Supplement to "Two-Sample Test of High Dimensional Means under Dependency"

Tony Cai[1], Weidong Liu[1,2] and Yin Xia[1]

Abstract

This supplement provides theoretical and numerical comparisons of the three oracle tests and presents more extensive simulation results comparing the numerical performance of the proposed test with that of other tests in the non-sparse cases as well as for non-Gaussian distributions. We also prove Propositions 1-3 and the technical results, Lemmas 3 and 4, which are used in the proofs of the main theorems.

1 Comparison of the Oracle Tests

In this section, we consider both the theoretical and numerical performance of the three oracle maximum-type tests. The results show that the test $\Phi_\alpha(\mathbf{\Omega})$ significantly outperforms the other two tests under the sparse alternatives in the oracle setting.

1.1 Theoretical Comparisons of the Three Oracle Tests

The test $\Phi_\alpha(\mathbf{\Omega})$ is shown in Section 3.2 to be minimax rate optimal for testing against sparse alternatives. We now compare the power of the test $\Phi_\alpha(\mathbf{\Omega})$ with that of $\Phi_\alpha(\mathbf{\Omega}^{\frac{1}{2}})$

[1]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, tcai@wharton.upenn.edu. The research was supported in part by NSF FRG Grant DMS-0854973.
[2]Department of Mathematics and Institute of Natural Sciences, Shanghai Jiao Tong University, China.

and $\Phi_\alpha(\boldsymbol{I})$ under the same alternative H_1 as in Section 3.2. Let

$$\mathcal{A} = \{1 \le i \le p : (\boldsymbol{\Omega}^{\frac{1}{2}})_{ij} = 0 \text{ for all } j \ne i\}.$$

That is, $i \in \mathcal{A}$ if and only if all the entries in the i-th row of $\boldsymbol{\Omega}^{\frac{1}{2}}$ are zero except for the diagonal entry.

Proposition 1 *(i) Suppose (C1)-(C3) hold. Then under H_1 with $r < 1/6$, we have*

$$\varliminf_{p\to\infty} \frac{P_{H_1}(\Phi_\alpha(\boldsymbol{\Omega}) = 1)}{P_{H_1}(\Phi_\alpha(\boldsymbol{I}) = 1)} \ge 1. \tag{1}$$

(ii) Suppose (C1)-(C3) hold. Assume there exists a constant $\epsilon_0 > 0$, such that for each $i \in \mathcal{A}^c$ at least one non-diagonal element in the i-th row of $\boldsymbol{\Omega}^{\frac{1}{2}}$ has a magnitude larger than ϵ_0. Then, under H_1 with $k_p = O(p^\tau)$ for all $0 < \tau < 1$, we have

$$\varliminf_{p\to\infty} \frac{P_{H_1}(\Phi_\alpha(\boldsymbol{\Omega}) = 1)}{P_{H_1}(\Phi_\alpha(\boldsymbol{\Omega}^{\frac{1}{2}}) = 1)} \ge 1. \tag{2}$$

The condition on $\boldsymbol{\Omega}^{\frac{1}{2}}$ is mild. In fact, by the definition of $\mathcal{A}$, there is at least one nonzero and non-diagonal element in each i-th row of $\boldsymbol{\Omega}^{\frac{1}{2}}$ with $i \in \mathcal{A}^c$. In Proposition 1, we assume that these nonzero and non-diagonal elements have magnitudes larger than ϵ_0.

Proposition 1 shows that, under some sparsity condition on $\boldsymbol{\delta}$, $\Phi_\alpha(\boldsymbol{\Omega})$ is uniformly at least as powerful as both $\Phi_\alpha(\boldsymbol{\Omega}^{\frac{1}{2}})$ and $\Phi_\alpha(\boldsymbol{I})$.

We now briefly discuss the different conditions on r in the theoretical results. For the maximum type test statistics, the range $r < 1/2$ is nearly optimal. Indeed, in the mean testing problem, the case $r > 1/2$ is treated as the dense setting and $r < 1/2$ as the sparse setting, similar to other sequence estimation problems. In the dense setting, the sum square type test statistics may outperform the maximum type test statistics under certain conditions. The different conditions imposed on r in this section are due to the technical arguments used in the proofs. We believe these ranges for r can be improved to $r < 1/2$ but the proof would be difficult. When the assumption on r does not hold, the tests are still valid but the comparison results may fail.

The test $\Phi_\alpha(\mathbf{\Omega})$ can be strictly more powerful than $\Phi_\alpha(\mathbf{\Omega}^{\frac{1}{2}})$ and $\Phi_\alpha(\boldsymbol{I})$. Assume that

$$H_1^{'}: \quad \boldsymbol{\delta} \text{ has } m = p^r,\ r < 1/2 \text{ nonzero coordinates with}$$

$$\frac{|\delta_i|}{\sqrt{\sigma_{i,i}}} = \sqrt{\frac{2\beta \log p}{n}}, \quad \text{where } \beta \in (0,1) \tag{3}$$

if $\delta_i \neq 0$. The nonzero locations $l_1 < l_2 < \cdots < l_m$ are randomly and uniformly drawn from $\{1, 2, \ldots, p\}$.

Proposition 2 *(i). Suppose that (C1) and (C2) hold. Then, under $H_1^{'}$ with $\beta \geq (1 - \sqrt{r})^2 + \varepsilon$ for some $\varepsilon > 0$, we have*

$$\lim_{p\to\infty} P_{H_1'}\Big(\Phi_\alpha(\boldsymbol{I}) = 1\Big) = 1.$$

If $\beta < (1 - \sqrt{r})^2$, then

$$\overline{\lim_{p\to\infty}}\, P_{H_1'}\Big(\Phi_\alpha(\boldsymbol{I}) = 1\Big) \leq \alpha.$$

(ii). Suppose that (C1) and (C3) hold and $r < 1/4$. Then, under $H_1^{'}$ with

$$\beta \geq (1 - \sqrt{r})^2 / (\min_{1\leq i\leq p} \sigma_{i,i}\omega_{i,i}) + \varepsilon \quad \textit{for some } \varepsilon > 0, \tag{4}$$

we have

$$\lim_{p\to\infty} P_{H_1'}\Big(\Phi_\alpha(\mathbf{\Omega}) = 1\Big) = 1.$$

The condition $r < 1/4$ can be weakened if we assume some stronger condition on $\mathbf{\Omega}$. In fact, based on the proof, we can see that it can be weakened to $r < 1/2$ if $\mathbf{\Omega}$ is s_p-sparse and $s_p = O(p^\tau)$, $\forall \tau > 0$.

Note that $\sigma_{i,i}\omega_{i,i} \geq 1$ for $1 \leq i \leq p$. When the variables are correlated, $\omega_{i,i}$ can be strictly larger than $1/\sigma_{i,i}$. For example, let $\mathbf{\Sigma} = (\phi^{|i-j|})$ with $|\phi| < 1$. Then $\min_{1\leq i\leq p} \sigma_{i,i}\omega_{i,i} \geq (1 - \phi^2)^{-1} > 1$. That is, $M_{\mathbf{\Omega}}$ is strictly more powerful than $M_{\boldsymbol{I}}$ under $H_1^{'}$.

In Proposition 2, the comparison between $\Phi_\alpha(\mathbf{\Omega})$ and $\Phi_\alpha(\boldsymbol{I})$ is restricted to H_1', which is special. However, the proof of Proposition 2 in fact implies the following more general result. Suppose that $\min_{1\le i\le p}\sigma_{i,i}\omega_{i,i} \ge 1+\varepsilon_1$ for some $\varepsilon_1 > 0$. Let β_0 and β_1 be any constants satisfying

$$\frac{(1-\sqrt{r})^2}{\min_{1\le i\le p}\sigma_{i,i}\omega_{i,i}} + \epsilon \le \beta_0 < \beta_1 < (1-\sqrt{r})^2$$

for some $\epsilon > 0$. Replacing (3) by $\sqrt{\frac{2\beta_0 \log p}{n}} \le \frac{|\delta_i|}{\sqrt{\sigma_{i,i}}} \le \sqrt{\frac{2\beta_1 \log p}{n}}$, we have

$$\lim_{p\to\infty} \mathsf{P}_{H_1'}\Big(\Phi_\alpha(\mathbf{\Omega}) = 1\Big) = 1,$$

and

$$\overline{\lim_{p\to\infty}}\, \mathsf{P}_{H_1'}\Big(\Phi_\alpha(\boldsymbol{I}) = 1\Big) \le \alpha.$$

We now turn to the comparison of the power of $\Phi_\alpha(\mathbf{\Omega})$ with that of $\Phi_\alpha(\mathbf{\Omega}^{\frac{1}{2}})$ under the alternative

$$\begin{aligned} H_1'' : \quad & \boldsymbol{\delta} \text{ has } m = p^r,\ r < 1/7 \text{ nonzero coordinates with} \\ & \max_j |a_{ji}\delta_i| = \sqrt{\frac{2\beta \log p}{n}}, \quad \text{where } \beta \in (0,1) \end{aligned} \tag{5}$$

if $\delta_i \ne 0$, where $\mathbf{\Omega}^{\frac{1}{2}} = (a_{ij})$. The nonzero locations $l_1 < l_2 < \cdots < l_m$ are randomly and uniformly drawn from $\{1, 2, \ldots, p\}$.

Proposition 3 *(i) Suppose (C1) holds. Then under H_1'' with $\beta < (1-\sqrt{r})^2$, we have*

$$\overline{\lim_{p\to\infty}}\, P_{H_1''}\Big(\Phi_\alpha(\mathbf{\Omega}^{\frac{1}{2}}) = 1\Big) \le \alpha.$$

(ii) Suppose that (C1) and (C3) hold. Then under H_1'' with

$$\beta \ge (1-\sqrt{r})^2/(\min_{1\le i\le p}(\omega_{i,i}/\max_j a_{ji}^2)) + \varepsilon \quad \textit{for some } \varepsilon > 0,$$

we have

$$\lim_{p\to\infty} P_{H_1''}\Big(\Phi_\alpha(\mathbf{\Omega}) = 1\Big) = 1.$$

It is easy to check that $\omega_{i,i}/(\max_j a_{ji}^2) \geq 1$ for all $1 \leq i \leq p$. When the variables are correlated, $\omega_{i,i}$ can be much larger than $\max_j a_{ji}^2$. For example, if for every row of $\mathbf{\Omega}^{\frac{1}{2}}$, there are at least 2 nonzero a_{ij}, then $\omega_{i,i} = \sum_{j=1}^{p} a_{ij}^2 > \max_j a_{ji}^2$. In this case, $M_{\mathbf{\Omega}}$ is strictly more powerful than $M_{\mathbf{\Omega}^{\frac{1}{2}}}$.

As the discussion below Proposition 2, the condition (5) can be generalized. Suppose that $\min_{1\leq i\leq p}(\omega_{i,i}/\max_j a_{ji}^2) > 1 + \varepsilon_1$ for some $\varepsilon_1 > 0$. Let β_0 and β_1 be any constants satisfying

$$\frac{(1-\sqrt{r})^2}{\min_{1\leq i\leq p}(\omega_{i,i}/\max_j a_{ji}^2)} + \epsilon \leq \beta_0 < \beta_1 < (1-\sqrt{r})^2$$

for some constant $\epsilon > 0$. If (5) is replaced by $\sqrt{\frac{2\beta_0 \log p}{n}} \leq \max_j |a_{ji}\delta_i| \leq \sqrt{\frac{2\beta_1 \log p}{n}}$, then

$$\lim_{p\to\infty} \mathsf{P}_{H_1''}\Big(\Phi_\alpha(\mathbf{\Omega}) = 1\Big) = 1$$

and

$$\overline{\lim_{p\to\infty}}\, \mathsf{P}_{H_1''}\Big(\Phi_\alpha(\mathbf{\Omega}^{\frac{1}{2}}) = 1\Big) \leq \alpha.$$

1.2 Numerical Comparisons of the Three Oracle Tests

We now make a numerical comparison of the three oracle tests $\Phi_\alpha(\mathbf{\Omega})$, $\Phi_\alpha(\mathbf{\Omega}^{\frac{1}{2}})$ and $\Phi_\alpha(\boldsymbol{I})$. Models 1 and 4 in Cai, Liu and Xia (2013) are considered. Besides, we also study the following two additional models.

