Biometrika (2018), **1**, 1, *pp.* 1–17
Printed in Great Britain Advance Access publication on 21 Sep 2018

Supplemental Materials for "Differential Markov Random Field Analysis with an Application to Detecting Differential Microbial Community Networks"

BY T. T. CAI

Department of Statistics, University of Pennsylvania, 3720 Walnut Street, Philadelphia, Pennsylvania 19104, U.S.A.
tcai@wharton.upenn.edu

H. LI, J. MA

Department of Biostatistics, University of Pennsylvania, 423 Guardian Drive, Philadelphia, Pennsylvania 19104, U.S.A.
hongzhe@upenn.edu jinma@upenn.edu

AND Y. XIA

Department of Statistics, Fudan University, Shanghai 200433, China
xiayin@fudan.edu.cn

A. SUPPLEMENTARY MATERIAL

A·1. *Auxiliary lemmas*

We first prove a few technical lemmas. For completeness, we repeat the statement of Lemma 1 here.

LEMMA A1. *Under Assumption* 2*,*

$$|\mathrm{cor}(v^o_{r,t,k}\varepsilon_{r,k}, v^o_{r',t,k}\varepsilon_{r',k})| \leq \frac{4C_0}{c^*}\{|\sinh(2\theta_{r,r',k})| + |\sinh(2\theta_{r,t,k})\sinh(2\theta_{r',t,k})|\}, \tag{A1}$$

$$|\mathrm{cor}(v^o_{r,t,k}\varepsilon_{r,k}, v^o_{r,t',k}\varepsilon_{r,k})| \leq \frac{4C_0}{c^*}\{|\sinh(2\theta_{t,t',k})| + |\sinh(2\theta_{r,t,k})\sinh(2\theta_{r,t',k})|\}, \tag{A2}$$

$$|\mathrm{cor}(v^o_{r,t,k}\varepsilon_{r,k}, v^o_{r',t',k}\varepsilon_{r',k})| \leq \frac{C_1}{c^*}\left\{\sum_{a=r',t'}|\sinh(2\theta_{r,a,k})|\right\}\left\{\sum_{b=r',t'}|\sinh(2\theta_{t,b,k})|\right\}, \tag{A3}$$

for $r \neq r'$ *and* $t \neq t'$*, where* $C_0 = 2\cosh^2(C_w)$ *and*

$$C_1 = \max[4\sinh(2C_w)(C_0-1) + 2(C_0-1)^2\{\sinh(2C_w) + 2C_0\}, 10(C_0-1)^2].$$

Proof. We show that the correlation between $v^o_{r,t,k}\varepsilon_{r,k}$ and $v^o_{r,t',k}\varepsilon_{r,k}$ for $t \neq t'$ is bounded as in (A2). The derivations of (A1) and (A3) can be done similarly. Without loss of generality, let $k = 1$.

Recall the residual $\varepsilon_{r,1} = X_r - \dot{f}(X_{-r}\theta_{r,-r,1})$ satisfies $E(\varepsilon_{r,1} \mid X_{-r}) = 0$ and $E(\varepsilon^2_{r,1} \mid X_{-r}) = \ddot{f}(X_{-r}\theta_{r,-r,1})$. By definition, $E(v^o_{r,t,1}\varepsilon_{r,1}) = 0$ and $E\{(v^o_{r,t,1}\varepsilon_{r,1})^2\} = E\{(v^o_{r,t,1})^2\ddot{f}(X_{-r}\theta_{r,-r,1})\} = F_{r,t,1}/4$. Therefore $\mathrm{var}(v^o_{r,t,1}\varepsilon_{r,1})$ is bounded from above by the fact that $|v^o_{r,t,1}| \leq 1, |\ddot{f}(u)| \leq 1$ and bounded from below by Assumption 2. It suffices to show that $\mathrm{cov}(v^o_{r,t,1}\varepsilon_{r,1}, v^o_{r,t',1}\varepsilon_{r,1})$ is bounded. To this end, we introduce several simplified notation: $\theta_{r,t,1} = a, \theta_{r,t',1} = b, \theta_{t,t',1} = c$, and for $l \in \{r, t, t'\}$,

$$A_l = \sum_{j:j \neq \{r,t,t'\}} \theta_{l,j,1} X_j.$$

Let $\theta_{r,-t,1} = \{\theta_{r,j,1} : j \neq r, t\}$ denote the $(p-2)$-dimensional subvector. The following facts are frequently used:

$$\dot{f}(u) = \frac{\sinh(u)}{\cosh(u)}, \quad \ddot{f}(u) = \frac{1}{\cosh^2(u)}.$$

By the conditional distribution of X_t given $X_{-\{r,t\}}$, the score vector $v^o_{r,t,1}$ can be rewritten as

$$\begin{aligned} v^o_{r,t,1} =& \frac{X_t + 1}{2} - \frac{E\{\ddot{f}(X_{-r}\theta_{r,-r,1})(X_t+1)/2 \mid X_{-\{r,t\}}\}}{E\{\ddot{f}(X_{-r}\theta_{r,-r,1}) \mid X_{-\{r,t\}}\}} \\ =& \frac{X_t + 1}{2} - \frac{\ddot{f}(a + X_{-\{r,t\}}\theta_{r,-t,1}) \cosh(a + X_{-\{r,t\}}\theta_{r,-t,1}) \exp(X_{-\{r,t\}}\theta_{t,-r,1})}{\sum_{\delta=\pm 1} \ddot{f}(\delta a + X_{-\{r,t\}}\theta_{r,-t,1}) \cosh(\delta a + X_{-\{r,t\}}\theta_{r,-t,1}) \exp(\delta X_{-\{r,t\}}\theta_{t,-r,1})} \\ =& \frac{X_t \exp(-X_t X_{-\{r,t\}}\theta_{t,-r,1}) \cosh(a X_t + X_{-\{r,t\}}\theta_{r,-t,1})}{\sum_{\delta=\pm 1} \exp(\delta X_{-\{r,t\}}\theta_{t,-r,1}) \cosh(-\delta a + X_{-\{r,t\}}\theta_{r,-t,1})}. \end{aligned} \tag{A4}$$

Let $\mathcal{D}_{r,t}(X_{t'})$ denote the denominator of $v^o_{r,t,1}$ and $\mathcal{D}_{r,t'}(X_t)$ the denominator of $v^o_{r,t',1}$. We also need the conditional distribution

$$\operatorname{pr}\left(X_r, X_t, X_{t'} \mid X_{-\{r,t,t'\}}\right) = \frac{\exp\{X_r(aX_t + bX_{t'} + A_r) + X_t(cX_{t'} + A_t) + X_{t'}A_{t'}\}}{\mathcal{D}_{r,t,t'}}, \tag{A5}$$

where

$$\mathcal{D}_{r,t,t'} = \sum_{X_r, X_t, X_{t'} \in \{-1,1\}^3} \exp\{X_r(aX_t + bX_{t'} + A_r) + X_t(cX_{t'} + A_t) + X_{t'}A_{t'}\}.$$

By (A4), (A5) and the definition

$$\varepsilon_{r,1} = X_r - \dot{f}(X_{-r}\theta_{r,-r,1}) = X_r \frac{\exp\{-X_r(aX_t + A_r + bX_{t'})\}}{\cosh(aX_t + A_r + bX_{t'})},$$

one can calculate

$$E(v^o_{r,t,1} v^o_{r,t',1} \varepsilon^2_{r,1} \mid X_{-\{r,t,t'\}}) = \frac{1}{\mathcal{D}_{r,t,t'}} \sum_{X_t, X_{t'}} X_t X_{t'} \frac{\exp(-cX_tX_{t'}) 2\cosh(aX_t + A_r + bX_{t'})}{\mathcal{D}_{r,t}(X_{t'})\mathcal{D}_{r,t'}(X_t)}.$$

With a bit algebraic exercise, we obtain

$$E(v^o_{r,t,1} v^o_{r,t',1} \varepsilon^2_{r,1} \mid X_{-\{r,t,t'\}}) = \frac{2\Gamma_{r,t,t'}}{\mathcal{D}_{r,t'}(1)\mathcal{D}_{r,t'}(-1)} \frac{e^{A_{t'}}\mathcal{D}_{r,t}(-1) + e^{-A_{t'}}\mathcal{D}_{r,t}(1)}{\mathcal{D}_{r,t,t'}\mathcal{D}_{r,t}(1)\mathcal{D}_{r,t}(-1)}, \tag{A6}$$

where $\Gamma_{r,t,t'} = \sinh(-2c)\{\cosh(2A_r) + \cosh(2a)\cosh(2b)\} + \cosh(2a)\sinh(2a)\sinh(2b)$. The second term on the right-hand side of (A6) is equal to $E(|v^o_{r,t,1}\varepsilon_{r,1}| \mid X_{-\{r,t,t'\}})$, which is finite because $|v^o_{r,t,1}| \leq 1$ and $|\varepsilon_{r,1}| \leq 2$. Because $\mathcal{D}_{r,t'}(\pm 1) \geq 2$,

$$\left|E(v^o_{r,t,1} v^o_{r,t',1} \varepsilon^2_{r,1} \mid X_{-\{r,t,t'\}})\right| \leq \left|\Gamma_{r,t,t'}\right|.$$

Therefore by Jensen's inequality and Assumption 2,

$$\begin{aligned} |\operatorname{cov}(v^o_{r,t,1}\varepsilon_{r,1}, v^o_{r,t',1}\varepsilon_{r,1})| = |E(v^o_{r,t,1} v^o_{r,t',1} \varepsilon^2_{r,1})| &\leq E\{|E(v^o_{r,t,1} v^o_{r,t',1} \varepsilon^2_{r,1} \mid X_{-\{r,t,t'\}})|\} \\ &\leq C_0 |\sinh(2c)| + |\sinh(2a)\sinh(2b)| \cosh(2C_w) \\ &\leq C_0\{|\sinh(2\theta_{t,t',1})| + |\sinh(2\theta_{r,t,1})\sinh(2\theta_{r,t',1})|\}. \end{aligned}$$

Because $\mathrm{var}(v_{r,t,1}\varepsilon_{r,1}) = E\{(v^o_{r,t,1}\varepsilon_{r,1})^2\} = F_{r,t,1}/4$ and $F_{r,t,1} > c^*$ for all (r,t) by Assumption 2,

$$|\mathrm{cor}(v_{r,t,1}\varepsilon_{r,1}, v_{r,t',1}\varepsilon_{r',1})| \le \frac{4C_0}{c^*}\{|\sinh(2\theta_{t,t',1})| + |\sinh(2\theta_{r,t,1})\sinh(2\theta_{r,t',1})|\}.$$

This completes the proof of (A2). □

LEMMA A2. *Under Assumptions* 1 *and* 2*, we have uniformly for all* $1 \le r < t \le p$ *that* $\check\theta_{r,t,1}$ *defined in* (7) *and similarly defined* $\check\theta_{r,t,2}$ *are nearly unbiased. Moreover, for any* $1 \le r < t \le p$,

$$|\check\theta_{r,t,k} - \tilde\theta_{r,t,k}| = o_{\mathrm{p}}\{(n_k \log p)^{-1/2}\} \quad (k = 1, 2). \tag{A7}$$

Proof. We first show that $\check\theta_{r,t,k}$ $(k = 1, 2)$ are nearly unbiased. Without loss of generality, let $k = 1$.

