Supplementary Appendix of "Simultaneous Covariance Inference for Multimodal Integrative Analysis"

Yin Xia, Lexin Li, Samuel N. Lockhart, and William J. Jagust

In this supplementary appendix, we first present additional simulation results, then provide a collection of technical lemmas and the proofs of the theorems in the paper. Note that the proof of Proposition 1 follows that of Theorem 1, and is thus omitted.

A1 Additional Simulation Results

We further consider the setting when the variables within $\boldsymbol{X}_k$ are correlated. Note that, by construction, the variables within $\boldsymbol{Y}_k$ are always correlated. Specifically, we adopt a similar simulation setup as in Section 4.1 of the paper, and add a covariance structure $\boldsymbol{\Sigma}_X$ for the entries of $\boldsymbol{X}_k$. We set $\boldsymbol{\Sigma}_X = c^{-1}\mathsf{Cov}(\boldsymbol{X}_k) = (\sigma_{X,i,j})$, where $c = \mathsf{Var}\{N(0,1)\} = 1$ for Distribution 1 and $c = \mathsf{Var}\{t(10)\} = 1.25$ for Distribution 2, and $\sigma_{X,i,i} = 1$, $\sigma_{X,i,j} = 0.5$ for $5(k-1)+1 \le i,j \le 5k$, with $k = 1,\ldots,[p/5]$, and $\sigma_{X,i,j} = 0$ otherwise. We have chosen this block covariance structure mostly to simplify the computation in the data generation step, since it allows us to efficiently compute the true covariance structure in the simulations. Then we generate n copies of the second modality, $\{\boldsymbol{Y}_k\}_{k=1}^n$ in the same way as the paper. In this case, we have $\mathsf{Cov}(\boldsymbol{X}_k, \boldsymbol{Y}_k) = c\boldsymbol{\Sigma}_X\boldsymbol{\Sigma}_{1,2}$, with the following three covariance structures similar as those in the paper:

Structure 1: $\boldsymbol{\Sigma}_{1,2} = (\sigma_{i,j})$ where $\sigma_{i,j} = 0.8$ Bernoulli$(1, 0.05)$ for $1 \le i,j \le p$;

Structure 2: $\boldsymbol{\Sigma}_{1,2} = (\sigma_{i,j})$ where $\sigma_{i,i} =$Uniform$(0.5, 2)$, $\sigma_{i,j} = 0.8$ for $5(k-1)+1 \le i,j \le 5k$, with $k = 1,\ldots,[p/5]$, and $\sigma_{i,j} = 0$ otherwise;

Structure 3: $\boldsymbol{\Sigma}_{1,2} = (\sigma_{i,j})$ where $\sigma_{i,i} =$Uniform$(0.5, 2)$, $\sigma_{i,i+1} = \sigma_{i+1,i} = 0.8$ for $i = 1,\ldots,p-1$, and $\sigma_{i,j} = 0$ otherwise.

For those pairs of regions with fewer or equal than 5 nonzero correlations, we set the corresponding submatrix equal to zero. We report the empirical FDR and power, both in percentages, based on 100 data replications, in Table S1 for $n = 100$, and in Table S2 for $n = 150$. We observe essentially the same pattern as before, in that our proposed test obtains an empirical FDR well controlled under the nominal level, and achieves a much higher empirical power than the competing methods.

Normal distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
	Empirical FDR					
Xie and Kang	0.0	0.0	0.0	0.3	0.0	0.0
Sparse CCA	97.3	17.6	10.8	96.1	16.6	14.1
Our test	5.4	4.3	4.5	5.1	4.3	4.4
	Empirical power					
Xie and Kang	5.8	28.9	17.2	33.2	44.2	26.8
Sparse CCA	19.3	2.9	1.4	20.0	1.2	1.6
Our test	37.1	99.6	90.6	84.4	99.6	91.2
t distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
	Empirical FDR					
Xie and Kang	0.0	0.0	0.0	0.5	0.0	0.0
Sparse CCA	96.8	14.2	15.3	98.3	21.0	12.8
Our test	3.1	3.3	3.2	2.0	3.2	3.4
	Empirical power					
Xie and Kang	4.3	19.0	13.3	10.2	28.6	17.7
Sparse CCA	14.6	1.7	2.0	21.4	1.3	0.7
Our test	44.2	97.0	80.2	47.5	96.8	79.5

Table S1: Empirical FDR and empirical power, in percentages, for the proposed testing procedure. It is also compared with the testing method of Xie and Kang (2017) and sparse CCA. The results are based on 100 data replications. The significance level is set at $\alpha = 5\%$. The sample size is $n = 100$.

A2 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*

$$\lim_{c \to \infty} \frac{P(X > c, Y > c)}{\{2\pi(1-\rho)^{1/2} c^2\}^{-1} \exp\left(-\frac{c^2}{1+\rho}\right) (1+\rho)^{3/2}} = 1,$$

Normal distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
	Empirical FDR					
Xie and Kang	0.0	0.0	0.0	0.1	0.0	0.0
Sparse CCA	98.3	13.8	9.5	95.8	14.1	8.5
Our test	4.0	3.9	3.9	4.3	4.0	3.9
	Empirical power					
Xie and Kang	52.6	93.8	68.7	91.8	98.6	82.3
Sparse CCA	16.4	0.6	0.6	25.8	0.8	0.6
Our test	93.6	100.0	99.9	99.3	100.0	99.9
t distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
	Empirical FDR					
Xie and Kang	0.0	0.0	0.0	0.0	0.0	0.0
Sparse CCA	92.7	9.8	4.1	96.9	7.7	13.2
Our test	2.7	3.0	3.1	2.5	3.0	2.9
	Empirical power					
Xie and Kang	46.1	76.6	46.6	67.8	90.8	63.9
Sparse CCA	17.2	1.1	0.1	22.4	1.1	1.2
Our test	95.3	100.0	99.6	97.6	100.0	99.6

Table S2: Empirical FDR and empirical power, in percentages, for the proposed testing procedure. It is also compared with the testing method of Xie and Kang (2017) and sparse CCA. The results are based on 100 data replications. The significance level is set at $\alpha = 5\%$. The sample size is $n = 150$.

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

Lemma 3 *Under Assumption (A2), there exists a constant $C > 0$, such that*

$$P\left(\max_{i,j\in\mathcal{S}_{l,g}} |\hat{\theta}_{i,j} - \theta_{i,j}|/\sigma_{i,i}\sigma_{j,j} \geq C\frac{\varepsilon_n}{\log p_{l,g}}\right) = O\left(p_{l,g}^{-1} + n^{-\epsilon/8}\right),$$

where $\varepsilon_n = \max\{(\log p_{l,g})^{1/6}/n^{1/2}, (\log p_{l,g})^{-1}\} \to 0$ as $n, p_{l,g} \to \infty$.

Lemma 3 is about the large deviation for $\hat{\theta}_{i,j}$. Its proof is given in Cai et al. (2013).