- Model 9: $\mathbf{\Sigma}^* = (\sigma_{ij}^*)$ where $\sigma_{i,i}^* = 1$, $\sigma_{ij}^* = 0.5$ for $i \neq j$. $\mathbf{\Sigma} = \boldsymbol{D}^{1/2}\mathbf{\Sigma}^*\boldsymbol{D}^{1/2}$.
- Model 10: $\mathbf{\Sigma}^* = (\sigma_{ij}^*)$ where $\sigma_{i,i}^* = 1$, $\sigma_{ij}^* = \text{Unif}(0,1)$ for $i < j$ and $\sigma_{ji}^* = \sigma_{ij}^*$. $\mathbf{\Sigma} = \boldsymbol{D}^{1/2}(\mathbf{\Sigma}^* + \delta\boldsymbol{I})/(1+\delta)\boldsymbol{D}^{1/2}$ with $\delta = |\lambda_{\min}(\mathbf{\Sigma}^*)| + 0.05$.

We can see from Table 1 that the estimated sizes are reasonably close to the nominal level 0.05 for all three tests, and the test $\Phi_\alpha(\mathbf{\Omega})$ has the highest power in all four models over all dimensions ranging from 50 to 200 and outperforms both $\Phi_\alpha(\boldsymbol{I})$ and $\Phi_\alpha(\mathbf{\Omega}^{\frac{1}{2}})$.

p	50	100	200	50	100	200	50	100	200	50	100	200
	Model 1			Model 4			Model 9			Model 10		
	Size											
$\Phi_\alpha(\boldsymbol{I})$	0.06	0.05	0.04	0.03	0.04	0.04	0.03	0.02	0.02	0.03	0.04	0.04
$\Phi_\alpha(\boldsymbol{\Omega}^{\frac{1}{2}})$	0.05	0.05	0.03	0.04	0.04	0.04	0.03	0.04	0.04	0.03	0.04	0.04
$\Phi_\alpha(\boldsymbol{\Omega})$	0.05	0.04	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.03	0.05
	Power when $m = 0.05p$											
$\Phi_\alpha(\boldsymbol{I})$	0.11	0.22	0.42	0.08	0.10	0.29	0.03	0.10	0.08	0.04	0.18	0.37
$\Phi_\alpha(\boldsymbol{\Omega}^{\frac{1}{2}})$	0.23	0.63	0.87	0.16	0.20	0.64	0.03	0.36	0.26	0.06	0.44	0.79
$\Phi_\alpha(\boldsymbol{\Omega})$	0.32	0.73	0.94	0.22	0.29	0.81	0.04	0.37	0.26	0.05	0.60	0.92
	Power when $m = \sqrt{p}$											
$\Phi_\alpha(\boldsymbol{I})$	0.18	0.32	0.49	0.15	0.29	0.51	0.08	0.06	0.18	0.11	0.26	0.18
$\Phi_\alpha(\boldsymbol{\Omega}^{\frac{1}{2}})$	0.57	0.80	0.92	0.33	0.80	0.88	0.21	0.22	0.66	0.87	0.66	0.33
$\Phi_\alpha(\boldsymbol{\Omega})$	0.70	0.89	0.96	0.47	0.97	0.97	0.21	0.22	0.66	1.00	1.00	0.48

Table 1: Empirical sizes and powers for three oracle maximum-type tests for Model 1, 4, 9 and 10 with $\alpha = 0.05$ and $n = 100$. Based on 1000 replications.

2 Additional Simulation Results

In this section we present additional simulation results comparing the numerical performance of the proposed test with that of other tests. Non-Gaussian distributions are also considered. In addition, we compare the performance of the proposed test with the procedure using the covariance matrix estimators given in Ledoit and Wolf (JMVA, 2004) and Kubokawa and Srivastava (JMVA, 2008).

More extensive simulations are carried out for a range of non-sparse settings. Specifically, we consider non-sparse covariance structures by adding to the covariance/precision matrices in Models 1-5 a perturbation of a non-sparse matrix $\boldsymbol{E}$, where $\boldsymbol{E}$ is a symmetric matrix with 30% random nonzero entries drawn from Unif$(-0.2, 0.2)$. Furthermore, we carried out simulations for five additional general non-sparse covariance models. The comparisons were consistent with the cases reported in Cai, Liu and Xia (2013).

2.1 Non-Sparse Cases

We now consider additional non-sparse covariance models. We will first study Models $1'$-$6'$, where the covariance matrix or the precision matrix is a sparse matrix considered in Cai, Liu and Xia (2013) with a non-sparse perturbation, and then consider five more general non-sparse models, Models 11-15.

Let $\boldsymbol{E}$ be a symmetric matrix with the support of the off-diagonal entries chosen independently according to the Bernoulli(0.3) distribution with the values of the nonzero entries drawn randomly from Unif$(-0.2, 0.2)$. The following 6 models are considered, where each of them is a sparse matrix with a perturbation of the matrix $\boldsymbol{E}$. Thus all of these covariance/precision matrices are non-sparse.

- Model $1'$: $\boldsymbol{\Sigma}^* = (\sigma^*_{ij})$ where $\sigma^*_{i,i} = 1$, $\sigma^*_{ij} = 0.8$ for $2(k-1)+1 \leq i \neq j \leq 2k$, where $k = 1, ..., [p/2]$ and $\sigma^*_{ij} = 0$ otherwise. $\boldsymbol{\Omega}^* = \boldsymbol{\Sigma}^{*-1} + \boldsymbol{E}$. $\boldsymbol{\Omega} = \boldsymbol{\Omega}^* + \delta \boldsymbol{I}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^*)| + 0.05$.

- Model $2'$: $\boldsymbol{\Sigma}^* = (\sigma^*_{ij})$ where $\sigma^*_{ij} = 0.6^{|i-j|}$ for $1 \leq i, j \leq p$. $\boldsymbol{\Omega}^* = \boldsymbol{\Sigma}^{*-1} + \boldsymbol{E}$. $\boldsymbol{\Omega} = \boldsymbol{\Omega}^* + \delta \boldsymbol{I}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^*)| + 0.05$.

- Model $3'$: $\boldsymbol{\Omega}^* = (\omega_{ij})$ where $\omega^*_{i,i} = 2$ for $i = 1, ..., p$, $\omega^*_{ii+1} = 0.8$ for $i = 1, ..., p-1$, $\omega^*_{ii+2} = 0.4$ for $i = 1, ..., p-2$, $\omega^*_{ii+3} = 0.4$ for $i = 1, ..., p-3$, $\omega^*_{ii+4} = 0.2$ for $i = 1, ..., p-4$, $\omega^*_{ij} = \omega^*_{ji}$ for $i, j = 1, ..., p$ and $\omega^*_{ij} = 0$ otherwise. $\boldsymbol{\Omega} = \boldsymbol{\Omega}^* + \boldsymbol{E} + \delta \boldsymbol{I}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^*)| + 0.05$.

- Model $4'$: $\boldsymbol{\Sigma}^* = (\sigma^*_{ij})$ where $\sigma^*_{i,i} = 1$, $\sigma^*_{ij} = 0.8$ for $2(k-1)+1 \leq i \neq j \leq 2k$, where $k = 1, ..., [p/2]$ and $\sigma^*_{ij} = 0$ otherwise. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^*\boldsymbol{D}^{1/2} + \boldsymbol{E} + \delta \boldsymbol{I}$ with $\delta = |\lambda_{\min}(\boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^*\boldsymbol{D}^{1/2} + \boldsymbol{E})| + 0.05$.

- Model $5'$: $\boldsymbol{\Omega} = (\omega_{ij})$ where $\omega_{ij} = 0.6^{|i-j|}$ for $1 \leq i, j \leq p$. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Omega}^{-1}\boldsymbol{D}^{1/2} + \boldsymbol{E} + \delta \boldsymbol{I}$ with $\delta = |\lambda_{\min}(\boldsymbol{D}^{1/2}\boldsymbol{\Omega}^{-1}\boldsymbol{D}^{1/2} + \boldsymbol{E})| + 0.05$.

- Model $6'$: $\boldsymbol{\Omega}^{1/2} = (a_{ij})$ where $a_{i,i} = 1$, $a_{ij} = 0.8$ for $2(k-1)+1 \leq i \neq j \leq 2k$, where $k = 1, ..., [p/2]$ and $a_{ij} = 0$ otherwise. $\boldsymbol{\Omega} = \boldsymbol{D}^{1/2}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{1/2}\boldsymbol{D}^{1/2}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1} + \boldsymbol{E} + \delta \boldsymbol{I}$

with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^{-1} + \boldsymbol{E})| + 0.05$.

Simulations results for sizes and powers are reported in Table 2. For reasons of space, we only list the case when the magnitudes of the signals can vary for the power analysis. When the magnitude is fixed, the performance of the tests is similar. For those non-sparse models, similar phenomenon has been observed as the sparse cases. It can be seen from Table 2 that the new test $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ still outperforms all the sum of square type tests. Thus, our method is not restricted to the sparse cases.

p	50	100	200	50	100	200	50	100	200
	Model 1$'$			Model 2$'$			Model 3$'$		
T^2	0.05	0.05	–	0.06	0.05	–	0.06	0.05	–
BS	0.06	0.07	0.07	0.06	0.07	0.07	0.06	0.06	0.04
SD	0.06	0.07	0.07	0.06	0.07	0.07	0.06	0.06	0.04
CQ	0.06	0.07	0.08	0.06	0.07	0.07	0.06	0.06	0.04
$\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$	0.04	0.04	0.05	0.04	0.04	0.05	0.08	0.04	0.05
	Model 4$'$			Model 5$'$			Model 6$'$		
T^2	0.05	0.05	–	0.04	0.03	–	0.05	0.06	–
BS	0.07	0.07	0.05	0.05	0.05	0.07	0.06	0.07	0.06
SD	0.08	0.05	0.06	0.05	0.05	0.07	0.06	0.07	0.07
CQ	0.07	0.07	0.05	0.05	0.06	0.07	0.06	0.07	0.07
$\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$	0.05	0.05	0.05	0.06	0.07	0.07	0.03	0.04	0.05

Table 2: Empirical sizes based on 1000 replications with $\alpha = 0.05$ and $n = 100$.

We now study five more general non-sparse models. Let $\boldsymbol{D} = (d_{ij})$ be a diagonal matrix with diagonal elements $d_{i,i} = \text{Unif}(1, 3)$ for $i = 1, ..., p$. The following five models are studied where the magnitude of the signals can vary. Three different numbers of nonzero entries of the signal are considered: $0.05p$, $\sqrt{p}$ and $0.4p$. In the last case, the signal is also not sparse.

- Model 11: $\boldsymbol{\Sigma}^* = (\sigma^*_{ij})$ where $\sigma^*_{i,i} = 1$ and $\sigma^*_{ij} = |i - j|^{-5}/2$ otherwise. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^*\boldsymbol{D}^{1/2}$.

p	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
	$m = 0.05p$																	
	Model 1′			Model 2′			Model 3′			Model 4′			Model 5′			Model 6′		
T^2	0.13	0.41	–	0.14	0.30	–	0.18	0.26	–	0.14	0.37	–	0.18	0.30	–	0.13	0.26	–
BS	0.07	0.10	0.10	0.07	0.09	0.16	0.16	0.08	0.09	0.10	0.22	0.24	0.09	0.27	0.34	0.07	0.09	0.10
SD	0.06	0.12	0.11	0.07	0.10	0.17	0.16	0.09	0.10	0.10	0.21	0.26	0.11	0.29	0.43	0.07	0.09	0.10
CQ	0.07	0.09	0.10	0.07	0.09	0.16	0.16	0.08	0.09	0.10	0.22	0.23	0.09	0.26	0.34	0.07	0.08	0.10
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.06	0.16	0.21	0.07	0.20	0.37	0.35	0.13	0.25	0.16	0.40	0.46	0.23	0.53	0.68	0.18	0.40	0.23
	$m = \sqrt{p}$																	
	Model 1′			Model 2′			Model 3′			Model 4′			Model 5′			Model 6′		
T^2	0.84	0.81	–	0.54	0.51	–	0.25	0.66	–	0.47	0.23	–	0.24	0.42	–	0.67	0.84	–
BS	0.11	0.13	0.11	0.10	0.12	0.16	0.20	0.14	0.13	0.17	0.13	0.46	0.16	0.27	0.42	0.10	0.10	0.13
SD	0.15	0.17	0.13	0.11	0.13	0.18	0.21	0.16	0.16	0.14	0.14	0.59	0.16	0.30	0.46	0.10	0.12	0.14
CQ	0.11	0.13	0.12	0.10	0.12	0.16	0.20	0.14	0.13	0.16	0.13	0.46	0.16	0.27	0.42	0.10	0.10	0.13
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.17	0.27	0.20	0.19	0.22	0.38	0.38	0.35	0.47	0.24	0.26	0.77	0.20	0.43	0.61	0.44	0.66	0.53

Table 3: Powers of the tests based on 1000 replications with $\alpha = 0.05$ and $n = 100$.