Applying local Taylor expansion to $\dot f(u^{(i)}_{r,1})$ around $\hat u^{(i)}_{r,1}$, the data generating model in (6) can be rewritten as

$$X^{(i)}_r - \dot f(\hat u^{(i)}_{r,1}) + \ddot f(\hat u^{(i)}_{r,1})X^{(i)}_{-r}\hat\theta_{r,-r,1} = \ddot f(\hat u^{(i)}_{r,1})X^{(i)}_{-r}\theta_{r,-r,1} + Re_i + \varepsilon^{(i)}_{r,1}, \tag{A8}$$

where Re_i is the remainder term from the Taylor expansion and is of order $o_{\mathrm{p}}\{(n_1 \log p)^{-1/2}\}$ under the conditions in (15) and (16). Thus (A8) is an approximately linear model where the response is $X^{(i)}_r - \dot f(\hat u^{(i)}_{r,1}) + \ddot f(\hat u^{(i)}_{r,1})X^{(i)}_{-r}\hat\theta_{r,-r,1}$ and the predictors are $\ddot f(\hat u^{(i)}_{r,1})X^{(i)}_{-r}$. Applying the projection-based bias correction procedure for linear model (Zhang & Zhang, 2014), one immediately obtains (7). Let $\Delta_{r,1} = \hat\theta_{r,-r,1} - \theta_{r,-r,1}$ and let $h_{(i)} = X^{(i)}_{-r}\Delta_{r,1}$. To show that $\check\theta_{r,t,1}$ in (7) is nearly unbiased, we plug (A8) in (7) and obtain an equivalent definition:

$$\begin{aligned}\check\theta_{r,t,1} &= \theta_{r,t,1} + \frac{n_1^{-1}\sum_{i=1}^{n_1} v^{(i)}_{r,t,1}\varepsilon^{(i)}_{r,1}}{n_1^{-1}\sum_{i=1}^{n_1} v^{(i)}_{r,t,1}\ddot f(\hat u^{(i)}_{r,1})X^{(i)}_t} + \frac{REM_1 + REM_2}{n_1^{-1}\sum_{i=1}^{n_1} v^{(i)}_{r,t,1}\ddot f(\hat u^{(i)}_{r,1})X^{(i)}_t} \\ &= \theta_{r,t,1} + L_0 + L_{12},\end{aligned} \tag{A9}$$

where the two remainder terms are,

$$\begin{aligned}REM_1 =& \frac{1}{n_1}\sum_{i=1}^{n_1} v^{(i)}_{r,t,1}Re_i = \frac{1}{n_1}\sum_{i=1}^{n_1} v^{(i)}_{r,t,1}\ddot f(\hat u^{(i)}_{r,1})h_{\{i\}} - \frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^1 v^{(i)}_{r,t,1}\ddot f(\hat u^{(i)}_{r,1} - zh_{\{i\}})h_{\{i\}}dz,\\ REM_2 =& \frac{1}{n_1}\sum_{i=1}^{n_1} v^{(i)}_{r,t,1}\ddot f(\hat u^{(i)}_{r,1})X^{(i)}_{-\{r,t\}}\Delta_{\{r,-t\},1}.\end{aligned}$$

Since $E(v^{(i)}_{r,t,1}\varepsilon^{(i)}_{r,1}) = 0$, the bias of $\check\theta_{r,t,1}$ is approximately L_{12}.

In the low-dimensional setting where $\mathrm{rank}(X) = p < n_1$, one can choose $v_{r,t,1} = X^{\perp}_t$ as the projection of X_t onto the orthogonal complement of the column space of $X_{-\{r,t\}}$ to ensure $REM_2 = 0$. Since REM_1 is negligible, the resulting estimator $\check\theta_{r,t,1}$ in (A9) is nearly unbiased under Assumption 1. However, in the high-dimensional regime where $p \gg n$, $X^{\perp}_t$ is no longer a valid choice of the score vector as $\mathrm{rank}(X) < p$, and the condition $v_{r,t,1} \perp X_{-\{r,t\}}$ forces $v_{r,t,1}$ to be zero. On the other hand, the full strength of $v_{r,t,1} \perp X_{-\{r,t\}}$ is not necessary to ensure that the term REM_2 is also negligible. In fact, the score vector in (8) is sufficient to guarantee that $\check\theta_{r,t,1}$ is nearly unbiased.

To show (A7), by definition of $\check\theta_{r,t,1}$ in (A9),

$$\begin{aligned}\check\theta_{r,t,1}-\tilde\theta_{r,t,1} &= \underbrace{\frac{n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\varepsilon_{r,1}^{(i)}}{n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\ddot f(\hat u_{r,1}^{(i)})X_t^{(i)}}-\frac{n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\varepsilon_{r,1}^{(i)}}{F_{r,t,1}/2}}_{(i)} \\ &\quad+\underbrace{\frac{n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\varepsilon_{r,1}^{(i)}}{F_{r,t,1}/2}-\frac{n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i),o}\varepsilon_{r,1}^{(i)}}{F_{r,t,1}/2}}_{(ii)}+\underbrace{\frac{REM_1+REM_2}{n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\ddot f(\hat u_{r,1}^{(i)})X_t^{(i)}}}_{(iii)} \\ &= a_1+a_2+a_3.\end{aligned}$$

To bound $|\check\theta_{r,t,1}-\tilde\theta_{r,t,1}|$, it suffices to bound $|a_k|$ $(k=1,2,3)$ separately. To this end, we first bound $\ddot f(\hat u_{r,1}^{(i)})-\ddot f(u_{r,1}^{(i)})$ and $v_{r,t,1}^{(i)}-v_{r,t,1}^{(i),o}$.

Let $h_{(i),1}=X_{-\{r,t\}}^{(i)}\Delta_{\{r,-t\},1}$ and let $h_{(i),2}=X_{-\{r,t\}}^{(i)}\Delta_{\{t,-r\},1}$. Under Assumption 1, we can show via Hölder's inequality that

$$\max\left\{h_{(i)},|h_{(i),1}|,|h_{(i),2}|\right\}=o\{(\log p)^{-1}\}, \tag{A10}$$

and

$$\frac{1}{n_1}\sum_{i=1}^{n_1}\ddot f(u_{r,1}^{(i)})(h_{(i),1}^2+h_{(i),2}^2+\Delta_{\{r,t\},1}^2)=o\{(n_1\log p)^{-1/2}\}. \tag{A11}$$

Using the fact that for any $h\geq 0$,

$$\exp(-2h)\ddot f(x)\leq \ddot f(x+h)\leq \exp(2h)\ddot f(x), \tag{A12}$$

we have

$$|\ddot f(u_{r,1}^{(i)})-\ddot f(\hat u_{r,1}^{(i)})|\leq|\exp(2|h_{(i)}|)-1|\ddot f(u_{r,1}^{(i)})=o\{(\log p)^{-1}\}. \tag{A13}$$

In the meanwhile,

$$\begin{aligned}\max_{i=1,\ldots,n_1}|v_{r,t,1}^{(i)}-v_{r,t,1}^{(i),o}| &= \max_{i=1,\ldots,n_1}\left|g(X_{-\{r,t\}}^{(i)},\hat\theta_{r,-r,1},\hat\theta_t,1)-g(X_{-\{r,t\}}^{(i)},\theta_{r,-r,1},\theta_{t,-t,1})\right| \\ &\leq \max_{i=1,\ldots,n_1}\left|\exp(|h_{(i),1}|+|h_{(i),2}|+|\Delta_{\{r,t\},1}|)-1\right|=o\{(\log p)^{-1}\}, \end{aligned}\tag{A14}$$

where the inequality follows from Lemma 1 of an unpublished 2016 technical report from Zhao Ren.

