Appendix for “On high-dimensional constrained maximum likelihood inference”

A Technical details of the counter example

Lemma 1 *(A counter example) In (5) in the main text, we write $y = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}$, where $\boldsymbol{x} = (x_1, \ldots, x_p)$ are independently distributed from $N(\mu_i, 1)$ with $\mu_1 = 0$ and $\mu_j = 1$; $2 \le j \le p$, and ϵ is $N(0, 1 - n^{-1})$, independent of $\boldsymbol{x}$. Assume that $\beta_0 = 0$ and $\boldsymbol{\beta} = n^{-1/2}, 0, \ldots, 0)$, or, $y = n^{-1/2} x_1 + \epsilon$. Then Assumption 3 is violated. Now consider a hypothesis test of $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$. If $\frac{\log p}{n} \to 0$ as $n, p \to \infty$, then $\Lambda_n(B) \xrightarrow{p} \infty$ as $n, p \to \infty$, with $B = \{0\}$.*

Proof of Lemma 1. Under the linear model, we have that

$$y_i = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}_i + \epsilon_i; i = 1, \ldots, n\,, \tag{A.1}$$

where $\boldsymbol{\beta} = (\beta_1, 0, \cdots, 0)$ and $\beta_0 = 0$, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip}) \sim N(\boldsymbol{\mu}, \boldsymbol{I}_{p\times p})$, and $\epsilon_i \sim N(0, 1 - \beta_1^2)$ and is independent of $\boldsymbol{x}_i$. Then, the constrained MLE for β_0 is

$$\hat{\beta}_0^{(1)} = \operatorname*{argmin}_{\sum_{i=1}^p \mathbb{I}(\beta_i \neq 0) \le 1} \sum_{i=1}^n (y_i - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{x}_i)^2 = \bar{y} - \widehat{\text{cor}}(x_{\cdot j^\star}, y) \frac{s_y}{s_{x_{\cdot j^\star}}} \bar{x}_{\cdot j^\star}\,, \tag{A.2}$$

where $x_{\cdot j}$ denotes a n-dimensional vector $(x_{1j}, \ldots, x_{nj})$, $\widehat{\text{cor}}$ denotes the sample correlation between two vectors, $\bar{x}$ and s_x denote the sample mean and sample covariance of a vector x, respectively, and

$$j^\star = \underset{1\le j\le p}{\operatorname{argmax}}\ \widehat{\operatorname{cor}}(x_{\cdot j}, y) \tag{A.3}$$

denotes the index of which feature has the largest sample correlation between y. For each observation $(y_i, \boldsymbol{x}_i)$, it is easy to write out its joint distribution

$$(y_i, x_{i1}, \ldots, x_{ip}) \sim N\left((\beta_1\mu_1, \mu_1, \ldots, \mu_p)^\top, \begin{pmatrix} 1 & \beta_1 & 0 & \cdots & 0 \\ \beta_1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}\right). \tag{A.4}$$

Hence, the conditional distribution of $\boldsymbol{x}_i$ given y_i is

$$\boldsymbol{x}_i|y_i \sim N\left((\beta_1(y_i - \beta_1\mu_1) + \mu_1, \mu_2, \ldots, \mu_p)^\top, \begin{pmatrix} 1-\beta_1^2 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}\right) \tag{A.5}$$

from which we can easily see that components of $\boldsymbol{x}_i$ are conditionally independent given y_i. Note that

$$\widehat{\operatorname{cor}}(x_{\cdot j}, y) = \frac{(n-1)^{-1}\sum_{i=1}^n x_{ij}(y_i - \bar{y})}{s_{\cdot j}s_y}, j = 1, \ldots, p \tag{A.6}$$

and $\operatorname{Var}(y) = \operatorname{Var}(x_{ij}) = 1$. Hence,

$$\sqrt{n}\widehat{\operatorname{cor}}(x_{\cdot j}, y)|y \overset{d}{=} Z_j + o_p(1)\,, \tag{A.7}$$

where $Z_j = \frac{\sum_{i=1}^n x_{ij}(y_i - \bar{y})}{(n-1)s_y}$, $j = 1, \ldots, p$, and Z_j's are independent and normally distributed conditioned on $\boldsymbol{y}$. By (A.5), we have that

$$Z_1 \sim N\left(\beta_1 s_y, 1-\beta_1^2\right) \text{ and } Z_j \sim N(0,1) \text{ for } j=2,\dots,p\,. \tag{A.8}$$