- Model 12: $\boldsymbol{\Sigma}^* = (\sigma^*_{ij})$ where $\sigma^*_{i,i} = 1$, $\sigma^*_{ii+1} = \sigma^*_{i+1i} = 0.5$ and $\sigma^*_{ij} = 0.05$ otherwise. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^*\boldsymbol{D}^{1/2}$.

- Model 13: $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}(\boldsymbol{F}+2\boldsymbol{u}\boldsymbol{u}')\boldsymbol{D}^{1/2}$, where $\boldsymbol{F} = (f_{ij})$ is a $p\times p$ matrix with $f_{i,i} = 1$, $f_{ii+1} = f_{i+1i} = 0.5$ and $f_{ij} = 0$ otherwise, and $\boldsymbol{u}$ is a standardized vector.

- Model 14: $\boldsymbol{\Sigma}^* = (\sigma^*_{ij})$ where $\sigma^*_{i,i} = 1$, $\sigma^*_{ii+1} = \sigma^*_{i+1i} = 0.5$, $\sigma^*_{ii+2} = \sigma^*_{i+2i} = 0.4$ and $\sigma^*_{ij} = 0.05$ otherwise. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^*\boldsymbol{D}^{1/2}$.

- Model 15: $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}(\boldsymbol{F} + \boldsymbol{u}_1\boldsymbol{u}_1' + \boldsymbol{u}_2\boldsymbol{u}_2' + \boldsymbol{u}_3\boldsymbol{u}_3')\boldsymbol{D}^{1/2}$, where $\boldsymbol{F} = (f_{ij})$ is a $p \times p$ matrix with $f_{i,i} = 1$, $f_{ii+1} = f_{i+1i} = 0.5$ and $f_{ij} = 0$ otherwise, and $\boldsymbol{u}_i$ are orthogonal standardized vectors for $i = 1, 2, 3$.

For Model 11, the entries of $\boldsymbol{\Sigma}$ are decaying when they are further and further away from the diagonal. The corresponding $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are not sparse. For Model 12 and 14, all the entries of $\boldsymbol{\Sigma}$ have magnitude at least equal to 0.05 and thus both $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are not sparse. For Model 13 and 15, $\boldsymbol{\Sigma}$ is a sparse matrix plus one or three rank one non-sparse matrices and thus it is not sparse. The corresponding precision matrix $\boldsymbol{\Omega}$ is also not sparse. We can see from Table 4 that for those non-sparse models and non-sparse alternatives, similar phenomena as the sparse cases are observed. Our test still performs reasonably well and it outperforms all the sum of square type tests.

p	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
	Model 11			Model 12			Model 13			Model 14			Model 15		
	Size														
T^2	0.05	0.05	–	0.06	0.05	–	0.05	0.06	–	0.04	0.06	–	0.04	0.04	–
BS	0.07	0.06	0.05	0.05	0.06	0.06	0.08	0.05	0.06	0.05	0.07	0.06	0.06	0.06	0.06
SD	0.08	0.06	0.05	0.05	0.06	0.06	0.08	0.05	0.06	0.05	0.06	0.06	0.05	0.05	0.06
CQ	0.07	0.06	0.06	0.05	0.06	0.06	0.08	0.05	0.06	0.05	0.06	0.07	0.06	0.06	0.06
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.02	0.02	0.03	0.03	0.03	0.03	0.02	0.03	0.04	0.04	0.03	0.04	0.03	0.02	0.03
	Power when $m = 0.05p$														
T^2	0.17	0.43	–	0.32	0.33	–	0.23	0.99	–	0.06	0.16	–	0.31	0.40	–
BS	0.09	0.12	0.20	0.14	0.14	0.23	0.12	0.12	0.67	0.06	0.10	0.13	0.07	0.10	0.16
SD	0.09	0.13	0.20	0.14	0.14	0.23	0.12	0.12	0.68	0.05	0.09	0.14	0.07	0.09	0.16
CQ	0.09	0.12	0.20	0.14	0.13	0.22	0.12	0.12	0.67	0.05	0.10	0.13	0.07	0.11	0.16
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.14	0.41	0.80	0.26	0.50	0.64	0.25	0.82	1.00	0.04	0.11	0.25	0.20	0.50	0.84
	Power when $m = \sqrt{p}$														
T^2	0.49	0.59	–	0.87	0.96	–	0.99	1.00	–	0.46	0.56	–	0.53	1.00	–
BS	0.16	0.18	0.20	0.24	0.43	0.46	0.24	0.31	0.52	0.10	0.16	0.16	0.11	0.15	0.16
SD	0.16	0.17	0.20	0.24	0.43	0.46	0.25	0.31	0.52	0.11	0.17	0.17	0.12	0.14	0.16
CQ	0.15	0.18	0.20	0.24	0.42	0.45	0.24	0.30	0.51	0.10	0.16	0.16	0.11	0.14	0.16
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.37	0.57	0.53	0.75	0.88	0.98	0.73	0.96	1.00	0.34	0.43	0.27	0.38	0.73	0.85
	Power when $m = 0.4p$														
T^2	0.77	1.00	–	0.99	0.99	–	1.00	1.00	–	0.98	0.99	–	1.00	1.00	–
BS	0.24	0.68	0.99	0.32	0.78	0.99	0.86	0.99	1.00	0.44	0.75	0.99	0.35	0.67	0.93
SD	0.29	0.78	0.99	0.40	0.87	0.99	0.86	0.99	1.00	0.57	0.84	0.99	0.39	0.81	0.96
CQ	0.24	0.67	0.99	0.31	0.78	0.99	0.86	0.99	1.00	0.43	0.74	0.98	0.34	0.67	0.93
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.50	0.93	1.00	0.84	0.93	1.00	0.99	1.00	1.00	0.59	0.90	0.97	0.95	1.00	1.00

Table 4: Empirical sizes and powers for Model 11-15 with $\alpha = 0.05$ and $n = 100$. Based on 1000 replications.

2.2 Non-Gaussian Distributions

For Model 6, we generate two independent random samples $\{\boldsymbol{X}_k\}_{k=1}^{n_1}$ and $\{\boldsymbol{Y}_k\}_{k=1}^{n_2}$ from multivariate models $\boldsymbol{X}_k = \Gamma \boldsymbol{Z}_k^{(1)} + \boldsymbol{\mu}_1$ and $\boldsymbol{Y}_k = \Gamma \boldsymbol{Z}_k^{(2)} + \boldsymbol{\mu}_2$, with $\Gamma\Gamma' = \boldsymbol{\Sigma}$, where the components of $\boldsymbol{Z}_k^{(i)} = (Z_{k1}, ..., Z_{kp})'$ are i.i.d. standardized Gamma(10,1) random variables. We consider the case when $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ has $m = \sqrt{p}$ nonzero elements with the same signal strength for each coordinate. The results are summarized in Table 5.

p	50	100	200	50	100	200
	Empirical Sizes			Empirical Powers		
T^2	0.041	0.051	–	0.716	0.456	–
BS	0.055	0.069	0.062	0.153	0.183	0.192
SD	0.060	0.063	0.063	0.170	0.177	0.207
CQ	0.058	0.072	0.064	0.152	0.180	0.186
$\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$	0.050	0.039	0.048	0.875	0.722	0.597

Table 5: Sizes and Powers of tests based on 1000 replications with $\alpha = 0.05$ and $n = 100$ for Model 6. $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ has $m = \sqrt{p}$ nonzero elements. Signal strength keeps the same.

2.3 The Effects of Covariance Matrix Estimators

In this section we consider the effects of the covariance matrix estimators on the test by comparing the proposed procedure with the test using the covariance estimator given in Ledoit and Wolf (JMVA, 2004) and the Stein-type estimator in Kubokawa and Srivastava (JMVA, 2008). The size and power results for our test $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ and tests based on these two estimators (LW and KS in short) are shown in Table 6. It can be seen that LW either suffers from serious size distortions or has powers much lower than that of the proposed test, while KS always suffers from serious size distortions. We only list the case when the magnitudes of the signals vary and when dimension $p = 200$, because KS mainly focuses on the case when $p > n_1 + n_2 - 2$.

Models	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
	Size							
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.06	0.06	0.07	0.05	0.04	0.04	0.03	0.03
LW	0.06	0.09	0.07	0.06	0.96	0.11	0.24	0.20
KS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Power: $m = 0.05p$							
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.90	0.82	0.80	1.00	1.00	0.63	0.67	0.89
LW	0.28	0.48	0.60	0.65	1.00	0.37	0.46	0.44
KS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Power: $m = \sqrt{p}$							
$\Phi_\alpha(\widehat{\mathbf{\Omega}})$	0.97	0.93	0.89	0.99	1.00	0.64	0.59	0.94
LW	0.31	0.63	0.71	0.59	1.00	0.40	0.50	0.59
KS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 6: Size and power based on 1000 replications with $p = 200$ and $n = 100$.

3 Proof of Technical Lemmas

Lemma 1 (Bonferroni inequality) *Let $A = \cup_{t=1}^{p} A_t$. For any $k < [p/2]$, we have*

$$\sum_{t=1}^{2k} (-1)^{t-1} E_t \leq P(A) \leq \sum_{t=1}^{2k-1} (-1)^{t-1} E_t,$$

where $E_t = \sum_{1 \leq i_1 < \cdots < i_t \leq p} P(A_{i_1} \cap \cdots \cap A_{i_t})$.

Lemma 2 *[Berman (1962)] If X and Y have a bivariate normal distribution with expectation zero, unit variance and correlation coefficient ρ, then*

$$\lim_{c \to \infty} \frac{P\Big(X > c, Y > c\Big)}{[2\pi(1-\rho)^{\frac{1}{2}} c^2]^{-1} \exp\Big(-\frac{c^2}{1+\rho}\Big)(1+\rho)^{\frac{1}{2}}} = 1,$$

uniformly for all ρ such that $|\rho| \leq \delta$, for any δ, $0 < \delta < 1$.

Lemma 3 *Suppose (C1) holds. Then for p^r-sparse $\boldsymbol{\delta}$, with $r < 1/4$ and nonzero locations $l_1, ..., l_m$, $m = p^r$, randomly and uniformly drawn from $\{1, ..., p\}$, we have, for any $2r < a < 1 - 2r$, as $p \to \infty$,*

$$P\Big(\max_{i \in H} \Big| \frac{(\boldsymbol{\Omega\delta})_i}{\sqrt{\omega_{i,i}}} - \sqrt{\omega_{i,i}}\delta_i \Big| = O(p^{r-a/2}) \max_{i \in H} |\delta_i| \Big) \to 1, \tag{6}$$

and

$$P\Big(\max_{i \in H} \Big| (\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i - a_{i,i}\delta_i \Big| = O(p^{r-a/2}) \max_{i \in H} |\delta_i| \Big) \to 1, \tag{7}$$

where $\boldsymbol{\Omega}^{\frac{1}{2}} =: (a_{ij})$ and H is the support of $\boldsymbol{\delta}$.