Consequently, by (A13), (A14) and Lemma 3 of an unpublished 2016 technical report from Zhao Ren,

$$\begin{aligned}\left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\ddot f(\hat u_{r,1}^{(i)})X_t^{(i)}-\frac12 F_{r,t,1}\right| &\leq \left|\frac{1}{n_1}\sum_{i=1}^{n_1}(v_{r,t,1}^{(i)}-v_{r,t,1}^{(i),o})\ddot f(\hat u_{r,1}^{(i)})X_t^{(i)}\right| \\ &\quad+\left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i),o}\{\ddot f(\hat u_{r,1}^{(i)})-\ddot f(u_{r,1}^{(i)})\}X_t^{(i)}\right| \\ &\quad+\left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i),o}\ddot f(u_{r,1}^{(i)})X_t^{(i)}-\frac12 F_{r,t,1}\right| \\ &\leq o\{(\log p)^{-1}\}. \end{aligned}\tag{A15}$$

By definition of the score vector, $E(v_{r,t,1}^{(i),o}\varepsilon_{r,1}^{(i)})=0$ and $E\{(v_{r,t,1}^{(i),o}\varepsilon_{r,1}^{(i)})^2\}=F_{r,t,1}/4$. Therefore by Hoeffding's inequality, there exists $C_2>0$ such that for any $\xi>0$,

$$\mathrm{pr}\left\{\left|n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i),o}\varepsilon_{r,1}^{(i)}\right|\geq C_2(\log p/n_1)^{1/2}\right\}\leq O(p^{-\xi}).$$

Similarly there exists $C_3 > 0$ such that

$$\mathrm{pr}\left\{\left|n_1^{-1}\sum_{i=1}^{n_1}\varepsilon_{r,1}^{(i)}\right| \geq C_3(\log p/n_1)^{1/2}\right\} \leq O(p^{-\xi}).$$

Hence we have $|n_1^{-1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\varepsilon_{r,1}^{(i)}| = O_{\mathrm{p}}\{(\log p/n_1)^{1/2}\}$, and

$$|a_1| = O_{\mathrm{p}}\{(\log p/n_1)^{1/2}\}\left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})X_t^{(i)} - \frac{1}{2}F_{r,t,1}\right| = o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\}.$$

For term a_2, recall that $F_{r,t,1} > c^* > 0$ by Assumption 2 and $v_{r,t,1}^{(i)} - v_{r,t,1}^{(i),o}$ is independent from $\varepsilon_{r,1}^{(i)}$, applying Hoeffding's inequality directly to bounded random variables $(v_{r,t,1}^{(i)} - v_{r,t,1}^{(i),o})\varepsilon_{r,1}^{(i)}$ yields

$$|a_2| = \frac{2}{F_{r,t,1}}\left|\frac{1}{n_1}\sum_{i=1}^{n_1}(v_{r,t,1}^{(i)} - v_{r,t,1}^{(i),o})\varepsilon_{r,1}^{(i)}\right| = o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\}.$$

Finally, we bound a_3. Because its denominator is of constant order by (A15), it suffices to bound the numerators REM_1 and REM_2 separately. To bound REM_2, by (A14), (A11), and equation (68) in an unpublished 2016 technical report from Zhao Ren,

$$\begin{aligned}\left|\frac{1}{n_1}\sum_{i=1}^{n_1}(v_{r,t,1}^{(i)} - v_{r,t,1}^{(i),o})\ddot{f}(\hat{u}_{r,1}^{(i)})h_{(i),1}\right| &\leq \frac{C}{n_1}\sum_{i=1}^{n_1}\left(|h_{(i),1}| + |h_{(i),2}| + |\Delta_{\{r,t\},1}|\right)\ddot{f}(\hat{u}_{r,1}^{(i)})|h_{(i),1}| \\ &= o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\}.\end{aligned}$$

Thus by Lemma 3 of an unpublished 2016 technical report from Zhao Ren,

$$\begin{aligned}|REM_2| &\leq o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\} + \left|\frac{1}{n_1}\sum_{i=1}^{n_1} v_{r,t,1}^{(i),o}\ddot{f}(\hat{u}_{r,1}^{(i)})h_{(i),1}\right| \\ &\leq o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\} + (\log p/n_1)^{1/2}\|\Delta_{r,1}\|_1 = o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\}.\end{aligned}$$

For REM_1, by equations (A10), (A11) and (A12) and the fact that $|v_{r,t,1}^{(i)}| \leq 1$,

$$\begin{aligned}|REM_1| &= \left|\sum_{i=1}^{n_1} v_{r,t,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)})h_{(i)} - \frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^1 v_{r,t,1}^{(i)}\ddot{f}(\hat{u}_{r,1}^{(i)} - zh_{\{i\}})h_{\{i\}}dz\right| \\ &= \left|\frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^1 v_{r,t,1}^{(i)}\{\ddot{f}(u_{r,1}^{(i)} + h_{\{i\}}) - \ddot{f}(u_{r,1}^{(i)} + zh_{\{i\}})\}h_{\{i\}}dz\right| \\ &\leq \frac{1}{n_1}\sum_{i=1}^{n_1}|v_{r,t,1}^{(i)}|\ddot{f}(u_{r,1}^{(i)})|\exp(2h_{\{i\}}) - 1||h_{\{i\}}| \\ &\leq \frac{C}{n_1}\sum_{i=1}^{n_1}|v_{r,t,1}^{(i)}|\ddot{f}(u_{r,1}^{(i)})|h_{\{i\}}^2| = o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\}.\end{aligned}$$

It follows immediately that $|a_3| = o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\}$.

In summary, we have shown that

$$|\tilde{\theta}_{r,t,1} - \check{\theta}_{r,t,1}| = o_{\mathrm{p}}\{(n_1\log p)^{-1/2}\}.$$

LEMMA A3. *Under Assumptions* 1 *and* 2*,*

$$\max_{1\leq r<t\leq p}|\check{s}_{r,t,k} - s_{r,t,k}| = o_{\mathrm{p}}\{(\log p)^{-1}\} \quad (k = 1, 2). \tag{A16}$$

Proof. We only prove (A16) for $k = 1$. By definition,

$$\check{s}_{r,t,1} = \left\{ 4n_1^{-1} \sum_{i=1}^{n_1} (v_{r,t,1}^{(i)})^2 \ddot{f}\big(\hat{u}_{r,1}^{(i)}\big) \right\}^{-1}.$$

It is easy to check that $4n_1^{-1} \sum_{i=1}^{n_1} (v_{r,t,1}^{(i)})^2 \ddot{f}\big(\hat{u}_{r,1}^{(i)}\big)$ is bounded. We first look at the difference

$$\frac{1}{n_1} \sum_{i=1}^{n_1} (v_{r,t,1}^{(i)})^2 \ddot{f}\big(X_{-r}^{(i)} \hat{\theta}_{r,1}\big) - \frac{1}{4} F_{r,t,1}.$$

Following the proof of Lemma A2, one can show that

$$\begin{aligned} \Big| \frac{1}{n_1} \sum_{i=1}^{n_1} (v_{r,t,1}^{(i)})^2 \ddot{f}(\hat{u}_{r,1}) - \frac{1}{4} F_{r,t,1} \Big| \leq & \Big| \frac{1}{n_1} \sum_{i=1}^{n_1} \{ (v_{r,t,1}^{(i)})^2 - (v_{r,t,1}^{(i),o})^2 \} \ddot{f}(\hat{u}_{r,1}) \Big| \\ & + \Big| \frac{1}{n_1} \sum_{i=1}^{n_1} (v_{r,t,1}^{(i),o})^2 \{ \ddot{f}(\hat{u}_{r,1}) - \ddot{f}(u_{r,1}) \} \Big| \\ & + \Big| \frac{1}{n_1} \sum_{i=1}^{n_1} (v_{r,t,1}^{(i),o})^2 \ddot{f}(u_{r,1}) - \frac{1}{4} F_{r,t,1} \Big| \\ = & b_1 + b_2 + b_3, \end{aligned}$$

where $b_1 = o_{\mathrm{p}}\{(\log p)^{-1}\}, b_2 = o_{\mathrm{p}}\{(\log p)^{-1}\}$ and $b_3 = O_{\mathrm{p}}\{(\log p/n_k)^{1/2}\} = o_{\mathrm{p}}\{(\log p)^{-1}\}$. Thus

$$\Big| \frac{1}{n_1} \sum_{i=1}^{n_1} (v_{r,t,1}^{(i)})^2 \ddot{f}(\hat{u}_{r,1}) - \frac{1}{4} F_{r,t,1} \Big| = o_{\mathrm{p}}\{(\log p)^{-1}\}.$$

The result in (A16) follows immediately because the denominator of $\check{s}_{r,t,1}$ is bounded below by a constant. □

LEMMA A4. *Under the conditions of Theorem* 1*, there exists some constant* $C > 0$ *such that for any* $\epsilon > 0$

$$\mathrm{pr}\left\{ \max_{(r,t)\in\Lambda} \frac{(\tilde{\theta}_{r,t,1} - \tilde{\theta}_{r,t,2} - \theta_{r,t,1} + \theta_{r,t,2})^2}{s_{r,t,1}/n_1 + s_{r,t,2}/n_2} \geq x^2 \right\} \leq C|\Lambda|\{1 - \Phi(x)\} + O(p^{-\epsilon}),$$

uniformly for $0 \leq x \leq (8 \log p)^{1/2}$ *and* $\Lambda \subset \{(r,t) : 1 \leq r < t \leq p\}$.