Lemma 4 *Under Assumption (A2), there exists a constant $C > 0$, such that*

$$P\left(\max_{(i,j)\in\Lambda} \frac{(\tilde{\sigma}_{i,j} - \sigma_{i,j})^2}{\theta_{i,j}/n} \geq y^2\right) \leq C|\Lambda|\{1 - \phi(y)\} + O\left(p_{l,g}^{-1} + n^{-\epsilon/8}\right),$$

uniformly for $0 \le y \le (8\log p_{l,g})^{1/2}$*, and* $\Lambda \subseteq \{(i,j) : i, \in \mathcal{S}_l, j \in \mathcal{S}_g\}$*, where* Λ *denotes any subset of* $\{(i,j) : i, \in \mathcal{S}_l, j \in \mathcal{S}_g\}$*, and* $\tilde{\sigma}_{i,j}$ *is the individual entry of the matrix,*

$$\tilde{\mathbf{\Sigma}}_{l,g} = (\tilde{\sigma}_{i,j})_{p_{l,g}\times p_{l,g}} = \frac{1}{n}\sum_{k=1}^{n}\left\{\boldsymbol{Z}_k^{(l,g)} - \boldsymbol{\mu}^{(l,g)}\right\}\left\{\boldsymbol{Z}_k^{(l,g)} - \boldsymbol{\mu}^{(l,g)}\right\}^{\mathsf{T}}.$$

Lemma 5 *For any random vector* $\boldsymbol{W} = (w_1, \ldots, w_b)$ *with* $\mathsf{E}\boldsymbol{W} = 0$*, and* $\boldsymbol{W} = \xi_1 + \cdots + \xi_n$*, where* $\{\xi_k = (\xi_{1,k}, \ldots, \xi_{b,k}), k = 1, \ldots, n\}$ *are independent random vectors with* $|\xi_{i,k}| \le \tau$*, for* $1 \le i \le b$*, we have, for any* $y, \epsilon > 0$*,*

$$\begin{aligned} \mathsf{P}(|\boldsymbol{W}| \ge y) &\le \mathsf{P}(|\boldsymbol{N}| \ge y - \epsilon) + c_1 b^{5/2}\exp\left(-\frac{\epsilon}{c_2 b^3 \tau}\right), \\ \mathsf{P}(|\boldsymbol{W}| \ge y) &\ge \mathsf{P}(|\boldsymbol{N}| \ge y + \epsilon) - c_1 b^{5/2}\exp\left(-\frac{\epsilon}{c_2 b^3 \tau}\right), \end{aligned}$$

for some absolute constants $c_1, c_2 > 0$*, where* $|\cdot|$ *is any vector norm,* $\boldsymbol{N}$ *is a normal random vector with* $\mathsf{E}\boldsymbol{N} = 0$ *and the same covariance matrix as* $\boldsymbol{W}$*.*

Lemma 5 is based on Theorem 1 of Zaïtsev (1987).

A3 Proof of Theorem 1

To prove this theorem, we truncate the random variables and use the normal approximation in Lemma 5 to translate the problem into deriving the limiting null distribution of normal random variables under the same dependence structure. We further divide the sets of pairs of random variables into small subsets, and show that the behavior of weakly correlated random variables dominates the rest, and the corresponding extreme value behavior is asymptotically the same as the maximum of independent normal random variables.

Without loss of generality, we assume that $\boldsymbol{\mu} = 0$ and $\sigma_{i,i} = 1$ for $1 \le i \le 2p$. Define

$$\hat{T}_{i,j} = \frac{\hat{\sigma}_{i,j}}{(\theta_{i,j}/n)^{1/2}} \quad \text{and} \quad \tilde{T}_{i,j} = \frac{\tilde{\sigma}_{i,j}}{(\theta_{i,j}/n)^{1/2}},$$

Let $\hat{M}_{l,g} = \max_{i\in\mathcal{S}_l^X, j\in\mathcal{S}_g^Y}\hat{T}_{i,j}^2$, and $\tilde{M}_{l,g} = \max_{i\in\mathcal{S}_l^X, j\in\mathcal{S}_g^Y}\tilde{T}_{i,j}^2$. By Lemma 3, we focus on the event that $|\hat{\theta}_{i,j} - \theta_{i,j}| \le C\varepsilon_n/\log p_{l,g}$. We have $|M_{l,g} - \hat{M}_{l,g}| \le C\hat{M}_{l,g}\epsilon_n/\log p_{l,g}$, and $|\hat{M}_{l,g} - \tilde{M}_{l,g}| \le Cn\max_{i\in\mathcal{S}_l^X, j\in\mathcal{S}_g^Y}\bar{X}_i^2\bar{Y}_j^2 + Cn^{1/2}\tilde{M}_{l,g}^{1/2}(\max_{i\in\mathcal{S}_l^X}\bar{X}_i^2 + \max_{j\in\mathcal{S}_g^Y}\bar{Y}_j^2)$. Thus by the fact that $\max_{i\in\mathcal{S}_l^X}|\bar{X}_i| + \max_{j\in\mathcal{S}_g^Y}|\bar{Y}_j| = O_{\mathsf{P}}((\log p_{l,g}/n)^{1/2})$, it suffices to show that

$$\mathsf{P}\left(\tilde{M}_{l,g} - 2\log(p_l p_g) + \log\log(p_l p_g) \le t\right) \to \exp\{-\pi^{-1/2}e^{-t/2}\}.$$

Recall that $\mathcal{A}^{(l,g)} = \left\{(i,j) : i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y\right\}$. Define

$$\begin{aligned} E_0 = &\left\{(i,j) : i \in D_0 \cap \mathcal{S}_l^X, j \in \mathcal{S}_g^Y\right\} \cup \left\{(i,j) : j \in D_0 \cap \mathcal{S}_g^Y, i \in \mathcal{S}_l^X\right\}, \\ &\text{with } |D_0| = o(\min\{p_l, p_g\}). \end{aligned}$$

Let $\tilde{M}_{l,g}^{\mathcal{A}^{(l,g)}\setminus E_0} = \max_{(i,j)\in\mathcal{A}^{(l,g)}\setminus E_0} \tilde{T}_{i,j}^2$, and $\tilde{M}_{l,g}^{E_0} = \max_{(i,j)\in E_0} \tilde{T}_{i,j}^2$. Let $y_{l,g}^2 = 2\log(p_l p_g) - \log\log(p_l p_g) + t$. Then we have

$$\left|\mathsf{P}\left(\tilde{M}_{l,g} \geq y_{l,g}^2\right) - \mathsf{P}\left(\tilde{M}_{l,g}^{\mathcal{A}^{(l,g)}\setminus E_0} \geq y_{l,g}^2\right)\right| \leq \mathsf{P}\left(M_{l,g}^{E_0} \geq y_{l,g}^2\right).$$

Noting that $|E_0| \leq \min\{o(p_g)p_l, o(p_l)p_g\} = o(p_l p_g)$. Thus by Lemma 4, we have

$$\mathsf{P}\left(M_{l,g}^{E_0} \geq y_{l,g}^2\right) \leq C|E_0|(p_l p_g)^{-1} + o(1) = o(1).$$