Proof. We only need to prove (6) because the proof of (7) is similar. We re-order $\omega_{i1}, \ldots, \omega_{ip}$ as $|\omega_{i(1)}| \geq ... \geq |\omega_{i(p)}|$ for $i = 1, ..., p$. Let a satisfy $2r < a < 1 - 2r$ with $r < 1/4$. Define $\mathcal{I} = \{1 \leq i_1 < ... < i_m \leq p\}$ and

$$\mathcal{I}_0 = \Big\{1 \leq i_1 < ... < i_m \leq p : \text{ there exist some } 1 \leq k \leq m \text{ and some } j \neq k \text{ with } 1 \leq j \leq m, \text{ such that } |\omega_{i_k i_j}| \geq |\omega_{i_k(p^a)}|\Big\}.$$

We can show that

$$|\mathcal{I}_0| = O\Big(p \cdot p^a \binom{p}{p^r - 2}\Big) \text{ and } |\mathcal{I}| = \binom{p}{p^r}.$$

Therefore

$$|\mathcal{I}_0|/|\mathcal{I}| = O(p^{a+2r-1}) = o(1). \tag{8}$$

For $1 \leq t \leq m$, write

$$(\boldsymbol{\Omega\delta})_{l_t} = \sum_{k=1}^{p} \omega_{l_t k} \delta_k = \omega_{l_t l_t} \delta_{l_t} + \sum_{j=1, j \neq t}^{m} \omega_{l_t l_j} \delta_{l_j}.$$

Note that for every $(l_1, ..., l_m) \in \mathcal{I}_0^c$,

$$\sum_{j=1, j \neq t}^{m} |\omega_{l_t l_j}| \leq p^r \sqrt{\frac{C_0}{p^a}}.$$

It follows that for $H \in \mathcal{I}_0^c$ and $i \in H$,

$$\Big| \frac{(\boldsymbol{\Omega\delta})_i}{\sqrt{\omega_{i,i}}} - \sqrt{\omega_{i,i}} \delta_i \Big| = O(p^{r-a/2}) \max_{i \in H} |\delta_i|. \tag{9}$$

By (8) and (9), (6) is proved. ∎

Let $Y_1, \ldots, Y_n$ be independent normal random variables with $\mathsf{E}Y_i = \mu_i$ and the same variance $\mathsf{Var}(Y_i) = 1$.

Lemma 4 *Let $a_n = o((\log n)^{-1/2})$. We have*

$$\sup_{x \in \mathbb{R}} \max_{1 \leq k \leq n} \Big| P\Big(\max_{1 \leq i \leq k} Y_i \geq x + a_n\Big) - P\Big(\max_{1 \leq i \leq k} Y_i \geq x\Big) \Big| = o(1) \tag{10}$$

uniformly in the means μ_i, $1 \leq i \leq n$. If Y_i is replaced by $|Y_i|$, then (10) still holds.

Proof. We have

$$\begin{aligned} \mathsf{P}\Big(\max_{1 \leq i \leq k} Y_i \geq x + a_n\Big) &= 1 - \prod_{i=1}^{k} \Big(1 - \mathsf{P}(Y_i \geq x + a_n)\Big) \\ &= 1 - \exp\Big(\sum_{i=1}^{k} \log\Big(1 - \mathsf{P}(Y_i \geq x + a_n)\Big)\Big). \end{aligned}$$

Let $0 < \varepsilon < 1/2$ be any small number and M be any large number. Define

$$E = \{1 \le i \le k : \mathsf{P}(Y_i \ge x + a_n) \ge \varepsilon\}.$$

We first consider those n and k such that $\mathrm{Card}(E) \le \varepsilon^{-2}$ and

$$\sum_{i=1}^{k} \mathsf{P}(Y_i \ge x + a_n) \le M.$$

For $i \in E^c$, by the inequality $|\log(1-x) + x| \le x^2$ with $|x| < 1/2$, we have

$$\Big|\frac{\log\Big(1 - \mathsf{P}(Y_i \ge x + a_n)\Big)}{-\mathsf{P}(Y_i \ge x + a_n)} - 1\Big| \le \varepsilon. \tag{11}$$

Write $a_n = \varepsilon_n (\log n)^{-1/2}$, where $\varepsilon_n \to 0$. Let $b_n = |\varepsilon_n|^{-1/2} (\log n)^{\frac{1}{2}}$. We have for any large $\alpha > 0$,

$$\begin{aligned}
\mathsf{P}(Y_i \ge x + a_n) &= \int_{y \ge x - \mu_i + a_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&= \int_{y \ge x - \mu_i} \frac{1}{2\sqrt{\pi}} e^{-\frac{(y-a_n)^2}{2}} dy \\
&= \int_{y \ge x - \mu_i, |y| \le b_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-a_n)^2}{2}} dy + \int_{y \ge x - \mu_i, |y| > b_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-a_n)^2}{2}} dy \\
&= (1 + o(1)) \int_{y \ge x - \mu_i, |y| \le b_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + O(n^{-\alpha}) \\
&= (1 + o(1)) \mathsf{P}(Y_i \ge x) + O(n^{-\alpha}),
\end{aligned} \tag{12}$$

where $O(1)$ and $o(1)$ are uniformly in i and μ_i. Thus, we have

$$\begin{aligned}
&\sum_{i \in E^c} \log\Big(1 - \mathsf{P}(Y_i \ge x + a_n)\Big) \\
&\quad \le -(1 - 2\varepsilon) \sum_{i \in E^c} \mathsf{P}(Y_i \ge x) + O(n^{-\alpha+1}) \\
&\quad \le (1 - 2\varepsilon)(1 + 2\varepsilon)^{-1} \sum_{i \in E^c} \log\Big(1 - \mathsf{P}(Y_i \ge x)\Big) + O(n^{-\alpha+1}),
\end{aligned} \tag{13}$$

where in the last inequality we used (11) with $a_n = 0$. By (12),

$$\sum_{i \in E} \log\Big(1 - \mathsf{P}(Y_i \ge x + a_n)\Big) = \sum_{i \in E} \log\Big(1 - \mathsf{P}(Y_i \ge x)\Big) + o(1)\varepsilon^{-2}. \tag{14}$$

Combining (13) and (14), we have

$$\sum_{i=1}^{k} \log\Big(1-\mathsf{P}(Y_i \geq x+a_n)\Big) \leq \sum_{i=1}^{k} \log\Big(1-\mathsf{P}(Y_i \geq x)\Big) + 4\varepsilon M + o(1)\varepsilon^{-2}.$$

Hence

$$\mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x+a_n\Big) \geq \mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x\Big) - |e^{4\varepsilon M+o(1)\varepsilon^{-2}} - 1|. \tag{15}$$

Note that if $\sum_{i=1}^{k} \mathsf{P}(Y_i \geq x+a_n) > M$, then

$$\mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x+a_n\Big) \geq 1 - e^{-M}. \tag{16}$$

If $\mathrm{Card}(E) > \varepsilon^{-2}$, then

$$\mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x+a_n\Big) \geq 1-(1-\varepsilon)^{\varepsilon^{-2}} \geq 1-e^{-\varepsilon^{-1}}. \tag{17}$$

By (15)-(17), we have

$$\mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x+a_n\Big) \geq \mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x\Big) - |e^{4\varepsilon M+o(1)\varepsilon^{-2}} - 1| - e^{-\varepsilon^{-1}} - e^{-M}.$$

Similarly, we can prove

$$\mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x+a_n\Big) \leq \mathsf{P}\Big(\max_{1\leq i\leq k} Y_i \geq x\Big) + |e^{4\varepsilon M+o(1)\varepsilon^{-2}} - 1| + e^{-\varepsilon^{-1}} + e^{-M}.$$

By letting $n \to \infty$ first, following by $\varepsilon \to 0$ and then $M \to \infty$, the lemma is proved. ∎

The following lemma comes from Baraud(2002).

Lemma 5 *Let $\mathcal{F}$ be some subset of $l_2(J)$. Let μ_ρ be some probability measure on*

$$\mathcal{F}_\rho = \{\theta \in \mathcal{F}, \|\theta\| \geq \rho\}$$

and let

$$P_{\mu_\rho} = \int P_\theta d\mu_\rho(\theta).$$

Assuming that P_{μ_ρ} is absolutely continuous with respect to P_0, we define

$$L_{\mu_\rho}(y) = \frac{dP_{\mu_\rho}}{dP_0}(y).$$

For all $\alpha > 0$, $\nu \in [0, 1-\alpha]$, *if*

$$\mathsf{E}_0(L^2_{\mu_{\rho^*}}(Y)) \leq 1 + 4(1-\alpha-\nu)^2,$$

then

$$\forall \rho \leq \rho^*, \quad \inf_{\phi_\alpha} \sup_{\theta \in \mathcal{F}_\rho} \mathsf{P}_\theta(\Phi_\alpha = 0) \geq \nu.$$

Let $(Z_1, \ldots, Z_p)'$ be a zero mean multivariate normal random vector with covariance matrix $\mathbf{\Omega} = (\omega_{ij})_{1 \leq i,j \leq p}$ and the diagonal $\omega_{i,i} = 1$ for $1 \leq i \leq p$.

Lemma 6 *Suppose that* $\max_{1 \leq i \neq j \leq p} |\omega_{ij}| \leq r < 1$ *and* $\lambda_{\max}(\mathbf{\Omega}) \leq C_0$. *We have*

$$\mathsf{P}\Big(\max_{1 \leq i \leq p} Z_i^2 - 2\log p + \log\log p \leq x\Big) \to \exp\Big(-\frac{1}{\sqrt{\pi}}\exp(-x/2)\Big), \tag{18}$$

and

$$\mathsf{P}\Big(\max_{1 \leq i \leq p} Z_i \leq \sqrt{2\log p - \log\log p + x}\Big) \to \exp\Big(-\frac{1}{2\sqrt{\pi}}\exp(-x/2)\Big), \tag{19}$$

for any $x \in \mathbb{R}$ *as* $p \to \infty$.

4 Proof of Propositions

Proof of Proposition 1 (i). Let $\boldsymbol{U}$ be a multivariate normal random vector with zero mean and covariance matrix $\mathbf{\Sigma}$. Let $\boldsymbol{Z} = \boldsymbol{\delta} + \boldsymbol{U}$, where $\boldsymbol{\delta}$ and $\boldsymbol{U}$ are independent. Without loss of generality, we assume that $\sigma_{i,i} = 1$ for $1 \leq i \leq p$. Then $\omega_{i,i} \geq 1$ for $1 \leq i \leq p$. Set $\boldsymbol{A} = \{\max_{1 \leq i \leq p} |\delta_i| \leq 6\sqrt{\log p}\}$. By Lemma 3, we have

$$\mathsf{P}\Big(\max_{1 \leq i \leq p} |(\mathbf{\Omega}\boldsymbol{\delta})_i / \sqrt{\omega_{i,i}}| \geq (1 - o(1)) \max_{1 \leq i \leq p} |\delta_i|\Big) \to 1. \tag{20}$$

Thus by Theorem 1 we have

$$\mathsf{P}\Big(M_{\mathbf{\Omega}} \in R_\alpha, \boldsymbol{A}^c\Big) \;\geq\; \mathsf{P}\Big(\max_{1 \leq i \leq p} |(\mathbf{\Omega}\boldsymbol{U})_i / \sqrt{\omega_{i,i}}| - 4\sqrt{\log p} \geq \sqrt{2\log p}, \boldsymbol{A}^c\Big) + o(1)$$

$$= \quad \mathsf{P}(\boldsymbol{A}^c) + o(1). \tag{21}$$

Similarly, we have

$$\mathsf{P}\Big(M_{\boldsymbol{I}} \in R_\alpha, \boldsymbol{A}^c\Big) = \mathsf{P}(\boldsymbol{A}^c) + o(1). \tag{22}$$

We next consider $\mathsf{P}\Big(M_{\boldsymbol{\Omega}} \in R_\alpha, \boldsymbol{A}\Big)$ and $\mathsf{P}\Big(M_{\boldsymbol{I}} \in R_\alpha, \boldsymbol{A}\Big)$. For notation briefness, we denote $\mathsf{P}(\boldsymbol{BA}|\boldsymbol{\delta})$ and $\mathsf{P}(\boldsymbol{B}|\boldsymbol{\delta})$ by $\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\boldsymbol{B})$ and $\mathsf{P}_{\boldsymbol{\delta}}(\boldsymbol{B})$ respectively for any event $\boldsymbol{B}$. Let $H = \text{supp}(\boldsymbol{\delta}) = \{l_1, ..., l_m\}$ with $m = p^r$ and $H^c = \{1, ..., p\} \setminus H$. We have

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(M_{\boldsymbol{I}} \in R_\alpha) = \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H}|Z_i| \geq \sqrt{x_p}) + \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H}|Z_i| < \sqrt{x_p}, \max_{j\in H^c}|Z_j| \geq \sqrt{x_p}), \tag{23}$$

where $x_p = 2\log p - \log\log p + x$. Define

$$H_1^c = \{j \in H^c : |\sigma_{ij}| \leq p^{-\xi} \text{ for any } i \in H\}, \quad H_1 = H^c - H_1^c$$

for $2r < \xi < (1-r)/2$. It is easy to see that $\text{Card}(H_1) \leq Cp^{r+2\xi}$. It follows that

$$\mathsf{P}\Big(\max_{j\in H_1}|Z_j| \geq \sqrt{x_p}\Big) \leq p^{r+2\xi}\mathsf{P}\Big(|N(0,1)| \geq \sqrt{x_p}\Big) = O(p^{r+2\xi-1}) = o(1). \tag{24}$$