Proof. Let

$$\begin{aligned} Z_{r,t,i} &= \frac{n_2}{n_1} \frac{2 v_{r,t,1}^{(i),o} \varepsilon_{r,1}^{(i)}}{F_{r,t,1}} \quad (i = 1, \ldots, n_1), \\ Z_{r,t,i} &= -\frac{2 v_{r,t,2}^{(i),o} \varepsilon_{r,2}^{(i)}}{F_{r,t,2}} \quad (i = n_1 + 1, \ldots, n_1 + n_2). \end{aligned}$$

Then

$$\frac{(\tilde{\theta}_{r,t,1} - \tilde{\theta}_{r,t,2} - \theta_{r,t,1} + \theta_{r,t,2})^2}{s_{r,t,1}/n_1 + s_{r,t,2}/n_2} = \frac{(\sum_{i=1}^{n_1+n_2} Z_{r,t,i})^2}{\sum_{i=1}^{n_1+n_2} Z_{r,t,i}^2} \frac{\sum_{i=1}^{n_1+n_2} Z_{r,t,i}^2}{n_2^2 s_{r,t,1}/n_1 + n_2 s_{r,t,2}}. \tag{A17}$$

It is easy to check that $n_1^{-1}\sum_{i=1}^{n_1} Z_{r,t,i}^2 \to (n_2/n_1)^2 s_{r,t,1}$ and $n_2^{-1}\sum_{i=1}^{n_2} Z_{r,t,i}^2 \to s_{r,t,2}$. In particular, since $Z_{r,t,i}^2$ are bounded for all i, we apply Hoeffding's inequality to obtain

$$\mathrm{pr}\left\{\frac{1}{n_1}\sum_{i=1}^{n_1} Z_{r,t,i}^2 - \frac{n_2^2}{n_1^2}s_{r,t,1} \geq C(\log p/n_1)^{1/2}\right\} \leq p^{-2C^2},$$
$$\mathrm{pr}\left\{\frac{1}{n_2}\sum_{i=1+n_1}^{n_1+n_2} Z_{r,t,i}^2 - s_{r,t,2} \geq C(\log p/n_2)^{1/2}\right\} \leq p^{-2C^2}.$$

Thus the second term on the right-hand side of (A17) can be treated as a constant, up to a small deviation in the order of $(\log p/n_k)^{1/2}$. It suffices to check the large deviation bound of the term

$$\frac{(\sum_{i=1}^{n_1+n_2} Z_{r,t,i})^2}{\sum_{i=1}^{n_1+n_2} Z_{r,t,i}^2}.$$

Since $Z_{r,t,i}$ $(i = 1, \ldots, n_1 + n_2)$ are all bounded and thus have finite $(2 + \delta)$th moment $(0 < \delta \leq 1)$, we can apply the self-normalized large deviation theorem for independent random variables in Jing et al. (2003) to obtain

$$\max_{1\leq r<t\leq p} \mathrm{pr}\left\{\frac{(\sum_{i=1}^{n_1+n_2} Z_{r,t,i})^2}{\sum_{i=1}^{n_1+n_2} Z_{r,t,i}^2} \geq x^2\right\} \leq C\{1 - \Phi(x)\},$$

uniformly for $0 \leq x \leq (8\log p)^{1/2}$. □

LEMMA A5. *Let $\widetilde{W}_{r,t} = (\tilde\theta_{r,t,1} - \tilde\theta_{r,t,2})(s_{r,t,1}/n_1 + s_{r,t,2}/n_2)^{-1/2}$. Under the conditions of Theorem 3, for any $\epsilon > 0$,*

$$\int_0^{b_p} \mathrm{pr}\left[\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{r,t}| \geq \tau) - \mathrm{pr}(|\widetilde{W}_{r,t}| \geq \tau)\}|}{q_0 G(\tau)} \geq \epsilon\right] d\tau = o(v_p), \tag{A18}$$

$$\sup_{0\leq\tau\leq b_p} \mathrm{pr}\left[\frac{|\sum_{(r,t)\in\mathcal{H}_0}\{I(|\widetilde{W}_{r,t}| \geq \tau) - \mathrm{pr}(|\widetilde{W}_{r,t}| \geq \tau)\}|}{q_0 G(\tau)} \geq \epsilon\right] = o(1), \tag{A19}$$

where $b_p = \{4\log p - 2\log(\log p)\}^{1/2}$ and $v_p = (c_p^2 \log p)^{-1/2}$ with $c_p = \log\log\log\log p$.

Proof. We only show (A18) as the proof of (A19) follows similarly. To this end, define

$$\mathcal{D}_0 = \{(r,t) : 1 \leq t \leq p, r \in \mathcal{A}_t(\xi)\}, \quad \mathcal{H}_{01} = \mathcal{H}_0 \cap \mathcal{D}_0, \quad \mathcal{H}_{02} = \mathcal{H}_0 \cap \mathcal{D}_0^c,$$

where $\mathcal{A}_t(\xi)$ is defined in Assumption 3. By Assumption 3, we have $|\mathcal{H}_{01}| = o(p^{1+\gamma})$ for $0 < \gamma < 1$. By (12) and the condition that $q_0 \geq c_1 p^2$,

$$E\left[\frac{|\sum_{(r,t)\in\mathcal{H}_{01}}\{I(|\widetilde{W}_{r,t}| \geq \tau) - \mathrm{pr}(|\widetilde{W}_{r,t}| \geq \tau)\}|}{q_0 G(\tau)}\right] \lesssim \frac{p^{1+\gamma}G(\tau)}{q_0 G(\tau)} = O(p^{-1+\gamma}). \tag{A20}$$

To evaluate (A18) over the set $\mathcal{H}_{02}$, we split $\mathcal{H}_{02}$ into several subsets as done in Cai & Liu (2016). The criterion for such a split is based on the function $\mathrm{Corr}_k(r,t,l,m)$, which is defined such that for $k \in \{1,2\}, r \neq t, l \neq m$ and $r,t,l,m \in \{1,\ldots,p\}$,

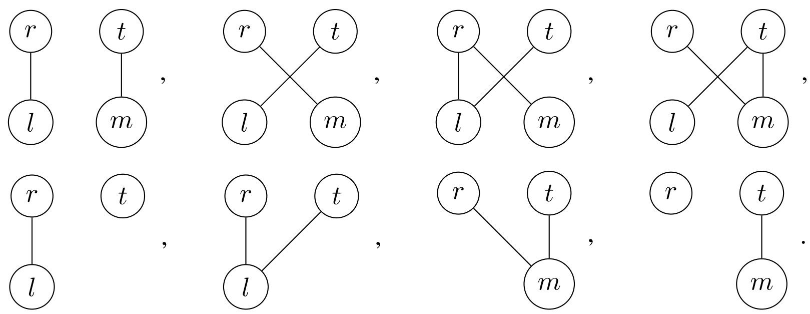
$$\mathrm{Corr}_k(r,t,l,m) = \begin{cases} 4(c^*)^{-1}C_0\{|\sinh(2\theta_{r,l,k})| + |\sinh(2\theta_{r,t,k})\sinh(2\theta_{l,t,k})|\}, & t = m, \\ 4(c^*)^{-1}C_0\{|\sinh(2\theta_{t,m,k})| + |\sinh(2\theta_{r,t,k})\sinh(2\theta_{r,m,k})|\}, & r = l, \\ (c^*)^{-1}C_1\{\sum_{a=m,l}|\sinh(2\theta_{r,a,k})|\}\{\sum_{a=m,l}|\sinh(2\theta_{t,a,k})|\}, & r \neq l, t \neq m. \end{cases}$$

By Lemma 1, $\mathrm{Corr}_1(r,t,l,m)$ and $\mathrm{Corr}_2(r,t,l,m)$ together approximately quantify the dependence between $\widetilde{W}_{r,t}$. For some large constant $C>0$, define the following events

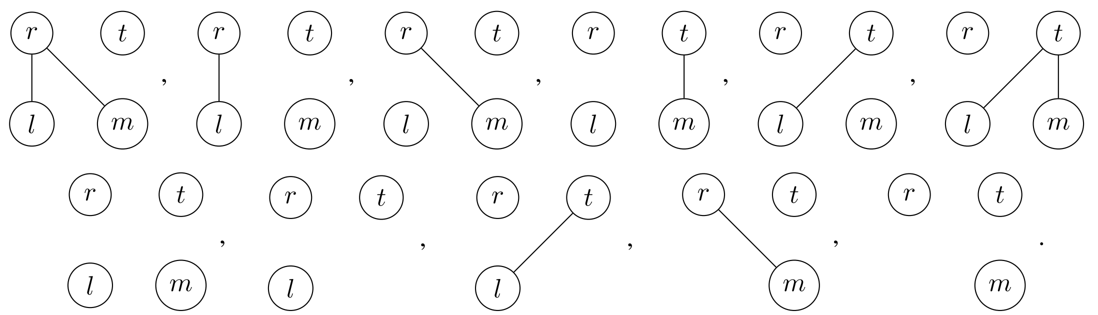
$$\mathcal{H}_{3X} = \{(r,t,l,m) : (r,t) \in \mathcal{H}_{02}, (l,m) \in \mathcal{H}_{02}, \mathrm{Corr}_1(r,t,l,m) \le C(\log p)^{-2-\xi}\},$$
$$\mathcal{H}_{4X} = \{(r,t,l,m) \notin \mathcal{H}_{3X} : (r,t) \in \mathcal{H}_{02}, (l,m) \in \mathcal{H}_{02}, \mathrm{Corr}_1(r,t,l,m) \le C\{0.5 + (\log p)^{-2-\xi}\}\},$$
$$\mathcal{H}_{5X} = \{(r,t,l,m) \notin \mathcal{H}_{3X} \cup \mathcal{H}_{4X} : (r,t) \in \mathcal{H}_{02}, (l,m) \in \mathcal{H}_{02}\}.$$

The events $\mathcal{H}_{3Y}, \mathcal{H}_{4Y}, \mathcal{H}_{5Y}$ are defined similarly. Let $\mathcal{H}_3 = \mathcal{H}_{3X} \cap \mathcal{H}_{3Y}$, $\mathcal{H}_5 = \mathcal{H}_{5X} \cup \mathcal{H}_{5Y}$, and $\mathcal{H}_4 = \mathcal{H}_{4X} \cup \mathcal{H}_{4Y} \backslash \mathcal{H}_5$.

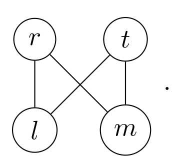
By Assumption 3, one can easily enumerate the different scenarios in $\mathcal{H}_3, \mathcal{H}_4$ and $\mathcal{H}_5$. Let $G_{rtlm} = (V_{rtlm}, E_{rtlm})$ be a graph, where $V_{rtlm} = \{r,t,l,m\}$ is the set of vertices and E_{rtlm} is the set of edges. There is an edge between $a \neq b \in \{r,t,l,m\}$ if and only if $|\sinh(\theta_{a,b,k})| \ge (\log p)^{-2-\xi}$. If the number of distinct vertices in V_{rtlm} is 3, then we call G_{rtlm} a three-vertex graph (3-G); similarly if there are 4 distinct vertices in V_{rtlm}, then G_{rtlm} is a four-vertex graph (4-G). Therefore the different scenarios in $\mathcal{H}_4$ consist of:



Similarly, the different scenarios in $\mathcal{H}_3$ can be listed as:



What remains goes into the set $\mathcal{H}_5$: either $(r,t) = (l,m)$ or



As a result, one can show that $|\mathcal{H}_4| = O(p^{2+2\gamma})$ and $|\mathcal{H}_5| = O(p^2 + p^{1+3\gamma})$.