Consequently, conditioned on $\boldsymbol{y}$,

$$\hat{\beta}_0^{(1)} = \bar{y} - \widehat{\text{cor}}(x_{\cdot j^\star}, y)\frac{s_y}{s_{x_{\cdot j^\star}}}\bar{x}_{\cdot j^\star} = \bar{y} - \beta_1\mu_1 + \beta_1\mu_1 - \widehat{\text{cor}}(x_{\cdot j^\star}, y) s_y \frac{\bar{x}_{\cdot j^\star} - \mu_{j^\star}}{s_{x_{\cdot j^\star}}} - \widehat{\text{cor}}(x_{\cdot j^\star}, y) s_y \frac{\mu_{j^\star}}{s_{x_{\cdot j^\star}}}$$

Now, we let $\mu_1 = 0$ and $\mu_2 = \cdots = \mu_p = 1$. Moreover, note that

$$\bar{y} - \beta_1\mu_1 = O_p\left(\frac{1}{\sqrt{n}}\right) \text{ and } \left|\frac{\bar{x}_{j^\star} - \mu_{j^\star}}{s_{x_{j^\star}}}\right| \le \max_{1\le j\le p}\left|\frac{\bar{x}_j - \mu_j}{s_{x_j}}\right| \le O\left(\sqrt{\frac{\log p}{n}}\right)\,. \tag{A.9}$$

Hence, if $\sqrt{\frac{\log p}{n}} \le O(1)$, then

$$\hat{\beta}_0^{(1)} = -\widehat{\text{cor}}(x_{\cdot j^\star}, y) s_y \frac{\mu_{j^\star}}{s_{x_{\cdot j^\star}}} + O_p\left(\frac{1}{\sqrt{n}}\right)\,. \tag{A.10}$$

Now we choose β_1 to be small number so that with nonzero probability $\{j^\star \neq 1\}$, that is, we need $\mathbb{P}(Z_1 \le \min_{2\le j\le p} Z_j)$ to be nonzero, which is easy to achieve when β_1 is chosen to be close to 0. Under the event $\{j^\star \ge 2\}$

$$\begin{aligned}
\hat{\beta}_0^{(1)} &= -\widehat{\text{cor}}(x_{\cdot j^\star}, y) s_y \frac{\mu_{j^\star}}{s_{x_{\cdot j^\star}}} + O_p\left(\frac{1}{\sqrt{n}}\right) = -\max_{2\le j\le p}\widehat{\text{cor}}(x_{\cdot j}, y)\frac{s_y}{s_{x_{\cdot j^\star}}} + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= O_p\left(\sqrt{\frac{\log p}{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\,,
\end{aligned}$$

because $\max_{2\le j\le p}\widehat{\text{cor}}(x_{\cdot j}, y) = O_p\left(\sqrt{\frac{\log p}{n}}\right)$ and $s_y \to 1$ in probability and $s_{x_{\cdot j^\star}} \to 1$ in probability. Hence, $n\left(\hat{\beta}_0^{(1)}\right)^2 \to \infty$ if $p \to \infty$ as $n \to \infty$. Next, we show that under this model, the log-likelihood ratio test statistic is of the same order as $n\hat{\beta}_0^2$ under the null model.