We claim that

$$\begin{aligned} &\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\max_{i\in H}|Z_i| < \sqrt{x_p}, \max_{j\in H_1^c}|Z_j| \geq \sqrt{x_p}\Big) \\ &\quad \leq \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big((\max_{i\in H}|Z_i| < \sqrt{x_p}\Big)\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\max_{j\in H_1^c}|Z_j| \geq \sqrt{x_p}\Big) + o(1). \end{aligned} \tag{25}$$

Throughout the proof, $O(1)$ and $o(1)$ are uniformly for $\boldsymbol{\delta}$. To prove (25), we set $\boldsymbol{E} = \{\max_{i\in H}|Z_i| < \sqrt{x_p}\}$, $\boldsymbol{F}_j = \{|Z_j| \geq \sqrt{x_p}\}$, $j \in H_1^c$. Then by Bonferroni inequality, we have for any fixed integer $k > 0$,

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\bigcup_{j\in H_1^c}\{\boldsymbol{E}\cap\boldsymbol{F}_j\}\Big) \leq \sum_{t=1}^{2k-1}(-1)^{t-1}\sum_{i_1<\cdots<i_t\in H_1^c}\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\boldsymbol{E}\cap\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t}\Big). \tag{26}$$

Let $\boldsymbol{W} = (w_{ij})$ be the covariance matrix of the vector $(Z_i, i \in H, Z_{i_1}, \ldots, Z_{i_t})'$ given $\boldsymbol{\delta}$. Note that $\boldsymbol{W}$ satisfies $|w_{ij}| \leq p^{-\xi}$ for $i \in H$ and $j = i_1, \ldots, i_t \in H_1^c$. Define the matrix $\tilde{\boldsymbol{W}} = (\tilde{w}_{ij})$ with $\tilde{w}_{ij} = w_{ij}$ for $i, j \in H$, $\tilde{w}_{ij} = w_{ij}$ for $i, j = i_1, \ldots, i_t \in H_1^c$ and $\tilde{w}_{ij} = \tilde{w}_{ji} = 0$ for $i \in H$ and $j = i_1, \ldots, i_t \in H_1^c$. Set $\boldsymbol{z} = (u_i, i \in H, z_{i_1}, \ldots, z_{i_t})'$ and

$$\mathcal{R} = \{|u_i + \delta_i| \leq \sqrt{x_p}, i \in H, |z_{i_1}| \geq \sqrt{x_p}, \ldots, |z_{i_t}| \geq \sqrt{x_p}\},$$

$$\mathcal{R}_1 = \mathcal{R} \cap \{|\boldsymbol{z}|_\infty \le 8\sqrt{t \log p}\},$$
$$\mathcal{R}_2 = \mathcal{R} \cap \{|\boldsymbol{z}|_\infty > 8\sqrt{t \log p}\}.$$

We have

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\boldsymbol{E} \cap \boldsymbol{F}_{i_1} \cap \cdots \cap \boldsymbol{F}_{i_t}\Big) = \frac{I\{\boldsymbol{A}\}}{(2\pi)^{p^r+t}|\boldsymbol{W}|^{\frac{1}{2}}} \int_{\mathcal{R}} \exp\Big(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}\Big)d\boldsymbol{z}. \tag{27}$$

By (C1) we have $C_0^{-1} \le \lambda_{\min}(\tilde{\boldsymbol{W}}) \le \lambda_{\max}(\tilde{\boldsymbol{W}}) \le C_0$. Note that $\|\boldsymbol{W} - \tilde{\boldsymbol{W}}\|_2 = O(p^{r-\xi})$ and $|\boldsymbol{W}| = (1 + O(p^{r-\xi}))^{p^r+t}|\tilde{\boldsymbol{W}}| = (1 + O(p^{2r-\xi}))|\tilde{\boldsymbol{W}}|$. This implies that

$$\begin{aligned} &\frac{1}{(2\pi)^{p^r+t}|\boldsymbol{W}|^{\frac{1}{2}}} \int_{\mathcal{R}_1} \exp\Big(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}\Big)d\boldsymbol{z} \\ &\quad = (1 + O(p^{2r-\xi}\log p))\frac{1}{(2\pi)^{p^r+t}|\tilde{\boldsymbol{W}}|^{\frac{1}{2}}} \int_{\mathcal{R}_1} \exp\Big(-\frac{1}{2}\boldsymbol{z}'\tilde{\boldsymbol{W}}^{-1}\boldsymbol{z}\Big)d\boldsymbol{z}. \end{aligned} \tag{28}$$

Furthermore, it is easy to see that

$$\begin{aligned} \frac{1}{(2\pi)^{p^r+t}|\boldsymbol{W}|^{\frac{1}{2}}} \int_{\mathcal{R}_2} \exp\Big(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{W}^{-1}\boldsymbol{z}\Big)d\boldsymbol{z} = O(p^{-32t}), \\ \frac{1}{(2\pi)^{p^r+t}|\tilde{\boldsymbol{W}}|^{\frac{1}{2}}} \int_{\mathcal{R}_2} \exp\Big(-\frac{1}{2}\boldsymbol{z}'\tilde{\boldsymbol{W}}^{-1}\boldsymbol{z}\Big)d\boldsymbol{z} = O(p^{-32t}). \end{aligned} \tag{29}$$

Thus, it follows from (27)-(29) that

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\boldsymbol{E} \cap \boldsymbol{F}_{i_1} \cap \cdots \cap \boldsymbol{F}_{i_t}\Big) = (1 + O(p^{2r-\xi}\log p))\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\boldsymbol{E})\mathsf{P}_{\boldsymbol{\delta}}\Big(\boldsymbol{F}_{i_1} \cap \cdots \cap \boldsymbol{F}_{i_t}\Big) + O(p^{-32t}).$$

As the proof of Lemma 6, we can show that

$$\sum_{i_1 < \cdots < i_t \in H_1^c} \mathsf{P}_{\boldsymbol{\delta}}\Big(\boldsymbol{F}_{i_1} \cap \cdots \cap \boldsymbol{F}_{i_t}\Big) = (1 + o(1))\pi^{-\frac{t}{2}}\frac{1}{t!}\exp\Big(-\frac{tq_\alpha}{2}\Big).$$

It follows from (26) that

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\bigcup_{j \in H_1^c}\{\boldsymbol{E} \cap \boldsymbol{F}_j\}\Big) \le \alpha\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\boldsymbol{E}) + o(1).$$

This, together with (23) and (24), implies that

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(M_{\boldsymbol{I}} \in R_\alpha) \le \alpha I\{\boldsymbol{A}\} + (1 - \alpha)\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\boldsymbol{E}^c) + o(1),$$

where $o(1)$ is uniformly for $\boldsymbol{\delta}$. Hence, we have

$$\mathsf{P}(M_{\boldsymbol{I}} \in R_\alpha, \boldsymbol{A}) \leq \alpha \mathsf{P}(\boldsymbol{A}) + (1-\alpha)\mathsf{P}(\boldsymbol{E}^c, \boldsymbol{A}) + o(1)$$

and

$$\mathsf{P}(M_{\boldsymbol{I}} \in R_\alpha) \leq \alpha \mathsf{P}(\boldsymbol{A}) + \mathsf{P}(\boldsymbol{A}^c) + (1-\alpha)\mathsf{P}(\boldsymbol{E}^c, \boldsymbol{A}) + o(1). \tag{30}$$

We next prove that

$$\mathsf{P}(M_{\boldsymbol{\Omega}} \in R_\alpha, \boldsymbol{A}) \geq \alpha \mathsf{P}(\boldsymbol{A}) + (1-\alpha)\mathsf{P}(\tilde{\boldsymbol{E}}^c, \boldsymbol{A}) + o(1), \tag{31}$$

and hence

$$\mathsf{P}(M_{\boldsymbol{\Omega}} \in R_\alpha) \geq \alpha \mathsf{P}(\boldsymbol{A}) + \mathsf{P}(\boldsymbol{A}^c) + (1-\alpha)\mathsf{P}(\tilde{\boldsymbol{E}}^c, \boldsymbol{A}) + o(1), \tag{32}$$

where $\tilde{\boldsymbol{E}} = \{\max_{i \in H} |Z_i^o| < \sqrt{x_p}\}$, $\boldsymbol{Z}^o = (Z_1^o, \ldots, Z_p^o)'$, and $Z_i^o = \frac{(\boldsymbol{\Omega Z})_i}{\sqrt{\omega_{i,i}}}$. It suffices to show that

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(M_{\boldsymbol{\Omega}} \in R_\alpha) \geq \alpha I\{\boldsymbol{A}\} + (1-\alpha)\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\tilde{\boldsymbol{E}}^c) + o(1). \tag{33}$$

Define $\tilde{H}_1^c = \{j \in H^c : |\omega_{ij}| \leq p^{-\xi} \text{ for any } i \in H\}$ for $2r < \xi < (1-r)/2$. It is easy to see that $\text{Card}(\tilde{H}_1^c) \geq p - O(p^{r+2\xi})$. Then

$$\begin{aligned}
\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(M_{\boldsymbol{\Omega}} \in R_\alpha) &= \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i \in H} |Z_i^o| \geq \sqrt{x_p}) + \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i \in H} |Z_i^o| < \sqrt{x_p}, \max_{j \in H^c} |Z_j^o| \geq \sqrt{x_p}) \\
&\geq \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i \in H} |Z_i^o| \geq \sqrt{x_p}) + \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i \in H} |Z_i^o| < \sqrt{x_p}, \max_{j \in \tilde{H}_1^c} |Z_j^o| \geq \sqrt{x_p}).
\end{aligned}$$

Note that on $\boldsymbol{A}$, $\max_{j \in \tilde{H}_1^c} |(\boldsymbol{\Omega\delta})_j| = \max_{j \in \tilde{H}_1^c} \left|\sum_{i \in H} \omega_{ji}\delta_i\right| \leq 4p^{r-\xi}\sqrt{\log p}$. It follows from the same arguments as above and using the left hand side of Bonferroni inequality that

$$\begin{aligned}
&\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i \in H} |Z_i^o| < \sqrt{x_p}, \max_{j \in \tilde{H}_1^c} |Z_j^o| \geq \sqrt{x_p}) \\
&\quad \geq \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\max_{i \in H} |Z_i^o| < \sqrt{x_p}, \max_{j \in \tilde{H}_1^c} |Z_j^o - (\boldsymbol{\Omega\delta})_j/\sqrt{\omega_{jj}}| \geq \sqrt{x_p} + Cp^{r-\xi}\sqrt{\log p}\Big) \\
&\quad \geq \alpha \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\tilde{\boldsymbol{E}}) + o(1)
\end{aligned}$$

Hence, (33) is proved.