Let $f_{rtlm}(\tau) = \mathrm{pr}(|\widetilde{W}_{r,t}| \ge \tau, |\widetilde{W}_{lm}| \ge \tau) - \mathrm{pr}(|\widetilde{W}_{r,t}| \ge \tau)\mathrm{pr}(|\widetilde{W}_{lm}| \ge \tau)$. Then

$$E\left(\left[\frac{|\sum_{(r,t)\in\mathcal{H}_{02}}\{I(|\widetilde{W}_{r,t}| \ge \tau) - \mathrm{pr}(|\widetilde{W}_{r,t}| \ge \tau)\}|}{q_0 G(\tau)}\right]^2\right) = \frac{\sum_{(r,t)\in\mathcal{H}_{02}} \sum_{(l,m)\in\mathcal{H}_{02}} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}. \tag{A21}$$

By Lemma 4 in the supplementary materials of Cai & Liu (2016),

$$\left|\frac{\sum_{(r,t,l,m)\in\mathcal{H}_3} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}\right| \le CA_n, \tag{A22}$$

where $A_n \le (\log p)^{-1-\gamma_2}$ for some $\gamma_2 > 0$. Further, we also have for $\eta = 0.5$,

$$\left|\frac{\sum_{(r,t,l,m)\in\mathcal{H}_4} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}\right| \le \frac{C(\tau+1)^{2/(1+\eta)-2}}{p^{2-2\gamma}\{G(\tau)\}^{2\eta/(1+\eta)}}, \tag{A23}$$

$$\left|\frac{\sum_{(r,t,l,m)\in\mathcal{H}_5} f_{rtlm}(\tau)}{q_0^2 G^2(\tau)}\right| \le \frac{C}{p^2 G(\tau)} + \frac{C}{p^{3-3\gamma}G(\tau)}. \tag{A24}$$

By the tail bound of standard normal random variable, $G(\tau) \ge G(b_p) \approx (2\pi b_p^2)^{-1/2}\exp(-b_p^2/2)$ for $0 \le \tau \le b_p$. Thus by Assumption 3, the following holds:

$$\int_0^{b_p}\left[p^{-1+\gamma} + A_n + \frac{C(\tau+1)^{2/(1+\zeta)-2}}{p^{2-2\gamma}\{G(\tau)\}^{2\zeta/(1+\zeta)}} + \frac{C}{p^2 G(\tau)} + \frac{C}{p^{3-3\gamma}G(\tau)}\right] d\tau = o(v_p). \tag{A25}$$

Combining (A20), (A21), (A22), (A23), (A24) and (A25), we prove (A18). □

A·2. *Proof of Theorem* 1

Proof. Recall the test statistic is

$$M_{n,p} = \max_{1\le r<t\le p} W_{r,t}^2 = \max_{1\le r<t\le p} \frac{(\check{\theta}_{r,t,1} - \check{\theta}_{r,t,2})^2}{\check{s}_{r,t,1}/n_1 + \check{s}_{r,t,2}/n_2}. \tag{A26}$$

Define

$$\widehat{M}_{n,p} = \max_{1\le r<t\le p} \widehat{W}_{r,t}^2 = \max_{1\le r<t\le p} \frac{(\check{\theta}_{r,t,1} - \check{\theta}_{r,t,2})^2}{s_{r,t,1}/n_1 + s_{r,t,2}/n_2},$$

$$\widetilde{M}_{n,p} = \max_{1\le r<t\le p} \widetilde{W}_{r,t}^2 = \max_{1\le r<t\le p} \frac{(\tilde{\theta}_{r,t,1} - \tilde{\theta}_{r,t,2})^2}{s_{r,t,1}/n_1 + s_{r,t,2}/n_2}.$$

Under the null, let $\theta_{rt} = \theta_{r,t,1} = \theta_{r,t,2}$. Recall the definition of $\mathcal{A}_t(\xi)$ in Assumption 3, and let $\mathcal{B}_0 = \{(r,t) : r \in \Lambda(\eta^*), r < t < p\} \cup \{(r,t) : t \in \Lambda(\eta^*), 1 \le r < t\}$ and let $\mathcal{D}_0 = \{(r,t) : t = 1,\ldots,p, r \in \mathcal{A}_t(\xi)\} \cup \mathcal{B}_0$. To prove Theorem 1, we first show that (i) $M_{n,p}$ is close to $\widetilde{M}_{n,p}$, and (ii) terms in $\mathcal{D}_0$ are negligible for defining the limiting distribution of $\widetilde{M}_{n,p}$. Consequently it suffices to study the limiting distribution of $\widetilde{M}_{n,p}$ on $\mathcal{A}\backslash\mathcal{D}_0$, where $\mathcal{A} = \{(r,t) : 1 \le r < t \le p\}$.

By definition, $|v_{r,t,k}^o| \le 1, |\ddot{f}(u_{r,k})| \le 1$ for $k = 1,2$. Therefore $F_{r,t,k} = 4E_{\Theta_k}\{(v_{r,t,k}^o)^2\ddot{f}(u_{r,k})\} \le 4$ and $s_{rt,k} = 1/F_{rt,k} \ge 1/4$. Hence

$$\begin{aligned}
W_{r,t} - \widetilde{W}_{r,t} =& W_{r,t} - \widehat{W}_{r,t} + \widehat{W}_{r,t} - \widetilde{W}_{r,t} \\
=& W_{r,t} o_{\mathrm{p}}\{(\log p)^{-1}\} + \frac{(\check{\theta}_{r,t,1} - \check{\theta}_{r,t,2}) - (\tilde{\theta}_{r,t,1} - \tilde{\theta}_{r,t,2})}{(s_{r,t,1}/n_1 + s_{r,t,2}/n_2)^{1/2}} \\
=& W_{r,t} o_{\mathrm{p}}\{(\log p)^{-1}\} + o_{\mathrm{p}}\{(\log p)^{-1/2}\} = o_{\mathrm{p}}\{(\log p)^{-1/2}\},
\end{aligned}$$

where we have used Lemma A2 and A3. Thus it suffices to show

$$\mathrm{pr}(\widetilde{M}_{n,p} - 4\log p + \log\log p \le z) \to \exp\{-(8\pi)^{-1/2}e^{-z/2}\},$$

as $n_1, n_2, p \to \infty$. Let $y_p = z + 4\log p - \log\log p$ and let

$$\widetilde{M}_{\mathcal{D}_0} = \max_{(r,t)\in\mathcal{D}_0} \frac{(\tilde{\theta}_{r,t,1} - \tilde{\theta}_{r,t,2})^2}{s_{r,t,1}/n_1 + s_{r,t,2}/n_2}, \quad \widetilde{M}_{\mathcal{A}\backslash\mathcal{D}_0} = \max_{(r,t)\in\mathcal{A}\backslash\mathcal{D}_0} \frac{(\tilde{\theta}_{r,t,1} - \tilde{\theta}_{r,t,2})^2}{s_{r,t,1}/n_1 + s_{r,t,2}/n_2}.$$

Then $\mathrm{pr}(\widetilde{M}_{n,p} \geq y_p) - \mathrm{pr}(\widetilde{M}_{\mathcal{A}\backslash\mathcal{D}_0} \geq y_p) \leq \mathrm{pr}(\widetilde{M}_{\mathcal{D}_0} \geq y_p)$. Assumption 3 and the bounded cardinality condition on $\Lambda(\eta^*)$ indicate that $|\mathcal{D}_0| = o(p^{1+\gamma})$. Thus by Lemma A4, for any fixed $x \in \mathbb{R}$, we must have

$$\mathrm{pr}(\widetilde{M}_{\mathcal{D}_0} \geq y_p) \leq |\mathcal{D}_0| C p^{-2} + o(1) = o(1).$$

Hence it suffices to show that

$$\mathrm{pr}(\widetilde{M}_{\mathcal{A}\backslash\mathcal{D}_0} - 4\log p + \log\log p \leq z) \to \exp\{-(8\pi)^{-1/2} e^{-z/2}\},$$

as $n_1, n_2, p \to \infty$.

Now let us rearrange the two-dimensional indices in the set $\{(r,t) : (r,t) \in \mathcal{A}\backslash\mathcal{D}_0\}$ as $\{(r_j, t_j) : j = 1, \ldots, q\}$, where $q = |\mathcal{A}\backslash\mathcal{D}_0|$. For $j = 1, \ldots, q$, let

$$Z_{ji} = \frac{n_2}{n_1} \frac{2 v^{(i),o}_{r_j,t_j,1} \varepsilon^{(i)}_{r_j,1}}{F_{r_j,t_j,1}} \quad (i = 1, \ldots, n_1),$$

$$Z_{ji} = -\frac{2 v^{(i),o}_{r_j,t_j,2} \varepsilon^{(i)}_{r_j,2}}{F_{r_j,t_j,2}} \quad (i = n_1 + 1, \ldots, n_1 + n_2).$$

Note the Z_{ji}'s are all bounded. Let

$$V_j = (n_2^2/n_1 s_{j1} + n_2 s_{j2})^{-1/2} \sum_{i=1}^{n_1+n_2} Z_{ji} \quad (j = 1, \ldots, q).$$

It suffices to show that for any $z \in \mathbb{R}$, as $n_1, n_2, p \to \infty$,

$$\mathrm{pr}\left(\max_{j=1,\ldots,q} V_j^2 - 4\log p + \log\log p \leq z\right) \to \exp\left\{-(8\pi)^{-1/2} \exp(-\frac{z}{2})\right\}. \tag{A27}$$