Thus it suffices to prove that, for any $t \in \mathbb{R}$,

$$\mathsf{P}\left(\tilde{M}_{l,g}^{\mathcal{A}^{(l,g)}\setminus E_0} - 2\log(p_l p_g) + \log\log(p_l p_g) \leq t\right) \to \exp\{-\pi^{-1/2}e^{-t/2}\},$$

as $n, p_{l,g} \to \infty$. We rearrange the two-dimensional indices $\mathcal{A}^{(l,g)} \setminus E_0$ in an arbitrary order and set them as $\{(i_m, j_m) : 1 \leq m \leq q\}$ with $q = |\mathcal{A}^{(l,g)} \setminus E_0|$. Define $\theta_m = \theta_{i_m,j_m}$. Let $W_{k,m} = (X_{k,i_m}Y_{k,j_m} - \sigma_{i_m,j_m})$, and $\hat{W}_{k,m} = W_{k,m}I(|W_{k,m}| \leq \tau_n) - \mathsf{E}\{W_{k,m}I(|W_{k,m}| \leq \tau_n)\}$, $V_m = 1/(n\theta_m)^{1/2}\sum_{k=1}^n W_{k,m}$, and $\hat{V}_m = 1/(n\theta_m)^{1/2}\sum_{k=1}^n \hat{W}_{k,m}$, where $\tau_n = \eta^{-1}8\log(p_{l,g}+n)$ if a sub-gaussian tail is assumed, and $\tau_n = n^{1/2}/(\log p_{l,g})^8$ if a polynomial tail is assumed. Then we have, under the null,

$$\tilde{M}_{l,g}^{\mathcal{A}^{(l,g)}\setminus E_0} = \max_{1\leq m\leq q} V_m^2.$$

We next show that $\mathsf{E}\{W_{k,m}I(|W_{k,m}| \leq \tau_n)\}$ is negligible. That is, we have

$$\begin{aligned}
&\max_{1\leq m\leq q} n^{-1/2}\sum_{k=1}^n \mathsf{E}\{|W_{k,m}|I(|W_{k,m}| \leq \tau_n)\} \leq Cn^{1/2}\max_{1\leq k\leq n}\max_{1\leq m\leq q}\mathsf{E}\{|W_{k,m}|I(|W_{k,m}| \leq \tau_n)\} \\
&\leq Cn^{1/2}(p_{l,g}+n)^{-4}\max_{1\leq k\leq n}\max_{1\leq m\leq q}\mathsf{E}\{|W_{k,m}|\exp(\eta|W_{k,m}|/2)\} \leq Cn^{1/2}(p_{l,g}+n)^{-4}.
\end{aligned}$$

if a sub-gaussian tail is assumed, and

$$\begin{aligned}
&\max_{1\leq m\leq q} n^{-1/2}\sum_{k=1}^n \mathsf{E}\{|W_{k,m}|I(|W_{k,m}| \leq \tau_n)\} \leq Cn^{1/2}\max_{1\leq k\leq n}\max_{1\leq m\leq q}\mathsf{E}\{|W_{k,m}|I(|W_{k,m}| \leq \tau_n)\} \\
&\leq Cn^{1/2}\frac{\mathsf{E}(|W_{k,m}||W_{k,m}|^{2\gamma_0+1+\epsilon/2})}{(n^{1/2}/(\log p)^8)^{2\gamma_0+1+\epsilon/2}} \leq Cn^{-\gamma_0-\epsilon/8},
\end{aligned}$$

if a polynomial tail is assumed. This in turn yields that

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\leq m\leq q}|V_m - \hat{V}_m| \geq (\log p_{l,g})^{-1}\right) \leq \mathsf{P}\left(\max_{1\leq m\leq q}\max_{1\leq k\leq n}|W_{k,m}| \geq \tau_n\right) \\
&\leq np_{l,g}\left\{\max_{i\in\mathcal{S}_l^X}\mathsf{P}(X_i^2 \geq \tau_n/2) + \max_{j\in\mathcal{S}_g^Y}\mathsf{P}(Y_j^2 \geq \tau_n/2)\right\} = O(p_{l,g}^{-1} + n^{-\epsilon/8}).
\end{aligned}$$

Due to the fact that

$$|\max_{1\leq m\leq q} V_m^2 - \max_{1\leq m\leq q}\hat{V}_m^2| \leq 2\max_{1\leq m\leq q}|V_m|\max_{1\leq m\leq q}|V_m - \hat{V}_m| + \max_{1\leq m\leq q}|V_m - \hat{V}_m|^2,$$

it is sufficient to show that, for any $t \in \mathbb{R}$,

$$\mathsf{P}\left(\max_{1\le m\le q}\hat{V}_m^2 - 2\log(p_l p_g) + \log\log(p_l p_g) \le t\right) \to \exp\{-\pi^{-1/2}e^{-t/2}\},$$

as $n, p_{l,g} \to \infty$. By Lemma 1, for any integer s with $0 < s < q/2$,

$$\begin{aligned}\sum_{d=1}^{2s}(-1)^{d-1}\sum_{1\le m_1<\cdots<m_d\le q}\mathsf{P}\left(\bigcap_{j=1}^{d}E_{m_j}\right) &\le \mathsf{P}\left(\max_{1\le m\le q}\hat{V}_m^2 \ge y_{l,g}^2\right)\\ &\le \sum_{d=1}^{2s-1}(-1)^{d-1}\sum_{1\le m_1<\cdots<m_d\le q}\mathsf{P}\left(\bigcap_{j=1}^{d}E_{m_j}\right), \quad \text{(S1)}\end{aligned}$$

where $E_{m_j} = \{\hat{V}_{m_j}^2 \ge y_{l,g}^2\}$. Let $\tilde{W}_{k,m} = \hat{W}_{k,m}/(\theta_m)^{1/2}$ for $1 \le m \le q$ and $\boldsymbol{W}_k = (\tilde{W}_{k,m_1}, \ldots, \tilde{W}_{k,m_d})$, for $1 \le k \le n$. Define $|\boldsymbol{a}|_{\min} = \min_{1\le i\le d}|a_i|$ for any vector $\boldsymbol{a} \in \mathbb{R}^d$. Then we have

$$\mathsf{P}\left(\bigcap_{j=1}^{d}E_{m_j}\right) = \mathsf{P}\left(\left|n^{-1/2}\sum_{k=1}^{n}\boldsymbol{W}_t\right|_{\min} \ge y_{l,g}\right).$$

By Lemma 5, we have

$$\begin{aligned}\mathsf{P}\left(\left|n^{-1/2}\sum_{k=1}^{n}\boldsymbol{W}_k\right|_{\min} \ge y_{l,g}\right) &\le \mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \ge y_{l,g} - \epsilon_n\{\log(p_l p_g)\}^{-1/2}\right)\\ &\quad + c_1 d^{5/2}\exp\left(-\frac{n^{1/2}\epsilon_n}{c_2 d^3\tau_n\{\log(p_l p_g)\}^{1/2}}\right), \quad \text{(S2)}\end{aligned}$$

where $c_1 > 0$ and $c_2 > 0$ are absolute constants, $\epsilon_n \to 0$ and is to be specified later, and $\boldsymbol{N}_d = (N_{m_1}, \ldots, N_{m_d})$ is a normal vector with $\mathsf{E}(\boldsymbol{N}_d) = 0$ and $\mathsf{Cov}(\boldsymbol{N}_d) = \mathsf{Cov}(\boldsymbol{W}_1)$. Note that d is a fixed integer not depending on n and $p_{l,g}$. Because $\log p_{l,g} = o(n^{1/5})$, we can let $\epsilon_n \to 0$ sufficiently slow such that

$$c_1 d^{5/2}\exp\left(-\frac{n^{1/2}\epsilon_n}{c_2 d^3\tau_n\{\log(p_l p_g)\}^{1/2}}\right) = O\left(p_{l,g}^{-M}\right) \quad \text{(S3)}$$

for any large $M > 0$. Combining (S1), (S2) and (S3), we have

$$\begin{aligned}&\mathsf{P}\left(\max_{1\le m\le q}\hat{V}_m^2 \ge y_{l,g}^2\right)\\ &\le \sum_{d=1}^{2s-1}(-1)^{d-1}\sum_{1\le m_1<\cdots<m_d\le q}\mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \ge y_{l,g} - \epsilon_n\{\log(p_l p_g)\}^{-1/2}\right) + o(1). \quad \text{(S4)}\end{aligned}$$