We next compare $\mathsf{P}(\tilde{\boldsymbol{E}}^c, \boldsymbol{A})$ with $\mathsf{P}(\boldsymbol{E}^c, \boldsymbol{A})$. Without loss of generality, we assume that for any $i \in H$, $\delta_i > 0$. By Lemma 3 we also can assume that, on the event $\boldsymbol{A}$, $\frac{(\boldsymbol{\Omega\delta})_i}{\sqrt{\omega_{i,i}}} := \delta_i^o \geq \delta_i - O(p^{r-a/2})$ for some $2r < a < 1-2r$. Note that

$$\Big|\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H}|Z_i| \geq \sqrt{x_p}) - \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H} Z_i \geq \sqrt{x_p})\Big| \leq \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\min_{i\in H} Z_i \leq -\sqrt{x_p}) = o(1)$$

and

$$\Big|\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H}|Z_i^o| \geq \sqrt{x_p}) - \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H} Z_i^o \geq \sqrt{x_p})\Big| \leq \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\min_{i\in H} Z_i^o \leq -\sqrt{x_p}) = o(1).$$

It suffices to show that

$$\mathsf{P}(\max_{i\in H} Z_i \geq \sqrt{x_p}, \boldsymbol{A}) \leq \mathsf{P}(\max_{i\in H} Z_i^o \geq \sqrt{x_p}, \boldsymbol{A}) + o(1). \tag{34}$$

Let $\mathcal{I}_0 = \{(i_1, ..., i_m) : \ \exists 1 \leq l < j \leq m, \text{ such that } |\sigma_{i_l,i_j}| \geq p^{-\xi}\}$ and let $\mathcal{I} = \{(i_1, ..., i_m) : 1 \leq i_1 < \cdots < i_m \leq p\}$. We can show that

$$|\mathcal{I}_0| \leq O\Big(p \cdot p^{2\xi}\binom{p}{p^r - 2}\Big).$$

By some simple calculations, for $\xi < \frac{1}{2}(1-2r)$, we have $|\mathcal{I}_0|/|\mathcal{I}| = o(1)$. Thus, $\mathsf{P}(\boldsymbol{\delta} \in \mathcal{I}_0) = o(1)$. For $\boldsymbol{\delta} \in \mathcal{I}_0^c$ with $2r < \xi < \frac{1}{2}(1-2r)$, using the same arguments from (27) to (29), we obtain that

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H} Z_i \geq \sqrt{x_p}) = I\{\boldsymbol{A}\} - I\{\boldsymbol{A}\}\prod_{i\in H}\Big(1 - \mathsf{P}_{\boldsymbol{\delta}}(Z_i \geq \sqrt{x_p})\Big) + o(1).$$

Similarly, let $\mathcal{I}_1 = \{(i_1, ..., i_m) : \ \exists 1 \leq l < j \leq m, \text{ such that } |\omega_{i_l,i_j}| \geq p^{-\xi}\}$, then we can get $|\mathcal{I}_1|/|\mathcal{I}| = o(1)$, and for $\boldsymbol{\delta} \in \mathcal{I}_1^c$,

$$\begin{aligned}
&\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H} Z_i^o \geq \sqrt{x_p}) \\
&\quad = I\{\boldsymbol{A}\} - I\{\boldsymbol{A}\}\prod_{i\in H}\Big(1 - \mathsf{P}_{\boldsymbol{\delta}}(Z_i^o \geq \sqrt{x_p})\Big) + o(1) \\
&\quad \geq I\{\boldsymbol{A}\} - I\{\boldsymbol{A}\}\prod_{i\in H}\Big(1 - \mathsf{P}_{\boldsymbol{\delta}}(Z_i \geq \sqrt{x_p} + O(p^{r-a/2}))\Big) + o(1),
\end{aligned} \tag{35}$$

for any a satisfying $2r < a < 1-2r$. By Lemma 4, we have for $\boldsymbol{\delta} \in \mathcal{I}_0^c \cap \mathcal{I}_1^c$,

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H} Z_i^o \geq \sqrt{x_p}) \geq I\{\boldsymbol{A}\} - I\{\boldsymbol{A}\}\prod_{i\in H}\Big(1 - \mathsf{P}_{\boldsymbol{\delta}}(Z_i \geq \sqrt{x_p})\Big) + o(1),$$

which, together with the fact $\mathsf{P}(\boldsymbol{\delta} \in \mathcal{I}_0) = o(1)$ and $\mathsf{P}(\boldsymbol{\delta} \in \mathcal{I}_1) = o(1)$, proves (34). Proposition 1 (i) is proved by (30),(32) and (34). ∎

Proof of Proposition 1 (ii). Define $M_{\mathbf{\Omega}}^{'} = \max_{i\in\mathcal{A}^c} |(\mathbf{\Omega}\boldsymbol{Z})_i/\sqrt{\omega_{i,i}}|$, $M_{\mathbf{\Omega}}^{''} = \max_{i\in\mathcal{A}} |(\mathbf{\Omega}\boldsymbol{Z})_i/\sqrt{\omega_{i,i}}|$, $M_{\mathbf{\Omega}^{\frac{1}{2}}}^{'} = \max_{i\in\mathcal{A}^c} |(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{Z})_i|$, and $M_{\mathbf{\Omega}^{\frac{1}{2}}}^{''} = \max_{i\in\mathcal{A}} |(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{Z})_i|$. By the definition of $\mathcal{A}$, we see that $M_{\mathbf{\Omega}}^{'}$ and $M_{\mathbf{\Omega}}^{''}$ are independent. Hence we have

$$\begin{aligned}\mathsf{P}\Big(M_{\mathbf{\Omega}} \geq \sqrt{x_p}\Big) &= \mathsf{P}\Big(M_{\mathbf{\Omega}}^{''} \geq \sqrt{x_p}\Big) + \mathsf{P}\Big(M_{\mathbf{\Omega}}^{''} < \sqrt{x_p}\Big)\mathsf{P}\Big(M_{\mathbf{\Omega}}^{'} \geq \sqrt{x_p}\Big) \\ &= \mathsf{P}\Big(M_{\mathbf{\Omega}^{\frac{1}{2}}}^{''} \geq \sqrt{x_p}\Big) + \mathsf{P}\Big(M_{\mathbf{\Omega}^{\frac{1}{2}}}^{''} < \sqrt{x_p}\Big)\mathsf{P}\Big(M_{\mathbf{\Omega}}^{'} \geq \sqrt{x_p}\Big).\end{aligned}$$

We next prove that

$$\mathsf{P}\Big(M_{\mathbf{\Omega}}^{'} \geq \sqrt{x_p}\Big) \geq \mathsf{P}\Big(M_{\mathbf{\Omega}^{\frac{1}{2}}}^{'} \geq \sqrt{x_p}\Big) + o(1). \tag{36}$$

From the proof of Proposition 1(i), we can assume that $\max_{1\leq i\leq p} |\delta_i| \leq 6\sqrt{\log p}$. Set $\boldsymbol{A} = \{\max_{i\in\mathcal{A}^c}(\max_j |a_{ij}\delta_i|) < \sqrt{2\beta_0 \log p}\}$ for some $\beta_0 < 1$ being sufficiently close to 1. Because $\omega_{i,i} = \sum_{j=1}^p a_{ij}^2$ and $\sum_{i=1}^p a_{ij}^2 \geq \max_j a_{ij}^2 + \epsilon_1^2$ for $i \in \mathcal{A}^c$ with some $\epsilon_1 > 0$, we have by Lemma 3,

$$\mathsf{P}\Big(\max_{i\in\mathcal{A}^c}\Big|\frac{(\mathbf{\Omega}\boldsymbol{\delta})_i}{\sqrt{\omega_{i,i}}}\Big| \;\geq\; (1+2\epsilon_2)\max_{i\in\mathcal{A}^c}(\max_j |a_{ij}\delta_i|) + o(1)\Big) \to 1,$$

for some constant $\epsilon_2 > 0$. Thus we have

$$\begin{aligned}\mathsf{P}\Big(M_{\mathbf{\Omega}}^{'} \in R_\alpha, \boldsymbol{A}^c\Big) &\geq \mathsf{P}\Big(|Z_{i_0}^o| \geq \sqrt{x_p}, \boldsymbol{A}^c\Big) \\ &\geq \mathsf{P}\Big(Z_{i_0}^o - \delta_i^o \geq \sqrt{x_p} - (1+\epsilon_2)\sqrt{2\beta_0 \log p}, \boldsymbol{A}^c\Big) - o(1) \\ &= \mathsf{P}(\boldsymbol{A}^c) - o(1),\end{aligned} \tag{37}$$

where $i_0 = \arg\max_{i\in\mathcal{A}^c}(\max_j |a_{ij}\delta_i|)$. We next consider $\mathsf{P}\Big(M_{\mathbf{\Omega}}^{'} \in R_\alpha, \boldsymbol{A}\Big)$ and $\mathsf{P}\Big(M_{\mathbf{\Omega}^{\frac{1}{2}}}^{'} \in R_\alpha, \boldsymbol{A}\Big)$. Let $Z_i^* = (\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{Z})_i$ and $\delta_i^* = (\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i$. Then

$$\begin{aligned}\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(M_{\mathbf{\Omega}^{\frac{1}{2}}}^{'} \in R_\alpha) &= \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H\cap\mathcal{A}^c} |Z_i^*| \geq \sqrt{x_p}) + \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H\cap\mathcal{A}^c} |Z_i^*| < \sqrt{x_p}, \max_{j\in H^c\cap\mathcal{A}^c} |Z_j^*| \geq \sqrt{x_p}) \\ &= \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H\cap\mathcal{A}^c} |Z_i^*| \geq \sqrt{x_p}) + \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H\cap\mathcal{A}^c} |Z_i^*| < \sqrt{x_p})\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{j\in H^c\cap\mathcal{A}^c} |Z_j^*| \geq \sqrt{x_p}).\end{aligned}$$

Let $H=\{i_1,\dots,i_m\}$ be uniformly drawn from $\{1,2,\dots,p\}$. We divide H^c into two sets H_1 and H_2, where

$$H_1=\cup_{k=1}^m\{1\le j\le p:|a_{i_kj}|>|a_{i_k(p^{4r})}|,\ j\ne i_k\}=:\cup_{k=1}^m H_{1k},\quad H_2=H^c\setminus H_1.$$

It is easy to see that Card$(H_1)\le p^{5r}$. We next show that $|(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i|\le\sqrt{2\beta_0\log p}+o(1)$ uniformly for $i\in H_1$ with probability tending to one. As in the proof of Lemma 3, we modify the definition of $\mathcal{I}_0$ to

$$\begin{aligned}\mathcal{I}_0 &= \Big\{1\le i_1<...<i_m\le p:\ \text{there exist some } i_k \text{ and } i_j\ne i_k\\ &\qquad\qquad \text{such that } |a_{i,i_j}|>|a_{i,(p^{4r})}| \text{ for some } i\in H_{1k}\Big\}.\end{aligned}$$

It's easy to show that Card$(\mathcal{I}_0)\le p^{1+9r}C_p^{m-2}=o(1)C_p^m$ since $m=p^r$ and r is arbitrarily small. Note that, for $i\in H_{1k}$ and $(i_1,\dots,i_m)\in\mathcal{I}_0^c$,

$$(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i=\sum_{j=1}^m a_{i,i_j}\delta_{i_j}=a_{i,i_k}\delta_{i_k}+\sum_{j\ne k}^m a_{i,i_j}\delta_{i_j}=a_{i,i_k}\delta_{i_k}+O(p^{-r}\sqrt{\log p}).$$

This implies that on $\boldsymbol{A}$, $\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\max_{i\in H_1\cap\mathcal{A}^c}|(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i|\le\sqrt{2\beta_0\log p}+O(p^{-r}\sqrt{\log p})\Big)\to 1$, which in turn yields that for supp$(\boldsymbol{\delta})\in\mathcal{I}_0^c$,

$$\begin{aligned}\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\max_{i\in H_1\cap\mathcal{A}^c}|Z_j^*|\ge\sqrt{x_p}\Big) &\le Card(H_1)\mathsf{P}\Big(|N(0,1)|\ge\sqrt{x_p}-\sqrt{2\beta_0\log p}+O(p^{-r}\sqrt{\log p})\Big)+o(1)\\ &= o(1).\end{aligned}$$

For $i\in H_2$, we have $(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i=\sum_{j=1}^m a_{i,i_j}\delta_{i_j}=O(p^{-r}\sqrt{\log p})$. Thus,

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{j\in H^c\cap\mathcal{A}^c}|Z_j^*|\ge\sqrt{x_p})=\mathsf{P}\Big(\max_{j\in H_2\cap\mathcal{A}^c}|Y_j|\ge\sqrt{x_p}\Big)I\{\boldsymbol{A}\}+o(1)=:\alpha_pI\{\boldsymbol{A}\}+o(1),$$

where $Y_1,\dots,Y_p$ are i.i.d. $N(0,1)$ random variables. Let $\boldsymbol{E}^*=\{\max_{i\in H\cap\mathcal{A}^c}|Z_i^*|<\sqrt{x_p}\}$. We have

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(M'_{\mathbf{\Omega}^{\frac{1}{2}}}\in R_\alpha)\le\alpha_p+(1-\alpha_p)\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\boldsymbol{E}^{*c})+o(1). \tag{38}$$

Without loss of generality, we assume that for any $i\in H$, $\delta_i>0$. By Lemma 3, we have $\delta_i^o\ge\delta_i^*-O(p^{r-a/2})$ with some $2r<a<1-2r$. Similarly as (33) and (34), it follows from Bonferroni's inequality that

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(M_{\mathbf{\Omega}}\in R_\alpha)\ge\alpha_p+(1-\alpha_p)\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H\cap\mathcal{A}^c}Z_i^*\ge\sqrt{x_p})+o(1). \tag{39}$$

By Lemma 3 and $\delta_i > 0$, we have $\mathsf{P}\Big(\min_{i\in H}\delta_i^* \geq 0 - O(p^{r-a/2})\Big) \to 1$. Hence

$$\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}\Big(\min_{i\in H} Z_i^* \leq -\sqrt{x_p}\Big) \leq \mathsf{P}\Big(\min_{1\leq i\leq m} Y_i \leq -\sqrt{x_p} + O(p^{r-a/2})\Big) = o(1).$$