Let $\mathcal{E}_{j_g} = \{V_{j_g}^2 \geq y_p\}$. By the Bonferroni inequality in Lemma 1 of Cai et al. (2013), for any integer m with $0 < m < q/2$,

$$\begin{aligned}\sum_{\ell=1}^{2m} (-1)^{\ell-1} \sum_{1 \leq j_1 < \cdots < j_\ell \leq q} \mathrm{pr}\left(\bigcap_{g=1}^{\ell} \mathcal{E}_{j_g}\right) &\leq \mathrm{pr}\left(\max_{j=1,\ldots,q} V_j^2 \geq y_p\right) \\ &\leq \sum_{\ell=1}^{2m-1} (-1)^{\ell-1} \sum_{1 \leq j_1 < \cdots < j_\ell \leq q} \mathrm{pr}\left(\bigcap_{g=1}^{\ell} \mathcal{E}_{j_g}\right). \end{aligned} \tag{A28}$$

Let

$$\tilde{Z}_{ji} = Z_{ji}/(n_2 s_{j1}/n_1 + s_{j2})^{1/2} \quad (j = 1, \ldots, q),$$

and $W_i = (\tilde{Z}_{j_1 i}, \ldots, \tilde{Z}_{j_\ell i})$ for all $i = 1, \ldots, n_1 + n_2$. Denote $|a|_{\min} = \min_{i=1,\ldots,\ell} |a_i|$ for any vector $a \in \mathbb{R}^\ell$. Then

$$\mathrm{pr}\left(\bigcap_{g=1}^{\ell} \mathcal{E}_{j_g}\right) = \mathrm{pr}\left(\left|n_2^{-1/2} \sum_{i=1}^{n_1+n_2} W_i\right|_{\min} \geq {y_p}^{1/2}\right).$$

By Theorem 1 in Zaitsev (1987),

$$\begin{aligned}\mathrm{pr}\left(\left|n_2^{-1/2} \sum_{i=1}^{n_1+n_2} W_i\right|_{\min} \geq {y_p}^{1/2}\right) \leq &\mathrm{pr}\left\{|N_\ell|_{\min} \geq {y_p}^{1/2} - \epsilon_n (\log p)^{-1/2}\right\} \\ &+ c_1 \ell^{5/2} \exp\left\{-\frac{n_2^{1/2} \epsilon_n}{c_2 \ell^{5/2} \tau_n (\log p)^{1/2}}\right\}. \end{aligned} \tag{A29}$$

where $c_1, c_2 > 0$ are absolute constants, $\epsilon_n \to 0$ whose rate will be specified later, $\tau_n > 0$ are constants, and $N_\ell = (N_{j_1}, \ldots, N_{j_\ell})$ is an ℓ-dimensional normal vector with mean zero and covariance $n_1\mathrm{cov}(W_1)/n_2 + \mathrm{cov}(W_{n_1+1})$. Recall that ℓ is a fixed integer that is independent of n_k and p. Since $\log p = o(n_k^{1/3})$ by Assumption 1, we can choose $\epsilon_n \to 0$ sufficiently slowly such that

$$c_1 \ell^{5/2} \exp\left\{-\frac{n_2^{1/2}\epsilon_n}{c_2 \ell^{5/2} \tau_n (\log p)^{1/2}}\right\} = o(p^{-2}). \tag{A30}$$

Thus by (A28), (A29), (A30),

$$\mathrm{pr}\left(\max_{j=1,\ldots,q} V_j^2 \geq y_p\right) \leq \sum_{\ell=1}^{2m-1} (-1)^{\ell-1} \sum_{1\leq j_1<\cdots<j_\ell\leq q} \mathrm{pr}\left\{|N_\ell|_{\min} \geq {y_p}^{1/2} - \epsilon_n (\log p)^{-1/2}\right\} + o(1). \tag{A31}$$

Similarly, by Theorem 1 in Zaitsev (1987),

$$\mathrm{pr}\left(\max_{j=1,\ldots,q} V_j^2 \geq y_p\right) \geq \sum_{\ell=1}^{2m} (-1)^{\ell-1} \sum_{1\leq j_1<\cdots<j_\ell\leq q} \mathrm{pr}\left\{|N_\ell|_{\min} \geq {y_p}^{1/2} - \epsilon_n (\log p)^{-1/2}\right\} + o(1). \tag{A32}$$

To complete the proof, we need to verify that Lemma 5 in Cai et al. (2013) also holds in our setting. A key condition is that the random variables Z_{ji} and $Z_{j'i}$ are only weakly dependent for $j \neq j'$. By Lemma 1 and Assumption 3, one can show that such weak dependence between Z_{ji} and $Z_{j'i}$ $(j \neq j')$ holds. In addition, Lemma A6 below is trivially satisfied since otherwise the Hessian of the likelihood function $Q_{r,k}$ would become singular. Hence Lemma 5 in Cai et al. (2013) also holds in the current context. Finally, applying this lemma to (A31) and (A32), one has for any positive integer $m > 0$,

$$\limsup_{n_k, p\to\infty} \mathrm{pr}\left(\max_{j=1,\ldots,q} V_j^2 \geq y_p\right) \leq \sum_{\ell=1}^{2m-1} (-1)^{\ell-1} \frac{1}{\ell!} \left\{(8\pi)^{-1/2} \exp\left(-\frac{z}{2}\right)\right\}^\ell \{1 + o(1)\},$$
$$\limsup_{n_k, p\to\infty} \mathrm{pr}\left(\max_{j=1,\ldots,q} V_j^2 \geq y_p\right) \geq \sum_{\ell=1}^{2m} (-1)^{\ell-1} \frac{1}{\ell!} \left\{(8\pi)^{-1/2} \exp\left(-\frac{z}{2}\right)\right\}^\ell \{1 + o(1)\}.$$

Letting $m \to \infty$, we obtain the desired convergence in (A27). □

LEMMA A6 (BERMAN, 1962). *If X and Y have a bivariate normal distribution with mean zero, unit variance and correlation coefficient ρ, then*

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

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

A·3. *Proof of Theorem* 2

Proof. Statement (i) follows from the proof of Theorem 2 in Xia et al. (2015). To prove the lower bound result in (ii), we first construct the worst case scenario to test between Θ_1 and Θ_2.

Let $\mathcal{M}(a_p, p-1)$ be the set of all subsets of $\{1, \ldots, p-1\}$ with cardinality $a_p = O(p^\gamma)$ for $\gamma < 1/2$. Let $\hat{m}$ be a random variable uniformly distributed over $\mathcal{M}(a_p, p-1)$. We construct a class $\mathcal{N} = \{\Theta_{\hat{m}}, \hat{m} \in \mathcal{M}(a_p, p-1)\}$, such that $\theta_{rt} = 0$ for $|r-t| > 1$, $\theta_{r,r+1} = \theta_{r+1,r} = \rho I(r \in \hat{m})$ and $\rho = c(\log p/n)^{1/2}$. Let Θ_1 be uniformly distributed over $\mathcal{N}$ and let Θ_2 be the zero matrix. Let μ_ρ be the distribution of $\Theta_1 - \Theta_2 = \Theta_1$, which is a probability measure on $\{\Delta \in \mathcal{S}(2a_p) : \|\Delta\|_F^2 = 2a_p\rho^2\}$, where $\mathcal{S}(2a_p)$ is the class of matrices with $2a_p$ nonzero entries. Let $d\mathrm{pr}_1\{(X^{(1)}, \ldots, X^{(n)})\}$ and $d\mathrm{pr}_2\{(Y^{(1)}, \ldots, Y^{(n)})\}$ denote the likelihood functions with partial correlation matrices Θ_1 and Θ_2, respectively. Then

$$L_{\mu_\rho} = E_{\mu_\rho}\left[\frac{d\mathrm{pr}_1\{(X^{(1)},\dots,X^{(n)})\}}{d\mathrm{pr}_2\{(Y^{(1)},\dots,Y^{(n)})\}}\right],$$

where the expectation is over the distribution of Θ_1. By the arguments in Baraud (2002), it suffices to show that $E(L^2_{\mu_\rho}) \le 1 + o(1)$.

It is easy to check that

$$L_{\mu_\rho} = E_{Y,\hat{m}}\left\{\prod_{i=1}^{n}\frac{2^p}{Z_{\hat{m}}(\Theta_{\hat{m}})}\exp\left(\rho\sum_{r\in\hat{m}}Y_r^{(i)}Y_{r+1}^{(i)}\right)\right\},$$

where $Y_1^{(i)},\dots,Y_p^{(i)}$ are independent Rademacher random variables and $Z_{\hat{m}}(\Theta_{\hat{m}})$ is the normalizing constant corresponding to $\mathrm{pr}_{\Theta_{\hat{m}}}$. Since $\Theta_{\hat{m}} \in \mathcal{N}$ and $|\hat{m}| = a_p$, we can express $Z_{\hat{m}}(\Theta_{\hat{m}})$ analytically as

$$\sum_{X_1,\dots,X_p}\exp(\rho\sum_{r\in\hat{m}}X_rX_{r+1}) = 2^p\{\cosh(\rho)\}^{a_p}.$$

Consequently, one can rewrite L_{μ_ρ} as

$$\begin{aligned}
L_{\mu_\rho} =& E_{Y,\hat{m}}\left[\prod_{i=1}^{n}\frac{2^p}{2^p\{\cosh(\rho)\}^{a_p}}\exp(\rho\sum_{r\in\hat{m}}Y_r^{(i)}Y_{r+1}^{(i)})\right]\\
=&\frac{1}{C_{p-1}^{a_p}}\sum_{m\in\mathcal{M}(a_p,p-1)}\frac{1}{\{\cosh(\rho)\}^{na_p}}E_Y\left\{\prod_{i=1}^{n}\exp(\rho\sum_{r\in m}Y_r^{(i)}Y_{r+1}^{(i)})\right\}.
\end{aligned}$$