Similarly, using Lemma 1 again, we get

$$\mathsf{P}\left(\max_{1\le m\le q}\hat{V}_m^2 \ge y_{l,g}^2\right)$$

$$\geq \sum_{d=1}^{2s}(-1)^{d-1}\sum_{1\leq m_1<\cdots<m_d\leq q}\mathsf{P}\left(|\boldsymbol{N}_d|_{\min}\geq y_{l,g}+\epsilon_n\{\log(p_lp_g)\}^{-1/2}\right)-o(1). \quad (S5)$$

Then it suffices to prove that, for any fixed integer $d \geq 1$ and any $t \in \mathbb{R}$,

$$\sum_{1\leq m_1<\cdots<m_d\leq q}\mathsf{P}\left(|\boldsymbol{N}_d|_{\min}\geq y_{l,g}\pm\epsilon_n\{\log(p_lp_g)\}^{-1/2}\right)=\frac{1}{d!}\left(-\pi^{-1/2}e^{-t/2}\right)^d\{1+o(1)\}. \quad (S6)$$

When $d = 1$, we have that,

$$\mathsf{P}\left(|\boldsymbol{N}_1|_{\min}\geq y_{l,g}\pm\epsilon_n\{\log(p_lq_l)\}^{-1/2}\right)=\{1+o(1)\}\frac{(p_lp_g)^{-1}}{\sqrt{\pi}}\exp(-t/2).$$

This yields (S6).

When $d \geq 2$, noting that for any $(i,j) \in \mathcal{A}^{(l,g)} \setminus B_0$ and $(i',j') \in \mathcal{A}^{(l,g)} \setminus B_0$, we have $\mathsf{Cov}(X_iY_j, X_{i'}Y_{j'}) = \mathsf{E}(X_iY_jX_{i'}Y_{j'}) + O((\log p_{l,g})^{-2-2\alpha_0})$. Define a graph $G_{abcd} = (V_{abcd}, E_{abcd})$, where $V_{abcd} = \{a,b,c,d\}$ is the set of vertices, and E_{abcd} is the set of edges. There is an edge between $i \neq j \in \{a,b,c,d\}$ if and only if $|\sigma_{i,j}| \geq (\log p_{l,g})^{-1-\alpha_0}$. We say G_{abcd} is a three-vertice graph (3-G) if the number of different vertices in V_{abcd} is 3. Similarly, G_{abcd} is a four-vertice graph (4-G) if the number of different vertices in V_{abcd} is 4. A vertex in G_{abcd} is said to be isolated if there is no edge connected to it. Note that for any $1 \leq s_1 \neq s_2 \leq q$, $G_{i_{s_1}j_{s_1}i_{s_2}j_{s_2}}$ is a 3-G or 4-G. We say a graph $\mathcal{G}$ satisfies (S7) if the following statement holds:

If $\mathcal{G}$ is a 4-G, then there is at least one isolated vertex in $\mathcal{G}$;
Otherwise $\mathcal{G}$ is a 3-G and $E_{i_{s_1}j_{s_1}i_{s_2}j_{s_2}} = \emptyset$. (S7)

Note that, under the null, we have $\mathsf{E}(X_{i_{s_1}}Y_{j_{s_1}}) = \mathsf{E}(X_{i_{s_2}}Y_{j_{s_2}}) = \mathsf{E}(X_{i_{s_1}}Y_{j_{s_2}}) = \mathsf{E}(X_{i_{s_2}}Y_{j_{s_1}}) = 0$. That is, there is no edge between X_{i_m} and Y_{j_m}. Thus there are at most 2 edges if it is 4-G, and there are at most 1 edge if it is 3-G. For any $G_{i_{s_1}j_{s_1}i_{s_2}j_{s_2}}$ satisfying (S7), by Assumption (A3),

$$|\mathsf{E}(X_{i_{s_1}}Y_{j_{s_1}}X_{i_{s_2}}Y_{j_{s_2}})| = O((\log p)^{-1-\alpha_0}), \quad (S8)$$

where $O(1)$ is uniformly for $i_{s_1}, j_{s_1}, i_{s_2}, j_{s_2}$. Denote by

$$\begin{aligned}
\mathcal{F} &= \{1 \leq m_1 < \cdots < m_d \leq q\},\\
\mathcal{F}_0 &= \{1 \leq m_1 < \cdots < m_d \leq q : \text{for some } s_1, s_2 \in \{m_1,\cdots,m_d\} \text{ with } s_1 \neq s_2\\
&\quad\ \mathcal{G} := G_{i_{s_1}j_{s_1}i_{s_2}j_{s_2}} \text{ does not satisfy (S7)}\},\\
\mathcal{F}_0^c &= \{1 \leq m_1 < \cdots < m_d \leq q : \text{for any } s_1, s_2 \in \{m_1,\cdots,m_d\} \text{ and } s_1 \neq s_2,\\
&\quad\ \mathcal{G} \text{ satisfies (S7)}\}.
\end{aligned}$$

It is easy to see that $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_0^c$. For any subset S of $\{m_1,\ldots,m_d\}$, we say that S satisfies (S9) if the following statement is true:

For any $s_1 \neq s_2 \in S$, $G_{i_{s_1}j_{s_1}i_{s_2}j_{s_2}}$ satisfies (S7). (S9)

For $2 \le t \le d$, let

$$\begin{aligned}\mathcal{F}_{0t} &= \{1 \le m_1 < \cdots < m_d \le q : \text{the largest cardinality of } S \text{ is } t, \text{ where} \\ &\quad S \text{ is any subset of } \{m_1 < \cdots < m_d\} \text{ satisfies (S9)}\}, \\ \mathcal{F}_{01} &= \{1 \le m_1 < \cdots < m_d \le q : \text{for any } s_1, s_2 \in \{m_1, \cdots, m_d\} \text{ with } s_1 \ne s_2 \\ &\quad \mathcal{G} := G_{i_{s_1} j_{s_1} i_{s_2} j_{s_2}} \text{ does not satisfy (S7)}\}.\end{aligned}$$

Then we have $\mathcal{F}_0^c = \mathcal{F}_{0d}$ and $\mathcal{F}_0 = \cup_{l=1}^{d-1} \mathcal{F}_{0t}$. It is easy to show that $|\mathcal{F}_{0t}| \le C_d q^{t+2\gamma(d-t)}$, and $|\mathcal{F}_0^c| = (1+o(1))\binom{q}{d}$. Thus it suffices to prove the following two statements hold:

$$\sum_{\mathcal{F}_0^c} \mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \ge y_{l,g} \pm \varepsilon_n \{\log(p_l p_g)\}^{-1/2}\right) = \{1+o(1)\} \frac{1}{d!} \left(-\pi^{-1/2} e^{-t/2}\right)^d, \qquad \text{(S10)}$$

and

$$\sum_{\mathcal{F}_0} \mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \ge y_{l,g} \pm \varepsilon_n \{\log(p_l p_g)\}^{-1/2}\right) = o(1). \qquad \text{(S11)}$$