This implies that

$$\Big|\mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H\cap\mathcal{A}^c}|Z_i^*| \geq \sqrt{x_p}) - \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\max_{i\in H\cap\mathcal{A}^c} Z_i^* \geq \sqrt{x_p})\Big| \leq \mathsf{P}_{\boldsymbol{\delta},\boldsymbol{A}}(\min_{i\in H} Z_i^* \leq -\sqrt{x_p}) = o(1). \quad (40)$$

By (37)-(40), (36) is proved. Hence we have $\mathsf{P}\Big(M_{\boldsymbol{\Omega}} \geq \sqrt{x_p}\Big) \geq \mathsf{P}\Big(M_{\boldsymbol{\Omega}^{\frac{1}{2}}} \geq \sqrt{x_p}\Big) + o(1)$ and this proves Proposition 1(ii). ∎

Proof of Proposition 2 (i). We first prove that for $\beta \geq (1-\sqrt{r})^2 + \varepsilon$,

$$\mathsf{P}\Big(M_{\boldsymbol{I}} \in R_\alpha\Big) \to 1. \quad (41)$$

Let $(Z_1,\ldots,Z_p)'$ be a multivariate normal random vector with p^r-sparse ($r < \frac{1}{2}$) mean $\sqrt{n}\boldsymbol{\delta} = \sqrt{n}(\delta_1,...,\delta_p)'$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$. We assume that the diagonal $\sigma_{i,i} = 1$ for $1 \leq i \leq p$, and $\boldsymbol{\Sigma}$ satisfies condition (C1) and (C2). Then it suffices to show $\mathsf{P}\Big(\max_{1\leq i\leq p}|Z_i| \geq \sqrt{x_p}\Big) \to 1$, where $x_p = 2\log p - \log\log p + q_\alpha$, and q_α is the $1-\alpha$ quantile of $\exp(-\frac{1}{\sqrt{\pi}}\exp(-x/2))$. Note that

$$\mathsf{P}\Big(\max_{1\leq i\leq p}|Z_i| \geq \sqrt{x_p}\Big) \geq \mathsf{P}\Big(\max_{i\in H}(sign(\delta_i)Z_i) \geq \sqrt{x_p}\Big),$$

where $H = \{i:\ \delta_i \neq 0, 1 \leq i \leq p\}$. Thus,

$$\mathsf{P}\Big(\max_{1\leq i\leq p}|Z_i| \geq \sqrt{x_p}\Big) \geq \mathsf{P}\Big(\max_{i\in H} U_i \geq \sqrt{x_p} - a\Big),$$

where $a = \sqrt{2\beta\log p}$ for $\beta \geq (1-\sqrt{r})^2 + \epsilon$ and U_i, $1 \leq i \leq p$, are $N(0,1)$ random variables with covariance matrix $\boldsymbol{\Sigma}$. Because

$$\begin{aligned}\sqrt{x_p} - a &= \sqrt{2\log p - \log\log p + q_\alpha} - \sqrt{2\beta\log p}\\ &\leq (\sqrt{2} - \sqrt{2\beta})\sqrt{\log p} < \sqrt{2r\log p - \log\log p^r - M}\end{aligned}$$

for any $M \in I\!R$, we have by Lemma 6

$$\mathsf{P}\Big(\max_{i\in H} U_i \geq \sqrt{x_p} - a\Big) \quad \geq \quad \mathsf{P}\Big(\max_{i\in H} U_i \geq \sqrt{2\log p^r - \log\log p^r - M}\Big)$$

$$\to \quad 1 - \exp\Big(- \frac{1}{2\sqrt{\pi}} \exp(M/2)\Big),$$

for arbitrary large M. By letting $M \to \infty$, we have $\mathsf{P}\Big(\max_{i \in H} U_i \geq \sqrt{x_p} - a\Big) \to 1$. Thus $\mathsf{P}\Big(M_{\boldsymbol{I}} \in R_\alpha\Big) \to 1$ for any $\beta \geq (1-\sqrt{r})^2 + \epsilon$. It remains to prove that for $\beta < (1-\sqrt{r})^2$,

$$\varlimsup_{p\to\infty} \mathsf{P}\Big(M_{\boldsymbol{I}} \in R_\alpha\Big) \leq \alpha.$$

By noting that $\mathsf{P}\Big(M_{\boldsymbol{I}} \in R_\alpha\Big) \leq \mathsf{P}(\max_{i \in H} Z_i^2 \geq x_p) + \mathsf{P}(\max_{i \in H^c} Z_i^2 \geq x_p)$, it suffices to show that for $\beta < (1-\sqrt{r})^2$,

$$\varlimsup_{p\to\infty} \mathsf{P}(\max_{i \in H^c} Z_i^2 \geq x_p) \leq \alpha, \tag{42}$$

and

$$\mathsf{P}(\max_{i \in H} Z_i^2 \geq x_p) \to 0. \tag{43}$$

Note that $\delta_i = 0$ for $i \in H^c$. It follows from Lemma 6 that (42) holds. For (43), we have

$$\mathsf{P}(\max_{i \in H} Z_i^2 \geq x_p) \leq p^r \mathsf{P}(|N(0,1)| \geq \sqrt{x_p} - \sqrt{2\beta \log p}) \leq C p^{r-(1-\sqrt{\beta})^2} (\log p)^2,$$

where C is a positive constant. Because $\beta < (1-\sqrt{r})^2$, we have (43). Combing (42) and (43), Proposition 2 (i) is proved. ∎

Proof of Proposition 2 (ii). To prove Proposition 2 (ii), we only need to prove the following lemma.

Lemma 7 *Consider $H_1^{'}$: $\boldsymbol{\delta}$ has $m = p^r$, $r < 1/4$ nonzero coordinates with $\sqrt{\omega_{i,i}}|\delta_i| \geq \sqrt{\frac{2\beta_\star \log p}{n}}$, where $\beta_\star > 0$ if $\delta_i \neq 0$. The nonzero locations $l_1 < l_2 < \cdots < l_m$ are randomly and uniformly drawn from $\{1, 2, \ldots, p\}$. If $\beta_\star \geq (1-\sqrt{r})^2 + \epsilon$ for some $\epsilon > 0$, then*

$$P\Big(M_{\boldsymbol{\Omega}} \in R_\alpha\Big) \to 1.$$

Note that $\sqrt{\omega_{i,i}}|\delta_i| = \frac{|\delta_i|}{\sqrt{\sigma_{i,i}}} \cdot \sqrt{\sigma_{i,i}\omega_{i,i}} = \sqrt{\frac{2\beta \log p}{n}} \cdot \sqrt{\sigma_{i,i}\omega_{i,i}} = \sqrt{\frac{2\beta^* \log p}{n}}$, where $\beta^* = \beta\sigma_{i,i}\omega_{i,i}$ and $\beta \geq (1-\sqrt{r})^2/(\min_{1\leq i\leq p} \sigma_{i,i}\omega_{i,i}) + \varepsilon$, we have $\beta^* \geq (1-\sqrt{r})^2 + \varepsilon$. Thus by Lemma 7, we have $\mathsf{P}\Big(M_{\boldsymbol{\Omega}} \in R_\alpha\Big) \to 1$.

We next prove Lemma 7. As the proof of (41), it suffices to show that

$$\mathsf{P}\Big(\min_{i\in H}\frac{|(\boldsymbol{\Omega}\boldsymbol{\delta})_i|}{\sqrt{\omega_{i,i}}}\geq\sqrt{\frac{2\beta\log p}{n}}\Big)\to 1$$

for some $\beta\geq(1-\sqrt{r})^2+\epsilon$. This follows from Lemma 3 immediately. ∎

Proof of Proposition 3 (i). Let $(Z_1,\ldots,Z_p)'$ be a multivariate normal random vector with mean $\boldsymbol{\delta}^o=\sqrt{n}\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\delta}$ and covariance matrix $\boldsymbol{I}_p$. Let $H=\{i_1,\ldots,i_m\}$ be the support of $\boldsymbol{\delta}$. Define

$$H_1=\cup_{k=1}^m\{1\leq j\leq p:|a_{i_kj}|>|a_{i_k(p^{r_1})}|\}=:\cup_{k=1}^m H_{1k},$$

where $r_1>0$ satisfies $\sqrt{r_1+r}<1-\sqrt{\beta}$. We have

$$\mathsf{P}(\max_{1\leq i\leq p}|Z_i|\geq\sqrt{x_p})\leq\mathsf{P}(\max_{i\in H_1}|Z_i|\geq\sqrt{x_p})+\mathsf{P}(\max_{i\in H_1^c}|Z_i|\geq\sqrt{x_p}).$$

Thus it suffices to show

$$\mathsf{P}(\max_{i\in H_1}|Z_i|\geq\sqrt{x_p})\to 0,\tag{44}$$

and

$$\overline{\lim_{p\to\infty}}\,\mathsf{P}(\max_{i\in H_1^c}|Z_i|\geq\sqrt{x_p})\leq\alpha.\tag{45}$$

Define

$$\begin{aligned}\mathcal{I}_1 \;=\; \Big\{&1\leq i_1<...<i_m\leq p:\ \text{there exist some } i_k \text{ and } i_j\neq i_k\\ &\qquad\text{such that } |a_{i,i_j}|>|a_{i,(p^a)}| \text{ for some } i\in H_{1k}\Big\},\end{aligned}$$

where $a>2r$ satisfies $3r+r_1+a<1$. Then Card($\mathcal{I}_1$)$\leq p^{1+r+r_1+a}C_p^{m-2}=o(1)C_p^m$. It follows that for $i\in H_{1k}$ and $(i_1,\ldots,i_m)\notin\mathcal{I}_1$,

$$\begin{aligned}\sqrt{n}|(\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i|=\sqrt{n}\Big|\sum_{k=1}^p a_{ik}\delta_k\Big| &= \sqrt{n}\Big|a_{i,i_k}\delta_{i_k}+\sum_{j\in H,j\neq i_k}a_{ij}\delta_j\Big|\\ &\leq \sqrt{2\beta\log p}+O(p^{r-a/2}(\log p)^{\frac{1}{2}}).\end{aligned}$$

Thus $\mathsf{P}\Big(\max_{i\in H_1} |Z_i| \geq \sqrt{x_p}\Big) \leq Cp^{r+r_1-(1-\sqrt{\beta})^2}(\log p)^2+o(1) = o(1)$ and this proves (44).

Set $H_{2k} = \{1 \leq j \leq p : |a_{i_k(p^{r_1})}| \geq |a_{i_kj}| > |a_{i_k(p^{r_2})}|\}$ and $H_2 = \cup_{k=1}^m H_{2k}$. Let

$$\mathcal{I}_2 = \Big\{1 \leq i_1 < ... < i_m \leq p : \text{ there exist some } i_k \text{ and } i_j \neq i_k \\ \text{such that } |a_{i,i_j}| > |a_{i,(p^a)}| \text{ for some } i \in H_{2k}\Big\},$$

where $a > 2r$ and $r_2 > 2r$ satisfy $3r + r_2 + a < 1$. We have Card($\mathcal{I}_2$)$\leq p^{1+r+r_2+a}C_p^{m-2} = o(1)C_p^m$. For $i \in H_{2k}$ and $(i_1, \dots, i_m) \notin \mathcal{I}_2$, we have

$$\begin{aligned}\sqrt{n}|(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i| = \sqrt{n}\Big|\sum_{k=1}^p a_{ik}\delta_k\Big| &= \sqrt{n}\Big|a_{i,i_k}\delta_{i_k} + \sum_{j\in H, j\neq i_k} a_{ij}\delta_j\Big| \\ &= O(p^{-r_1/2}(\log p)^{\frac{1}{2}}) + O(p^{r-a/2}(\log p)^{\frac{1}{2}}).\end{aligned}$$

Hence

$$\mathsf{P}\Big(\max_{i\in H_2} |Z_i| \geq \sqrt{x_p}\Big) \leq CCard(H_2)p^{-1} + o(1) = o(1). \tag{46}$$

For $i \in H_3 := (H_1 \cup H_2)^c$, we have $\sqrt{n}(\mathbf{\Omega}^{\frac{1}{2}}\boldsymbol{\delta})_i = \sqrt{n}\sum_{k=1}^p a_{ik}\delta_k = O(p^{r-r_2/2}(\log p)^{\frac{1}{2}})$. This, together with Lemma 6, implies that

$$\overline{\lim_{p\to\infty}}\, \mathsf{P}\Big(\max_{i\in H_3} |Z_i| \geq \sqrt{x_p}\Big) \leq \alpha. \tag{47}$$