Let $J(p,\rho) = C_{p-1}^{a_p}C_{p-1}^{a_p}\{\cosh(\rho)\}^{2na_p}$. By mutual independence of Y_r's,

$$\begin{aligned}
E(L^2_{\mu_\rho}) &= \frac{1}{J(p,\rho)}\sum_{m,m'\in\mathcal{M}(a_p,p-1)}E_Y\left\{\prod_{i=1}^{n}\exp\left(\rho\sum_{r\in m}Y_r^{(i)}Y_{r+1}^{(i)} + \rho\sum_{r\in m'}Y_r^{(i)}Y_{r+1}^{(i)}\right)\right\}\\
&= \frac{1}{J(p,\rho)}\sum_{m,m'\in\mathcal{M}(a_p,p-1)}E_Y\left\{\prod_{i=1}^{n}\exp\left(2\rho\sum_{r\in m\cap m'}Y_r^{(i)}Y_{r+1}^{(i)} + \rho\sum_{r\in m\Delta m'}Y_r^{(i)}Y_{r+1}^{(i)}\right)\right\}\\
&= \frac{1}{J(p,\rho)}\sum_{m,m'\in\mathcal{M}(a_p,p-1)}\prod_{i=1}^{n}E_Y\left\{\exp\left(2\rho\sum_{r\in m\cap m'}Y_r^{(i)}Y_{r+1}^{(i)}\right)\right\}E_Y\left\{\exp\left(\rho\sum_{r\in m\Delta m'}Y_r^{(i)}Y_{r+1}^{(i)}\right)\right\},
\end{aligned}$$

where $m\Delta m' = (m\backslash m')\cup(m'\backslash m)$. Further, one can show that

$$E_Y\Big\{\exp\left(\rho\sum_{r\in m}Y_r^{(i)}Y_{r+1}^{(i)}\right)\Big\} = \cosh^m(\rho).$$

Hence

$$\begin{aligned}
E(L^2_{\mu_\rho}) =&\frac{1}{J(p,\rho)}\sum_{m,m'\in\mathcal{M}(a_p,p-1)}\prod_{i=1}^{n}\{\cosh(2\rho)\}^{|m\cap m'|}\{\cosh(\rho)\}^{|m\Delta m'|}\\
=&\frac{1}{J(p,\rho)}\sum_{m,m'\in\mathcal{M}(a_p,p-1)}\{\cosh(2\rho)\}^{n|m\cap m'|}\{\cosh(\rho)\}^{n|m\Delta m'|}\\
=&\frac{1}{J(p,\rho)}\sum_{m,m'\in\mathcal{M}(a_p,p-1)}\left\{\frac{\cosh(2\rho)}{\cosh^2(\rho)}\right\}^{n|m\cap m'|}\{\cosh(\rho)\}^{2n|m\cap m'|+n|m\Delta m'|}\\
=&\frac{1}{C_{p-1}^{a_p}C_{p-1}^{a_p}}\sum_{m,m'\in\mathcal{M}(a_p,p-1)}\left\{\frac{2\cosh(2\rho)}{1+\cosh(2\rho)}\right\}^{n|m\cap m'|},
\end{aligned}$$

where the last equation follows from the fact that $|m \cap m'| + |m\Delta m'|/2 = a_p$. In view of the random variables $m, m' \in \mathcal{M}(a_p, p-1)$, we can further write

$$E(L^2_{\mu_\rho}) = \binom{p-1}{a_p}^{-2} \sum_{j=1}^{a_p} \binom{a_p}{j} \binom{p-1-a_p}{a_p - j} \left\{ \frac{2\cosh(2\rho)}{1+\cosh(2\rho)} \right\}^{nj}$$
$$= \{1+o(1)\} \left[1 + \frac{a_p}{p-1} \left\{ \frac{2\cosh(2\rho)}{1+\cosh(2\rho)} \right\}^n \right]^{a_p}.$$

Because $\rho = c(\log p/n)^{1/2}$, the term $\{1+\cosh(2\rho)\}^{-1}\{2\cosh(2\rho)\} \sim 1 + 2\rho^2$ by local Taylor approximation. Therefore,

$$E(L^2_{\mu_\rho}) \leq \{1+o(1)\} \exp\left[a_p \log\left\{ 1 + \frac{a_p}{p-1}(1+2\rho^2)^n \right\} \right]$$
$$\leq \{1+o(1)\} \exp\left(a_p^2 p^{2c^2-1} \right) = 1 + o(1),$$

for sufficiently small $c > 0$, where the last inequality follows from local Taylor approximation $\log(1+x) \sim x$ for x in a small neighborhood of $x_0 = 0$. □

A·4. *Proof of Theorem* 3

Proof. The key to the proof of Theorem 3 is Lemma A5. Conditioning on Lemma A5, we can follow the strategy developed in Xia et al. (2015) to establish the asymptotic control of false discovery proportion and false discovery rate in Theorem 3. □

B. ADDITIONAL TECHNICAL DETAILS

B·1. *Covariances between residuals*

For expositional clarity, consider generic parameters $(\theta_{rt})_{1\leq r<t\leq p}$ and the corresponding oracle score vectors $(v_{rt})_{1\leq r<t\leq p}$. We derive $\mathrm{cov}(\varepsilon_r, \varepsilon_t)$ for $r \neq t$.

The conditional distribution of X_r given X_{-r} is

$$\mathrm{pr}_\Theta(X_r \mid X_{-r}) = \frac{\exp(X_r \sum_{j\neq r} \theta_{rj} X_j)}{\exp(\sum_{j\neq r} \theta_{rj} X_j) + \exp(-\sum_{j\neq r} \theta_{rj} X_j)}.$$

Suppose $r < t$. The joint distribution of X_r and X_t given $X_{-\{r,t\}}$ is

$$\mathrm{pr}_\Theta(X_r, X_t \mid X_{-\{r,t\}}) = \frac{\mathrm{pr}_\Theta(X)}{\sum_{X_r}\sum_{X_t} \mathrm{pr}_\Theta(X_r, X_t, X_{-\{r,t\}})}$$
$$= \frac{\exp(\theta_{rt} X_r X_t + X_r \sum_{j\neq r,t} \theta_{rj} X_j + X_t \sum_{j\neq r,t} \theta_{tj} X_j)}{2\mathcal{D}_{rt}},$$

where $\mathcal{D}_{rt} = \exp(\theta_{rt}) \cosh\{\sum_{j\neq r,t}(\theta_{rj} + \theta_{tj})X_j\} + \exp(-\theta_{rt}) \cosh\{\sum_{j\neq r,t}(\theta_{rj} - \theta_{tj})X_j\}$. Therefore the conditional distribution of X_t given $X_{-\{r,t\}}$ is

$$\mathrm{pr}_\Theta(X_t \mid X_{-\{r,t\}}) = \frac{\mathrm{pr}_\Theta(X_r, X_t \mid X_{-\{r,t\}})}{\mathrm{pr}_\Theta(X_r \mid X_{-\{r\}})}$$
$$= \frac{\exp(X_t \sum_{j\neq r,t} \theta_{tj} X_j) \cosh(\sum_{j\neq r} \theta_{rj} X_j)}{\mathcal{D}_{rt}},$$

which again follows a logistic regression form.

Recall the residual ε_r is defined as

$$\varepsilon_r = X_r - \dot{f}(X_{-r}\theta_r),$$

whose conditional variance $\mathrm{var}(\varepsilon_r \mid X_{-r}) = \ddot{f}(X_{-r}\theta_r)$. Since $E(\varepsilon_r) = 0$ for all $r = 1, \ldots, p$, it suffices to calculate $E(\varepsilon_r \varepsilon_t)$. Let $u_{\{r,-t\}} = X_{-\{r,t\}}\theta_{\{r,-t\}}$. By the joint distribution of X_r and X_t given $X_{-\{r,t\}}$,

$$E(\varepsilon_r \varepsilon_t \mid X_{-\{r,t\}}) = \frac{A}{B}, \tag{B1}$$

where

$$\begin{aligned} A =& \sinh(-2\theta_{rt})\{e^{-\theta_{rt}} \cosh(u_{\{t,-r\}} + u_{\{r,-t\}}) + e^{\theta_{rt}} \cosh(u_{\{t,-r\}} - u_{\{r,-t\}})\}, \\ B =& 2\mathcal{D}_{rt} \cosh(\theta_{rt} + u_{\{r,-t\}}) \cosh(-\theta_{rt} + u_{\{r,-t\}}) \cosh(\theta_{tr} + u_{\{t,-r\}}) \cosh(-\theta_{tr} + u_{\{t,-r\}}). \end{aligned}$$

Further, we can calculate $E\{\ddot{f}(X_{-r}\theta_r) \mid X_{-\{r,t\}}\}$, which yields

$$\frac{1}{2}E\big\{\ddot{f}(X_{-r}\theta_r) \mid X_{-\{t,r\}}\big\} \sinh(-2\theta_{rt}) = \frac{A}{B}\Big\{ \cosh(\theta_{rt} + u_{\{t,-r\}}) \cosh(-\theta_{rt} + u_{\{t,-r\}})\Big\}. \tag{B2}$$

Combining (B1) and (B2), the desired conditional expectation becomes

$$E(\varepsilon_r \varepsilon_t \mid X_{-\{r,t\}}) = \frac{1}{2} \frac{\sinh(-2\theta_{rt}) E\big\{\ddot{f}(X_{-r}\theta_r) \mid X_{-\{r,t\}}\big\}}{\cosh(\theta_{rt} + u_{\{t,-r\}}) \cosh(-\theta_{rt} + u_{\{t,-r\}})}. \tag{B3}$$

The right-hand side of (B3) still involves one conditional expectation. It is thus not straightforward to obtain an explicit expression of $\mathrm{cov}(\varepsilon_r, \varepsilon_t)$. Nonetheless, we can obtain the following upper bound using the fact that $\cosh(x) \geq 1$:

$$|\mathrm{cov}(\varepsilon_r, \varepsilon_t)| = \big|E\left\{E(\varepsilon_r \varepsilon_t \mid X_{-\{r,t\}})\right\}\big| \leq \frac{1}{2}|\sinh(2\theta_{rt})|.$$