We first prove (S11). For $1 \le a \ne b \le q$, define the indicator function,

$$d\{(i_a, j_a), (i_b, j_b)\} = 1 \text{ if } G_{i_a j_a i_b j_b} \text{ does not satisfy (S7), and 0 otherwise.}$$

We further divide $\mathcal{F}_{0t}$ in the following way. Let $(m_1, \ldots, m_d) \in \mathcal{F}_{0t}$ and let $S_\star \subset (m_1, \ldots, m_d)$ be the largest cardinality subset satisfying (S9). If there is more than one subset that attains the largest cardinality, one can choose any one of them. Define

$$\begin{aligned}\mathcal{F}_{0t1} &= \{(m_1, \ldots, m_d) \in \mathcal{F}_{0t} : \text{there exists an } a \notin S_\star \text{ such that for some } b_1, b_2 \in S_\star \\ &\quad \text{with } b_1 \ne b_2,\ d((i_a, j_a), (i_{b_1}, j_{b_1})) = 1,\ d((i_a, j_a), (i_{b_2}, j_{b_2})) = 1\}, \\ \mathcal{F}_{0t2} &= \mathcal{F}_{0t} \setminus \mathcal{F}_{0t1}.\end{aligned}$$

Note that $\mathcal{F}_{011} = \emptyset$, and $\mathcal{F}_{012} = \mathcal{F}_{01}$. Recall that d is fixed and $t \le d-1$. We then obtain that $|\mathcal{F}_{0t1}| \le C q^{t-1+2\gamma(d-t+1)}$, and $|\mathcal{F}_{0t2}| \le C_d q^{t+2\gamma(d-t)}$. Write $S_\star = (b_1, \ldots, b_t)$ and $x_{l,g} = y_{l,g} \pm \epsilon_n (\log(p_l p_g))^{-1/2}$. For any $(m_1, \ldots, m_d) \in \mathcal{F}_{0t}$,

$$\begin{aligned}\mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \ge y_{l,g} \pm \epsilon_n \{\log(p_l p_g)\}^{-1/2}\right) &\le \mathsf{P}\left(|N_{b_1}| \ge x_{l,g}, \ldots, |N_{b_t}| \ge x_{l,g}\right) \\ &= \frac{1}{(2\pi)^{l/2} |\boldsymbol{U}_t|^{1/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_{l,g}} \exp\left(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{U}_t^{-1} \boldsymbol{y}\right) d\boldsymbol{y},\end{aligned}$$

where $\boldsymbol{U}_t$ is the covariance matrix of $(N_{b_1}, \ldots, N_{b_t})$, $\boldsymbol{y} = (y_1, \ldots, y_t)$. By (S8), we have $\|\boldsymbol{U}_t - \boldsymbol{I}_t\|_2 = O((\log p_{l,g})^{-1-\alpha_0})$. Let $|\boldsymbol{y}|_{\max} = \max_{1 \le i \le t} |y_i|$. Then

$$\begin{aligned}&\frac{1}{(2\pi)^{t/2} |\boldsymbol{U}_t|^{1/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_{l,g}} \exp\left(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{U}_t^{-1} \boldsymbol{y}\right) d\boldsymbol{y} \\ = &\frac{1}{(2\pi)^{t/2} |\boldsymbol{U}_t|^{1/2}} \int_{|\boldsymbol{y}|_{\min} \ge x_{l,g}, |\boldsymbol{y}|_{\max} \le (\log p_{l,g})^{1/2+\alpha_0/4}} \exp\left(-\frac{1}{2} \boldsymbol{y}^T \boldsymbol{U}_t^{-1} \boldsymbol{y}\right) d\boldsymbol{y}\end{aligned}$$

$$
\begin{aligned}
& + O\left(\exp\left\{-(\log p_{l,g})^{1+\alpha_0/2}/4\right\}\right) \\
= \; & \frac{1+O((\log p_{l,g})^{-\alpha_0/2})}{(2\pi)^{t/2}} \int_{|\boldsymbol{y}|_{\min} \geq x_{l,g}, |\boldsymbol{y}|_{\max} \leq (\log p_{l,g})^{1/2+\alpha_0/4}} \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y} \\
& + O\left(\exp\left\{-(\log p_{l,g})^{1+\alpha_0/2}/4\right\}\right) \\
= \; & \frac{1+O((\log p_{l,g})^{-\alpha_0/2})}{(2\pi)^{t/2}} \int_{|\boldsymbol{y}|_{\min} \geq x_{l,g}} \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y} \\
& + \; + \; O\left(\exp\left\{-(\log p_{l,g})^{1+\alpha_0/2}/4\right\}\right) \\
= \; & O\left((p_l p_g)^{-t}\right). \qquad \text{(S12)}
\end{aligned}
$$

Note that γ is sufficiently small. Thus it yields that

$$
\sum_{\mathcal{F}_{0t1}} \mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \geq y_{l,g} \pm \epsilon_n\{\log(p_l p_g)\}^{-1/2}\right) \leq C(p_l p_g)^{-1+2\gamma(d-t+1)} = o(1). \qquad \text{(S13)}
$$

For $(m_1, \ldots, m_d) \in \mathcal{F}_{0t2}$, let $\bar{a} = \min\{a : a \in (m_1, \ldots, m_d), a \notin S_\star\}$. Without loss of generality and for notation briefness, we can assume that $d((i_{\bar{a}}, j_{\bar{a}}), (i_{b_1}, j_{b_1})) = 1$. Then we have

$$
\sum_{\mathcal{F}_{0t2}} \mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \geq y_{l,g} \pm \epsilon_n\{\log(p_l p_g)\}^{-1/2}\right) \leq \sum_{\mathcal{F}_{0t2}} \mathsf{P}\left(|N_{\bar{a}}| \geq x_{l,g}, |N_{b_1}| \geq x_{l,g}, \ldots, |N_{b_t}| \geq x_{l,g}\right),
$$

