\varepsilon)}$$

for all sufficiently large n. The second probability in (5.18) can be estimated in exactly the same way, and hence the results of Lemmas 5.1 and 5.2 follow immediately. $\square$

PROOF OF LEMMA 5.3. In this case, (z_{11}, z_{13}) and (z_{12}, z_{14}) are independent. Then, for all $|r_1|, |r_2| \leq 1 - \delta$,

$$\mathbb{E}(z_{11}z_{12} + z_{13}z_{14})^2 = 2 + 2\mathbb{E}[z_{11}z_{13}] \cdot \mathbb{E}[z_{12}z_{14}] \leq 2 + 2(1 - \delta)^2.$$

Set $\varepsilon_\delta = (2\delta - \delta^2)/(4 - 2\delta + \delta^2)$. Applying (3.2) again, we have

$$P\left(\frac{|\sum_{k=1}^{n}(z_{k1}z_{k2} + z_{k3}z_{k4})|}{n^{1/2}} > 2n^{1/2}\tilde{y}_n\right)$$
$$\leq 2\exp\left\{-\frac{(1 - \varepsilon_\delta/2)n\tilde{y}_n^2}{1 + (1 - \delta)^2}\right\} \leq 2p^{-4(1-\varepsilon_\delta)/(1+(1-\delta)^2)}$$
$$= 2p^{-2(1+\varepsilon_\delta)}$$

for all sufficiently large n. This completes the proof. $\square$

REFERENCES

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

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210

CAI, T. T. and JIANG, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *J. Multivariate Anal.* **107** 24–39. MR2890430

CAI, T. T., WANG, L. and XU, G. (2010). Stable recovery of sparse signals and an oracle inequality. *IEEE Trans. Inform. Theory* **56** 3516–3522. MR2799010

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

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

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

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

JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215. MR2016616

LI, D., LIU, W.-D. and ROSALSKY, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148** 5–35. MR2653220

LI, D., QI, Y. and ROSALSKY, A. (2012). On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivariate Anal.* **111** 256–270. MR2944420

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

LINNIK, J. V. (1961). On the probability of large deviations for the sums of independent variables. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, *Vol. II* 289–306. Univ. California Press, Berkeley, CA. MR0137142

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

SHAO, Q.-M. (1997). Self-normalized large deviations. *Ann. Probab.* **25** 285–328. MR1428510

SHAO, Q. M. and ZHOU, W. X. (2012). Cramér type moderate deviation theorems for Studentized non-linear statistics. Unpublished manuscript.

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

DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
SHATIN, NT
HONG KONG
E-MAIL: qmshao@cuhk.edu.hk

DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE
    AND TECHNOLOGY
CLEAR WATER BAY
KOWLOON, HONG KONG
E-MAIL: zhouwenxin1986@gmail.com