By (46) and (47), we prove (45) and complete the proof of Proposition 3 (i). ∎

Proof of Proposition 3 (ii). It suffices to verify the condition in Lemma 7. Note that for $i \in H$,

$$\sqrt{\omega_{i,i}}|\delta_i| = \max_{1\leq j\leq p} |a_{ij}||\delta_i| \cdot \sqrt{\omega_{i,i}/\max_{1\leq j\leq p} a_{ij}^2} = \sqrt{\frac{2\beta \log p}{n}} \cdot \sqrt{\omega_{i,i}/\max_{1\leq j\leq p} a_{ij}^2} = \sqrt{\frac{2\beta^* \log p}{n}},$$

where $\beta^* = \beta\omega_{i,i}/\max_{1\leq j\leq p} a_{ij}^2$. If $\beta \geq (1-\sqrt{r})^2/(\min_{1\leq i\leq p}(\omega_{i,i}/\max_{1\leq j\leq p} a_{ij}^2)) + \varepsilon$, then we have $\beta^* \geq (1-\sqrt{r})^2 + \varepsilon$. Thus, by Lemma 7, we have $\mathsf{P}\Big(M_{\mathbf{\Omega}} \in R_\alpha\Big) \to 1$. and the proof of Proposition 3 (ii) is complete. ∎

Proof of Proposition 1 in Section 3.2.2. The proof that CLIME satisfies (8) in Section 3.2.2 follows from Cai, Liu and Luo (2011) and Cai, Liu and Zhou (2013). For the adaptive thresholding estimator, by the proof of Theorem 1 in Cai and Liu (2011), we have $\|\hat{\mathbf{\Sigma}}^\star - \mathbf{\Sigma}\|_{L_1} \le C s_{2,p}(\log p/n)^{(1-q)/2}$ with probability tending to one. By the inequality $\|(\hat{\mathbf{\Sigma}}^\star)^{-1} - \mathbf{\Omega}\|_{L_1} \le \|(\hat{\mathbf{\Sigma}}^\star)^{-1}\|_{L_1}\|\hat{\mathbf{\Sigma}}^\star - \mathbf{\Sigma}\|_{L_1}\|\mathbf{\Omega}\|_{L_1}$ and $|(\hat{\mathbf{\Sigma}}^\star)^{-1} - \mathbf{\Omega}|_\infty \le \|(\hat{\mathbf{\Sigma}}^\star)^{-1}\|_{L_1}|\hat{\mathbf{\Sigma}}^\star - \mathbf{\Sigma}|_\infty\|\mathbf{\Omega}\|_{L_1}$, we prove (8) in Section 3.2.2 holds.

We now consider $\mathbf{\Sigma} \in \mathcal{F}_\alpha(M_1, M_2)$. Let $|\sigma_{i,(1)}| \ge \ldots \ge |\sigma_{i,(p)}|$ be an arrangement of $\sigma_{i1}, \ldots, \sigma_{ip}$. Hence, we have $\max_i |\sigma_{i,(j)}| = O(j^{-1-\alpha})$. By the proof of Lemma 2 in Cai and Liu (2011), we have with probability tending to one,

$$\max_{1\le i,j\le p} |\frac{\hat{\sigma}_{ij} - \sigma_{ij}}{\hat{\theta}_{ij}^{1/2}}| \le 2\sqrt{\frac{\log p}{n}} - \frac{1}{\sqrt{n\log p}}.$$

Let $M > 0$ be a sufficiently large number and $k_n = [Mn\log p]^{1/(2(1+\alpha))}$. Then for any $j \ge k_n$, $|\sigma_{i,(j)}| \le CM^{-1}\sqrt{1/(n\log p)}$ and correspondingly, $|\hat{\sigma}_{i,(j)}| \le \lambda_{i,(j)}$, where $\lambda_{ij} = \delta\sqrt{\frac{\hat{\theta}_{ij}\log p}{n}}$ as defined in Cai, Liu and Xia (2013). Hence it is easy to see that, with probability tending to one,

$$\sum_{j=1}^p |\hat{\sigma}_{ij} I\{|\hat{\sigma}_{ij}| \ge \lambda_{ij}\} - \sigma_{ij}| \le C[Mn\log p]^{1/(2(1+\alpha))}\sqrt{\frac{\log p}{n}} = o(\frac{1}{\log p}) \tag{48}$$

by the condition $\log p = o(n^{\alpha/(4+3\alpha)})$. By (48) and the above arguments, we prove (8) in Section 3.2.2. ∎

Proof of Theorem 5. Note that the CLIME estimator only depends on $\mathbf{\Sigma}_n$. So $\hat{\mathbf{\Omega}}$ is independent with $\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}}$. Let $x_p = 2\log p - \log\log p + q_\alpha$. It yields that

$$\mathsf{P}_{H_0}\Big(M_{\hat{\mathbf{\Omega}}} \ge x_p|\mathbf{\Sigma}_n\Big) = \mathsf{P}_{H_0}\Big(|(\boldsymbol{D}_n\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}\boldsymbol{D}_n)^{1/2}\boldsymbol{Z}|_\infty \ge x_p|\mathbf{\Sigma}_n\Big),$$

where $\boldsymbol{D}_n = (diag(\hat{\mathbf{\Omega}}\mathbf{\Sigma}_n\hat{\mathbf{\Omega}}))^{-1/2}$ and $\boldsymbol{Z}$ are standard multivariate normal random vector which is independent with $\mathbf{\Sigma}_n$. By the Sidak's inequality, we have on the event $\{|diag(\boldsymbol{D}_n\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}\boldsymbol{D}_n) - \boldsymbol{I}_p|_\infty = o(1/\log p)\}$,

$$\begin{aligned}\mathsf{P}_{H_0}\Big(|(\boldsymbol{D}_n\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}\boldsymbol{D}_n)^{1/2}\boldsymbol{Z}|_\infty \ge x_p|\mathbf{\Sigma}_n\Big) &= 1 - \mathsf{P}_{H_0}\Big(|(\boldsymbol{D}_n\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}\boldsymbol{D}_n)^{1/2}\boldsymbol{Z}|_\infty < x_p|\mathbf{\Sigma}_n\Big) \\ &\le 1 - \prod_{i=1}^p \mathsf{P}(|\xi_i| \le x_p|\mathbf{\Sigma}_n)\end{aligned}$$

$$= \alpha + o(1),$$

where, given $\mathbf{\Sigma}_n$, ξ_i is the centered normal variable with variance $(diag(\boldsymbol{D}_n\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}\boldsymbol{D}_n))_{i,i}$. So it is enough to prove $|diag(\boldsymbol{D}_n\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}\boldsymbol{D}_n) - \boldsymbol{I}_p|_\infty = o_{\mathsf{P}}(1/\log p)$. By the definition of $\hat{\mathbf{\Omega}}$, we have $\mathsf{P}(\|\hat{\mathbf{\Omega}}\|_{L_1} \leq M_p) \to 1$. So under $M_p^2 = o(\sqrt{n}/(\log p)^{3/2})$, we have

$$|diag((\hat{\mathbf{\Omega}}\mathbf{\Sigma}_n\hat{\mathbf{\Omega}})) - diag((\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}))|_\infty = o_{\mathsf{P}}(1/\log p).$$

By Lemma 1 in Cai, Liu and Zhou (2013), we have $\mathsf{P}(|\hat{\mathbf{\Omega}}\mathbf{\Sigma}_n\hat{\mathbf{\Omega}} - \hat{\mathbf{\Omega}}|_\infty \leq CM_p\sqrt{\log p/n}) \to 1$. By the proof of Theorem 2 in Cai, Liu and Zhou (2013), $\mathsf{P}(|\hat{\mathbf{\Omega}} - \mathbf{\Omega}|_\infty \leq CM_p\sqrt{\log p/n}) \to 1$. Hence, under the condition

$$M_p^2 = o(\sqrt{n}/(\log p)^{3/2}), \tag{49}$$

we have $|diag(\boldsymbol{D}_n\hat{\mathbf{\Omega}}\mathbf{\Sigma}\hat{\mathbf{\Omega}}\boldsymbol{D}_n) - \boldsymbol{I}_p|_\infty = o_{\mathsf{P}}(1/\log p)$.

The above arguments imply that

$$\mathsf{P}\Big(n \max_i (\hat{\mathbf{\Omega}}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} - \boldsymbol{\delta})_i)^2/\hat{\omega}_{i,i}^{(0)} \leq 2\log p\Big) \to 1. \tag{50}$$

To show that $\mathsf{P}_{H_1}(\Phi_\alpha(\hat{\mathbf{\Omega}}) = 1) \to 1$, we first prove

$$\mathsf{P}\Big(\max_i \sum_{j=1}^p \hat{\omega}_{ij}^2 \leq C\Big) \to 1. \tag{51}$$

Actually, with probability tending to one, we have

$$\begin{aligned} \sum_{j=1}^p \hat{\omega}_{ij}^2 &\leq 2\sum_{j=1}^p (\hat{\omega}_{ij} - \omega_{ij})^2 + 2\sum_{j=1}^p \omega_{ij}^2 \\ &\leq CM_n^2\sqrt{\frac{\log p}{n}} + C. \end{aligned} \tag{52}$$

Hence (51) holds. By the proof of (6), we have $\max_{i\in H} |(\hat{\mathbf{\Omega}}\boldsymbol{\delta})_i/\hat{\omega}_{i,i}^{1/2} - \hat{\omega}_{i,i}^{1/2}\delta_i| \leq Cp^{-\epsilon}\max_{i\in H}|\delta_i|$ with probability tending to one. Without loss of generality, we can assume that $\max_{i\in H}|\delta_i| \leq C\sqrt{\log p/n}$ for some large $C > 0$. Otherwise, $\max_{i\in H} |(\hat{\mathbf{\Omega}}\boldsymbol{\delta})_i/\hat{\omega}_{i,i}^{1/2}| \geq C\sqrt{\log p/n}$ for some large $C > 0$ and by (50) we have $\mathsf{P}_{H_1}(\Phi_\alpha(\hat{\mathbf{\Omega}}) = 1) \to 1$. Let i_0 (may be random) be the index such that $|\delta_{i_0}/\sigma_{i_0 i_0}^{1/2}| \geq \sqrt{2\beta\log p/n}$. Then we have with probability tending to one,

$|(\hat{\mathbf{\Omega}}\boldsymbol{\delta})^2_{i_0}/\hat{\omega}_{i_0i_0}| \geq 2\sigma_{i_0i_0}\omega_{i_0i_0}(\beta+o(1))\log p/n \geq (2+\epsilon_1)\log p/n$ for some $\epsilon_1 > 0$. Note that, by the independence between $(\boldsymbol{X}-\boldsymbol{\mu}_1, \boldsymbol{Y}-\boldsymbol{\mu}_2)$ and the positions of nonzero locations in $\boldsymbol{\delta}$,

$$\begin{aligned}
&\mathsf{P}\Big(n(\hat{\mathbf{\Omega}}(\bar{\boldsymbol{X}}-\bar{\boldsymbol{Y}}-\boldsymbol{\delta})_{i_0})^2/\hat{\omega}^{(0)}_{i_0i_0} \geq \sqrt{\log p}\Big)\\
&= \sum_{i=1}^{p}\mathsf{P}\Big(n(\hat{\mathbf{\Omega}}(\bar{\boldsymbol{X}}-\bar{\boldsymbol{Y}}-\boldsymbol{\delta})_{i})^2/\hat{\omega}^{(0)}_{i,i} \geq \sqrt{\log p}\Big)\mathsf{P}(i_0=i)\\
&= o(1).
\end{aligned}$$

This proves $\mathsf{P}_{H_1}(\Phi_\alpha(\hat{\mathbf{\Omega}})=1) \to 1$.

References

[1] Berman, S.M. (1962). A law of large numbers for the maximum of a stationary Gaussian sequence. *Ann. Math. Statist.* 33: 894-908.

[2] Baraud, Y. (2002). Non-Asymptotic Minimax Rates of Testing in Signal Detection. *Bernoulli* 8: 577-606.

[3] Cai, T., Liu, W.D. and Zhou, H.B. (2013). Estimating Sparse Precision Matrix: Optimal Rates of Convergence and Adaptive Estimation. Technical report.

[4] Cai, T. and Liu, W.D. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* 106:672-684.

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