Because $(m_1, \ldots, m_d) \in \mathcal{F}_{0t2}$, by Assumption (A3), we can show that $\mathsf{Cov}(N_{\bar{a}}, N_{b_j}) = O((\log p_{l,g})^{-1-\alpha_0})$ for $2 \leq j \leq l$. Recall that $S_\star = (b_1, \ldots, b_t)$. We have $\mathsf{Cov}(N_{b_i}, N_{b_j}) = O((\log p_{l,g})^{-1-\alpha_0})$ for $1 \leq i \neq j \leq l$. Let $\boldsymbol{V}_t$ be the covariance matrix of $(N_{\bar{a}}, N_{b_1}, \ldots, N_{b_t})$. It follows that $\|\boldsymbol{V}_t - \bar{\boldsymbol{V}}_t\|_2 = O((\log p_{l,g})^{-1-\alpha_0})$, where $\bar{\boldsymbol{V}}_t = \text{diag}(\boldsymbol{D}, \boldsymbol{I}_{l-1})$ and $\boldsymbol{D}$ is the covariance matrix of $(N_{\bar{a}}, N_{b_1})$. We say the graph $G_{i_{\bar{a}} j_{\bar{a}} i_{b_1} j_{b_1}}$ is aG-bE if $G_{i_{\bar{a}} j_{\bar{a}} i_{b_1} j_{b_1}}$ is a-G and there are b edges in $E_{i_{\bar{a}} j_{\bar{a}} i_{b_1} j_{b_1}}$ for $a = 3, 4$ and $b = 0, 1, 2, 3, 4$. Under the null, we have that $G_{i_{\bar{a}} j_{\bar{a}} i_{b_1} j_{b_1}}$ can only be 3G-1E or 4G-2E. Using the similar argument as that in (S12), we obtain that

$$
\begin{aligned}
& \sum_{\mathcal{F}_{0t2}} \mathsf{P}\left(|N_{\bar{a}}| \geq x_{l,g}, |N_{b_1}| \geq x_{l,g}, \ldots, |N_{b_t}| \geq x_{l,g}\right) \\
\leq \; & C \sum_{\mathcal{F}_{0t2}} \left[\mathsf{P}\left(|N_{\bar{a}}| \geq x_{l,g}, |N_{b_1}| \geq x_{l,g}\right) \times (p_l p_g)^{-l+1} + \exp\left\{-(\log p_{l,g})^{1+\alpha_0/2}/4\right\}\right] \\
\leq \; & C \sum_{\mathcal{F}_{0t2}} \left[(p_l p_g)^{-1-(1-r)/(3+r)} \times (p_l p_g)^{-l+1} + \exp\left\{-(\log p_{l,g})^{1+\alpha_0/2}/4\right\}\right], \qquad \text{(S14)}
\end{aligned}
$$

where the last inequality follows from Lemma 2. Hence it follows from (S14) that

$$
\sum_{\mathcal{F}_{0t2}} \mathsf{P}\left(|N_{\bar{a}}| \geq x_{l,g}, |N_{b_1}| \geq x_{l,g}, \ldots, |N_{b_t}| \geq x_{l,g}\right) \leq C(p_l p_g)^{-(1-r)/(3+r)+(d-t)\gamma} = o(1). \qquad \text{(S15)}
$$

By (S13) and (S15), (S11) is proved. Next we prove (S10). By (S8), we have $\|\mathsf{Cov}(\boldsymbol{N}_d) - \boldsymbol{I}_d\|_2 = O((\log p_{l,g})^{-1-\alpha_0})$ uniformly for $(m_1, \ldots, m_d) \in \mathcal{F}_0^c$. Then following the same argument as in (S12), we obtain

$$
\mathsf{P}\left(|\boldsymbol{N}_d|_{\min} \geq y_{l,g} \pm \varepsilon_n\{\log(p_l p_g)\}^{-1/2}\right) = \{1+o(1)\}\left(2\pi^{-1/2} e^{-t/2}\right)^d (p_l p_g)^{-d}
$$

uniformly for $(m_1, \dots, m_d) \in \mathcal{F}_0^c$. Then (S10) holds by the fact $|\mathcal{F}_0^c| = \{1 + o(1)\}\binom{q}{d}$. Thus (S6) holds, and Theorem 1 follows. $\square$

A4 Proof of Theorem 2

The lower bound result can be directly obtained by Theorem 4 of Cai et al. (2013). Next we prove the upper bound result. Define

$$M_{l,g}^1 = \max_{i\in\mathcal{S}_l, j\in\mathcal{S}_g} \frac{(\hat{\sigma}_{i,j} - \sigma_{i,j})^2}{\hat{\theta}_{i,j}/n}.$$

By Lemma 4 and the proof of Theorem 1,

$$\mathsf{P}\left(M_n^1 \le 2\log(p_l p_g) - 1/2 \log\log(p_l p_g)\right) \to 1$$

as $n, p_{l,g} \to \infty$. Because

$$\max_{i\in\mathcal{S}_l, j\in\mathcal{S}_g} \frac{\sigma_{i,j}^2}{\hat{\theta}_{i,j}/n} \le 2M_{l,g}^1 + 2M_{l,g}, \quad \text{and} \quad \max_{i\in\mathcal{S}_l, j\in\mathcal{S}_g} \frac{\sigma_{i,j}^2}{\theta_{i,j}/n} \ge 16\log(p_{l,g}),$$

by Lemma 3, we have

$$\mathsf{P}\left(M_{l,g} \ge q_\alpha + 2\log(p_l p_g) - \log\log(p_l p_g)\right) \to 1$$

as $n, p_{l,g} \to \infty$. Then Theorem 2 is proved. $\square$

A5 Proof of Theorem 3

We first show that if $\hat{t}$ does not exist in the range $\left[0, \{2\log(L^2) - 2\log\log L\}^{1/2}\right]$, the thresholding of $\hat{t}$ at $\{2\log(L^2)\}^{1/2}$ leads to no false rejection with probability tending to 1. Thus we focus on the event $\mathcal{A} = \left\{\hat{t} \text{ exists in the range } [0, \{2\log(L^2) - 2\log\log L\}^{1/2}]\right\}$, and prove the FDP results by dividing the pairs of submatrices into small subsets. The key is to show that the weakly correlated pairs play the dominating role. We then show that under the condition on $|\mathcal{L}_\rho|$, the event $\mathcal{A}$ holds with probability tending to 1, and hence the FDR and FDP converge to $\alpha|\mathcal{H}_0|/L^2$ asymptotically.

Note that

$$\mathsf{P}\left\{\sum_{(l,g)\in\mathcal{H}_0} I\left(N_{l,g} \ge \sqrt{2\log L^2}\right) \ge 1\right\} \le |\mathcal{H}_0| \max_{(l,g)\in\mathcal{H}_0} \mathsf{P}\left(N_{l,g} \ge \sqrt{2\log L^2}\right).$$

By the proof of Theorem 1, we have

$$\mathsf{P}\left(\sum_{(l,g)\in\mathcal{H}_0} I\left(N_{l,g} \ge \sqrt{2\log L^2}\right) \ge 1\right) \quad \le \quad |\mathcal{H}_0| \max_{(l,g)\in\mathcal{H}_0} \mathsf{P}\left(N_{l,g} \ge \sqrt{2\log L^2}\right)$$

$$\leq \; |\mathcal{H}_0| G\left(\sqrt{2\log L^2}\right)\{1+o(1)\} = o(1),$$

where $G(t) = 1 - \Phi(t)$, which shows that if $N_{l,g}$ are thresholded at level $\sqrt{2\log L^2}$, the probability of false rejection is tending to 0 asymptotically. For that reason, we focus on the event $\{\hat{t}$ exists in the range $[0, \sqrt{2\log(L^2) - 2\log\log L}]\}$.