B·2. *Covariances between test statistics*

To complete the proof of Lemma 1, we show (A3) by calculating $\mathrm{cov}(v_{rt}\varepsilon_r, v_{ml}\varepsilon_m)$ for distinct indices r, t, m and l. For expositional clarity, we introduce several simplified notation: $\theta_{rt} = a, \theta_{rm} = b, \theta_{tm} = c, \theta_{rl} = d, \theta_{tl} = e$ and $\theta_{ml} = f$. For $k \in \{r, t, m, l\}$, let $A_k = \sum_{j:j\neq\{r,t,m,l\}} \theta_{kj} X_j$. The full conditional distribution is

$$\mathrm{pr}_\Theta \left(X_r, X_t, X_m, X_l \mid X_{-\{r,t,m,l\}}\right) = \frac{\mathcal{D}_{rtml}}{\sum_{X_r,X_t,X_m,X_l \in\{-1,1\}^4} \mathcal{D}_{rtml}},$$

where the numerator

$$\mathcal{D}_{rtml} = \exp\{X_r(aX_t + bX_m + dX_l + A_r) + X_t(cX_m + eX_l + A_t) + X_m(fX_l + A_m) + X_l A_l\}.$$

To evaluate the desired correlation, we need $E(v_{rt}v_{ml}\varepsilon_r\varepsilon_m \mid X_{-\{r,t,m,l\}})$. Let

$$\mathcal{D}_{ml}(X_r, X_t) = \sum_{\delta=\pm 1} e^{-\delta(A_l + dX_r + eX_t)} \cosh(\delta f + A_m + bX_r + cX_t).$$

By definition, $\mathcal{D}_{ml}(X_r, X_t) \geq 2\cosh\left(A_l + dX_r + eX_t\right) \geq 2$. Expanding terms in the conditional expectation yields

$$E\left(v_{rt}\varepsilon_r v_{ml}\varepsilon_m \mid X_{-\{r,t,m,l\}}\right) = \frac{E\left\{(Q + \gamma X_m X_l) v_{rt}\varepsilon_r \mid X_{-\{r,t,m,l\}}\right\}}{\mathcal{D}_{ml}(1,1)\mathcal{D}_{ml}(-1,1)\mathcal{D}_{ml}(1,-1)\mathcal{D}_{ml}(-1,-1)}, \tag{B4}$$

where

$$
\begin{aligned}
Q =&\nu \sinh(2X_mA_m - 2X_lA_l) + (\eta - \xi)X_m \sinh(2fX_l - 2A_m)\\
&+ 2\sinh(2b)\sinh(2c)e^{-2X_lA_l}\{e^{2A_l}\cosh(-2f+2A_m) + e^{-2A_l}\cosh(2f+2A_m)\}\\
&+ (\alpha-\beta)X_l \sinh(2fX_m - 2A_l)\\
&+ 2\sinh(2d)\sinh(2e)e^{-2X_mA_m}\{e^{2A_m}\cosh(-2f+2A_l) + e^{-2A_m}\cosh(2f+2A_l)\}\\
&+ (\mu-\psi)e^{-2fX_mX_l}\{e^{2f}\cosh(2A_l-2A_m) + e^{-2f}\cosh(2A_l+2A_m)\}\\
&+ (\eta+\xi)X_mX_l\cosh(2fX_l-2A_m) + (\alpha+\beta)X_mX_l\cosh(2fX_m-2A_l)\\
&+ (\mu+\psi)X_mX_le^{-2fX_mX_l}\{e^{2f}\cosh(2A_l-2A_m) + e^{-2f}\cosh(2A_l+2A_m)\},
\end{aligned}
$$

and

$$
\begin{aligned}
\gamma &= 4\sinh(2b)\sinh(2e)\cosh(2b)\cosh(2e) + 4\sinh(2c)\sinh(2d)\cosh(2c)\cosh(2d),\\
\nu &= 4\sinh(2b)\sinh(2c)\cosh(2d)\cosh(2e) - 4\sinh(2d)\sinh(2e)\cosh(2b)\cosh(2c),\\
\eta-\xi &= -2\sinh(2b)\sinh(2c)\{\cosh(4d)+\cosh(4e)\},\\
\alpha-\beta &= -2\sinh(2d)\sinh(2e)\{\cosh(4b)+\cosh(4c)\},\\
\mu-\psi &= 2\sinh(2b)\sinh(2c)\cosh(2d)\cosh(2e) + 2\sinh(2d)\sinh(2e)\cosh(2b)\cosh(2c),\\
\eta+\xi &= 4\sinh(2b)\sinh(2e)\cosh(2c)\cosh(2e) + 4\sinh(2c)\sinh(2d)\cosh(2b)\cosh(2d),\\
\alpha+\beta &= 4\sinh(2b)\sinh(2e)\cosh(2b)\cosh(2d) + 4\sinh(2c)\sinh(2d)\cosh(2c)\cosh(2e),\\
\mu+\psi &= 2\sinh(2b)\sinh(2e)\cosh(2c)\cosh(2d) + 2\sinh(2c)\sinh(2d)\cosh(2b)\cosh(2e).
\end{aligned}
$$

Since $|v_{rt}\varepsilon_r| \le 2$ and $\mathcal{D}_{ml}(X_r, X_t) \ge 2$, applying Jensen's inequality to (B4) implies

$$|\mathrm{cov}(v_{rt}\varepsilon_r, v_{ml}\varepsilon_m)| = |E\left(v_{rt}\varepsilon_r v_{ml}\varepsilon_m\right)| \le \frac{1}{8}E\big(|Q| + |\gamma|\big).$$

In the following, we bound each term in Q for Θ satisfying Assumption 2. To this end, we combine terms and write Q as a function of $\sinh(2b)\sinh(2c), \sinh(2d)\sinh(2e), \sinh(2b)\sinh(2e)$ and $\sinh(2c)\sinh(2d)$.

Specifically, the multipliers of $\sinh(2b)\sinh(2c)$ are bounded by

$$
\begin{aligned}
|4\cosh(2d)\cosh(2e)\sinh(2X_mA_m - 2X_lA_l)| &\le 4\sinh(4C_w),\\
2\{\cosh(4d)+\cosh(4e)\}|\sinh(2fX_l-2A_m)| &\le 4(C_0-1)^2\sinh(2C_w),\\
2e^{-2X_lA_l}\{e^{2A_l}\cosh(-2f+2A_m) + e^{-2A_l}\cosh(2f+2A_m)\} &\le 8(C_0-1)^3,\\
2\cosh(2d)\cosh(2e)e^{-2fX_mX_l}\{e^{2f}\cosh(2A_l-2A_m) + e^{-2f}\cosh(2A_l+2A_m)\} &\le 8(C_0-1)^2.
\end{aligned}
$$

And the multipliers of $\sinh(2b)\sinh(2e)$ are bounded by

$$
\begin{aligned}
4\cosh(2c)\cosh(2e)\cosh(2fX_l-2A_m) &\le 4(C_0-1)^2,\\
4\cosh(2b)\cosh(2d)\cosh(2fX_m-2A_l) &\le 4(C_0-1)^2,\\
2\cosh(2c)\cosh(2d)e^{-2fX_mX_l}\{e^{2f}\cosh(2A_l-2A_m) + e^{-2f}\cosh(2A_l+2A_m)\} &\le 8(C_0-1)^2.
\end{aligned}
$$

Similarly, one can bound the multipliers of $\sinh(2d)\sinh(2e)$ and $\sinh(2c)\sinh(2d)$. After combining terms, we can bound $|Q| + |\gamma|$ by

$$
\begin{aligned}
|Q| + |\gamma| \le &\sum_{k=m,l} J_{rk,tk}|\sinh(2\theta_{rk})\sinh(2\theta_{tk})|\\
&+ J_{rm,tl}|\sinh(2\theta_{rm})\sinh(2\theta_{tl})| + J_{rl,tm}|\sinh(2\theta_{rl})\sinh(2\theta_{tm})|,
\end{aligned}
$$

where for $k \in \{m, l\}$,

$$J_{rk,tk} = 4\sinh(4C_w) + 4(C_0-1)^2\sinh(2C_w) + 8(C_0-1)^3 + 8(C_0-1)^2$$
$$= 8\sinh(2C_w)(C_0-1) + 4(C_0-1)^2\{\sinh(2C_w) + 2C_0\},$$

and for $(k, k') \in \{(m, l), (l, m)\}$,

$$J_{rk,tk'} = 16(C_0-1)^2 + 4\cosh(2\theta_{tk'})\cosh(2\theta_{rk}) \leq 20(C_0-1)^2.$$

Let $C_1 = \max\{J_{rm,tm}/2, 10(C_0-1)^2\}$. The desired correlation is thus

$$|\mathrm{cor}(v_{rt}\varepsilon_r, v_{ml}\varepsilon_m)| = \frac{4|\mathrm{cov}(v_{rt}\varepsilon_r, v_{ml}\varepsilon_m)|}{(F_{rt}F_{ml})^{1/2}}$$
$$\leq \frac{1}{2c^*}E(|Q| + |\gamma|)$$
$$\leq \frac{C_1}{c^*}\left\{\sum_{a=m,l}|\sinh(2\theta_{ra})\right\}\left\{\sum_{b=m,l}\sinh(2\theta_{tb})|\right\}.$$

REFERENCES

BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8**, 577–606.

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. (2016). Large-scale multiple testing of correlations. *J. Am. Statist. Assoc.* **111**, 229–240.

CAI, T. T., LIU, W. & XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Am. Statist. Assoc.* **108**, 265–277.

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

XIA, Y., CAI, T. & CAI, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102**, 247–266.

ZAITSEV, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of sn bernstein's inequality conditions. *Probab. Theory Relat. Fields* **74**, 535–566.

ZHANG, C.-H. & ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Statist. Soc.* B **76**, 217–242.

[*Received 1 Jun* 2017. *Revised 21 Sep* 2018]