Note that, by the definition of $\hat{t}$, for any $t < \hat{t}$, we have

$$\frac{G(t)L^2}{\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq t, 1\}} > \alpha.$$

Because $\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq \hat{t}\}, 1\} \leq \max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq t\}, 1\}$, we have

$$\frac{G(t)L^2}{\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq \hat{t}\}, 1\}} > \alpha.$$

Thus, by letting $t \to \hat{t}$,

$$\frac{G(\hat{t})L^2}{\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq \hat{t}\}, 1\}} \geq \alpha.$$

On the other hand, based on the definition of $\hat{t}$, there exists a sequence $\{t_l\}$ with $t_l \geq \hat{t}$ and $t_l \to \hat{t}$, such that

$$\frac{G(t_l)L^2}{\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq t_l\}, 1\}} \leq \alpha.$$

Thus we have $\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq t_l\}, 1\} \leq \max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq \hat{t}\}, 1\}$, which implies

$$\frac{G(t_l)L^2}{\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq \hat{t}\}, 1\}} \leq \alpha.$$

Letting $t_l \to \hat{t}$, we have

$$\frac{G(\hat{t})L^2}{\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq \hat{t}\}, 1\}} \leq \alpha.$$

Thus by focusing on the event $\{\hat{t}$ exists in the range $[0, \sqrt{2\log(L^2) - 2\log\log L}]\}$, we have

$$\frac{G(\hat{t})L^2}{\max\{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g} \geq \hat{t}\}, 1\}} = \alpha.$$

Set $t_L = \sqrt{2\log(L^2) - 2\log\log L}$. Then it suffices to show that

$$\sup_{0\leq t\leq t_L} \left| \frac{\sum_{(l,g)\in\mathcal{H}_0} I\{N_{l,g} \geq t\} - |\mathcal{H}_0| G(t)}{L^2 G(t)} \right| \to 0$$

in probability. Let $0 \le t_0 < t_1 < \cdots < t_b = t_L$ such that $t_\iota - t_{\iota-1} = v_L$ for $1 \le \iota \le b-1$ and $t_b - t_{b-1} \le v_L$, where $v_L = 1/\sqrt{\log(L^2)(\log_4 L)}$. Then we have $b \sim t_L/v_L$. For any t such that $t_{\iota-1} \le t \le t_\iota$, we have

$$\frac{\sum_{(l,g)\in\mathcal{H}_0} I(N_{l,g} \ge t_\iota)}{|\mathcal{H}_0|G(t_\iota)} \frac{G(t_\iota)}{G(t_{\iota-1})} \le \frac{\sum_{(l,g)\in\mathcal{H}_0} I(N_{l,g} \ge t)}{|\mathcal{H}_0|G(t)} \le \frac{\sum_{(l,g)\in\mathcal{H}_0} I(N_{l,g} \ge t_{\iota-1})}{|\mathcal{H}_0|G(t_{\iota-1})} \frac{G(t_{\iota-1})}{G(t_\iota)}.$$

Thus it suffices to prove

$$\max_{0\le\iota\le b} \left| \frac{\sum_{(l,g)\in\mathcal{H}_0}[I(N_{l,g} \ge t_\iota) - G(t_\iota)]}{|\mathcal{H}_0|G(t_\iota)} \right| \to 0$$

in probability. Note that

$$\begin{aligned}
&\mathsf{P}\left(\max_{0\le\iota\le b} \left| \frac{\sum_{(l,g)\in\mathcal{H}_0}[I(N_{l,g} \ge t_\iota) - G(t_\iota)]}{|\mathcal{H}_0|G(t_\iota)} \right| \ge \epsilon \right) \\
&\le \sum_{\iota=1}^{b} \mathsf{P}\left(\left| \frac{\sum_{(l,g)\in\mathcal{H}_0}\{I(N_{l,g} \ge t_\iota) - G(t_\iota)\}}{|\mathcal{H}_0|G(t_\iota)} \right| \ge \epsilon \right) \\
&\le \frac{1}{v_L}\int_0^{t_L} \mathsf{P}\left(\left| \frac{\sum_{(l,g)\in\mathcal{H}_0}\{I(N_{l,g} \ge t) - G(t)\}}{|\mathcal{H}_0|G(t)} \right| \ge \epsilon \right) dt \\
&\quad + \sum_{\iota=b-1}^{b} \mathsf{P}\left(\left| \frac{\sum_{(l,g)\in\mathcal{H}_0}\{I(N_{l,g} \ge t_\iota) - G(t_\iota)\}}{|\mathcal{H}_0|G(t_\iota)} \right| \ge \epsilon \right).
\end{aligned}$$

Recall that $N_{l,g} = \Phi^{-1}\{1 - F^*(M_{l,g})\}$, where F^* is the corrected cumulative distribution function. Because $p \le cn^a$ for some $c > 0$ and $a > 0$, by the proof of (S3) – (S6) in Theorem 1, we have

$$\mathsf{P}(N_{l,g} \ge t) = \mathsf{P}\left(M_{l,g} \ge F^{*-1}\{\Phi(t)\}\right) = \{1 + o(1)\}G(t).$$

Thus it suffices to prove the following statements are true for any $\epsilon > 0$.

$$\int_0^{t_L} \mathsf{P}\left(\left| \frac{\sum_{(l,g)\in\mathcal{H}_0}\{I(N_{l,g} \ge t) - \mathsf{P}(N_{l,g} \ge t)\}}{L^2 G(t)} \right| \ge \epsilon \right) dt = o(v_L) \tag{S16}$$

and

$$\sup_{0\le t\le t_L} \mathsf{P}\left(\left| \frac{\sum_{(l,g)\in\mathcal{H}_0}\{I(N_{l,g} \ge t) - \mathsf{P}(N_{l,g} \ge t)\}}{L^2 G(t)} \right| \ge \epsilon \right) = o(1). \tag{S17}$$

We next prove (S16), and the proof of (S17) is similar. Note that the variance can be calculated as follows

$$\begin{aligned}
&\mathsf{E}\left[\frac{\sum_{(l,g)\in\mathcal{H}_0}\{I(N_{l,g} \ge t) - \mathsf{P}(N_{l,g} \ge t)\}}{(L^2)G(t)} \right]^2 \\
= &\frac{\sum_{(l,g),(l',g')\in\mathcal{H}_0}\{\mathsf{P}(N_{l,g} \ge t, N_{l',g'} \ge t) - \mathsf{P}(N_{l,g} \ge t)\mathsf{P}(N_{l',g'} \ge t)\}}{(L^2)^2 G^2(t)}.
\end{aligned}$$

In order to estimate the correlations of $N_{l,g}$ and $N_{l',g'}$, we first split the set $\mathcal{H}_0$ into three subsets. Similarly as defined in Theorem 1, let $G_{abcd} = (V_{abcd}, E_{abcd})$ denote a graph, where $V_{abcd} = \{a, b, c, d\}$ denotes the set of vertices and E_{abcd} the set of edges. There is an edge between $i \neq j \in \{a, b, c, d\}$ if and only if $|\sigma_{i,j}| \geq (\log L)^{-2-\gamma}$. Note that for any $i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y$ and $i' \in \mathcal{S}_{l'}^X, j' \in \mathcal{S}_{g'}^Y$, $(l, g) \neq (l', g')$, $G_{iji'j'}$ is a 3-G or 4-G. We say a graph $\mathcal{G} = G_{iji'j'}$ satisfies (S18) if the following statement holds:

$$\begin{array}{l}\text{If } \mathcal{G} \text{ is a 4-G, then there is at least one isolated vertex in } \mathcal{G};\\ \text{Otherwise } \mathcal{G} \text{ is a 3-G and } E_{iji'j'} = \emptyset.\end{array} \tag{S18}$$

Then similarly as in Theorem 1, for any $G_{iji'j'}$ satisfying (S18), uniformly for i, j, i', j', $|\mathsf{E}(X_i Y_j X_{i'} Y_{j'})| = O\{(\log L)^{-2-\gamma}\}$. Based on the definition of (S18), we further divide $\mathcal{H}_0$ into three subsets

$$\begin{aligned}
\mathcal{H}_{01} &= \{(l,g),(l',g') \in \mathcal{H}_0, (l,g) = (l',g')\},\\
\mathcal{H}_{02} &= \{(l,g),(l',g') \in \mathcal{H}_0, (l,g) \neq (l',g'), \forall i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y, i' \in \mathcal{S}_{l'}^X, j' \in \mathcal{S}_{g'}^Y,\\
&\quad\ G_{iji'j'} \text{ satisfies (S18)}\},\\
\mathcal{H}_{03} &= \{(l,g),(l',g') \in \mathcal{H}_0, (l,g) \neq (l',g'), \exists i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y, i' \in \mathcal{S}_{l'}^X, j' \in \mathcal{S}_{g'}^Y,\\
&\quad\ G_{iji'j'} \text{ does not satisfy (S18)}\}.
\end{aligned}$$

For the subset $\mathcal{H}_{01}$, the cardinality is small, and we have

$$\frac{\sum_{(l,g),(l',g')\in\mathcal{H}_{01}} \{\mathsf{P}(N_{l,g} \geq t, N_{l',g'} \geq t) - \mathsf{P}(N_{l,g} \geq t)\mathsf{P}(N_{l',g'} \geq t)\}}{(L^2)^2 G^2(t)} \leq \frac{C}{L^2 G(t)}. \tag{S19}$$

Recall that

$$\Lambda_l(\gamma) = \{g : 1 \leq g \leq L, \exists i \in \mathcal{S}_l^X \cup \mathcal{S}_l^Y, j \in \mathcal{S}_g^X \cup \mathcal{S}_g^Y, \text{ s.t. } |\sigma_{i,j}| \geq (\log L)^{-2-\gamma}\}.$$

and $\max_{1\leq l\leq L} |\Lambda_l(\gamma)| = o(L^\nu)$ for any $\nu > 0$. Thus we have $|\mathcal{H}_{03}| = O((L^2)^{1+\nu})$. Note that uniformly for $(l, g), (l', g') \in \mathcal{H}_{03}$, by Assumption (A1), we have $\mathsf{Corr}(N_{l,g}, N_{l',g'}) \leq r' < 1$, for some $r < r' < 1$. Thus, by truncations and the application of Lemma 5 to obtain normal approximations for $N_{l,g}$ and $N_{l',g'}$ similarly as in the proofs of Theorem 1, we have

$$\begin{aligned}
&\frac{\sum_{(l,g),(l',g')\in\mathcal{H}_{03}} \{\mathsf{P}(N_{l,g} \geq t, N_{l',g'} \geq t) - \mathsf{P}(N_{l,g} \geq t)\mathsf{P}(N_{l',g'} \geq t)\}}{(L^2)^2 G^2(t)}\\
\leq\ & C\frac{(L^2)^{1+\nu} t^{-2} \exp(-t^2/(1+r'))}{(L^2)^2 G(t)} \leq \frac{C}{(L^2)^{1-\nu}\{G(t)\}^{2r'/(1+r')}}.
\end{aligned} \tag{S20}$$

It remains to consider subset $\mathcal{H}_{02}$, in which $N_{l,g}$ and $N_{l',g'}$ are weakly correlated with each other. By applying Lemma 5, it is straightforward to check that

$$\max_{(l,g),(l',g')\in\mathcal{H}_{02}} \mathsf{P}\left(N_{l,g} \geq t, N_{l',g'} \geq t\right) = [1 + O\{(\log L)^{-1-\gamma}\}]G^2(t).$$

Thus, we have

$$\frac{\sum_{(l,g),(l',g')\in\mathcal{H}_{02}}\{\mathsf{P}(N_{l,g}\geq t, N_{l',g'}\geq t)-\mathsf{P}(N_{l,g}\geq t)\mathsf{P}(N_{l',g'}\geq t)\}}{(L^2)^2G^2(t)}$$
$$= O((\log L)^{-1-\gamma}). \tag{S21}$$

Combining (S19), (S20) and (S21), we have

$$\int_0^{t_L}\left[\frac{C}{(L^2)G(t)}+\frac{C}{(L^2)^{1-\nu}\{G(t)\}^{2r'/(1+r')}}+C(\log L)^{-1-\gamma}\right]dt = o(v_L).$$

Thus (S16) is proved. Accordingly, we have

$$\limsup_{n,L,(p_{l,g})_{l,g=1}^L\to\infty} \text{FDR}(\hat{t}) \leq \alpha|\mathcal{H}_0|/L^2,$$

and for any $\epsilon > 0$,

$$\lim_{n,L,(p_{l,g})_{l,g=1}^L\to\infty} \mathsf{P}\left(\text{FDP}(\hat{t}) \leq \alpha|\mathcal{H}_0|/L^2+\epsilon\right) = 1.$$

Finally we prove the FDR and FDP results under the condition on $|\mathcal{L}_\rho|$. It is easy to check that

$$\sum_{(l,g)\in\mathcal{H}} I\left\{N_{l,g}\geq\sqrt{2\log(L^2)}\right\} \geq \left(\frac{1}{\sqrt{8\pi}\alpha}+\delta\right)\sqrt{\log L},$$

with probability going to 1. Hence with probability going to one, we have

$$\frac{L^2}{\sum_{(l,g)\in\mathcal{H}} I\{N_{l,g}\geq\sqrt{2\log(L^2)}\}} \leq L^2\left(\frac{1}{\sqrt{8\pi}\alpha}+\delta\right)^{-1}(\log L)^{-1/2}.$$

Recall that $t_L = \sqrt{2\log(L^2)-2\log\log L}$. Because $1-\Phi(t_L)\sim\left(\sqrt{2\pi}t_L\right)^{-1}\exp(-t_L^2/2)$, we have $\mathsf{P}(0\leq\hat{t}\leq t_L)\to 1$ according to the definition of $\hat{t}$ in Algorithm 1 in Section 3.1. Namely, we have $\mathsf{P}(\hat{t}$ exists in $[0,t_L])\to 1$. Henceforth,

$$\lim_{n,L,(p_{l,g})_{l,g=1}^L\to\infty}\frac{\text{FDR}(\hat{t})}{\alpha|\mathcal{H}_0|/L^2} = 1,$$

and

$$\frac{\text{FDP}(\hat{t})}{\alpha|\mathcal{H}_0|/L^2}\to 1 \text{ in probability, as } n, L, (p_{l,g})_{l,g=1}^L\to\infty.$$

Then Theorem 3 is proved. □

References

Berman, S. M. (1962). A law of large numbers for the maximum in a stationary gaussian sequence. *Ann. Math. Statist.*, 33:93–97.

Cai, T. T., Liu, W., and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.*, 108(501):265–277.

Xie, J. and Kang, J. (2017). High-dimensional tests for functional networks of brain anatomic regions. *Journal of Multivariate Analysis*, 156(Supplement C):70 – 88.

Zaïtsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions. *Probab. Theory Rel.*, 74(4):535–566.