Toward this end, denote by $f(\beta_0) = \sup_{\|\boldsymbol{\beta}\|_0 \le 1, \sigma > 0} n^{-1} L_n(\beta_0, \boldsymbol{\beta}, \sigma)$. By definition of $\hat{\beta}_0^{(1)}$, it must maximizes $f(\beta_0)$ as a function of β_0 and hence must satisfies $f'(\hat{\beta}_0^{(1)}) = 0$. Moreover, we note that the log-likelihood ratio can be rewritten in terms of $f(\cdot)$

$$\Lambda_n(B) = 2n(f(\hat{\beta}_0^{(1)}) - f(0)) \tag{A.11}$$

Applying a Taylor expansion around $\hat{\beta}_0^{(1)}$, we obtain

$$\Lambda_n(B) = -n(\hat{\beta}_0^{(1)})^2 f''(\beta^\star) \tag{A.12}$$

where $\beta^\star$ is some number between 0 and $\hat{\beta}_0^{(1)}$. Under $\log p/n \to 0$, it is easy to show that $\hat{\beta}_0^{(1)}$ is consistent, hence converges to 0 in probability. Hence, $\Lambda_n(B) = -n(\hat{\beta}_0^{(1)})^2 (f''(0) + o_p(1)) \xrightarrow{\mathbb{P}} \infty$, which completes the proof.

B Proofs of Lemmas 2-9

This section provides detailed proofs of Lemmas 2-9 to be used in "On high-dimensional constrained maximum likelihood inference".

Lemma 2 *For any symmetric matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$, $\mathrm{vec}(\boldsymbol{C}_1)^\top \mathrm{vec}(\boldsymbol{C}_2) = \mathrm{tr}(\boldsymbol{C}_1 \boldsymbol{C}_2)$. Moreover, for any positive definite matrix $\boldsymbol{C} \succ 0$,*

$$\nabla \left(\log \det \boldsymbol{C}\right) = -\,\mathrm{vec}(\boldsymbol{C}^{-1})\,, \quad \nabla^2 \left(-\log \det \boldsymbol{\Omega}^0\right) = \boldsymbol{C}^{-1} \otimes_s \boldsymbol{C}^{-1}\,, \tag{B.1}$$

$$\boldsymbol{I} = \tfrac{1}{2} \boldsymbol{\Sigma}^0 \otimes_s \boldsymbol{\Sigma}^0, \tag{B.2}$$

$$\mathrm{Var}\left(\mathrm{vec}(\boldsymbol{X}\boldsymbol{X}^\top)\right) = 4\boldsymbol{I} \ \textit{with} \ \boldsymbol{X} \sim N(0, \boldsymbol{\Sigma}^0)\,, \tag{B.3}$$

$$\mathrm{vec}(\boldsymbol{C})^\top \boldsymbol{I}\, \mathrm{vec}(\boldsymbol{C}) = \tfrac{1}{2} \mathrm{tr}\left(\boldsymbol{\Sigma}^0 \boldsymbol{C} \boldsymbol{\Sigma}^0 \boldsymbol{C}\right)\,. \tag{B.4}$$

Proof of Lemma 2: By the definition, (B.1) follows from an identity:

$$\mathrm{vec}(\boldsymbol{C}_1)^\top \mathrm{vec}(\boldsymbol{C}_2) = \sum_{i\le j}(1+\mathbb{I}(i\neq j))\boldsymbol{S}_1(i,j)\boldsymbol{S}_2(i,j) = \sum_{i,j}\boldsymbol{S}_1(i,j)\boldsymbol{S}_2(i,j) = \mathrm{tr}(\boldsymbol{S}_1\boldsymbol{S}_2)\,.$$

Moreover, it follows from Taylor's expansion of the log det function that

$$\begin{aligned}
&\log\det(\boldsymbol{C}+\boldsymbol{\Delta}) - \log\det(\boldsymbol{C}) = \mathrm{tr}(\boldsymbol{C}^{-1}\boldsymbol{\Delta}) - \frac{1}{2}\,\mathrm{tr}\left((\boldsymbol{C}^{-1}\boldsymbol{\Delta})^2\right) + o(\|\boldsymbol{C}^{-1/2}\boldsymbol{\Delta}\boldsymbol{C}^{-1/2}\|_F^2)\\
= \;&\mathrm{vec}(\boldsymbol{C}^{-1})^\top \mathrm{vec}(\boldsymbol{\Delta}) - \frac{1}{2}\mathrm{vec}(\boldsymbol{\Delta})^\top \mathrm{vec}(\boldsymbol{C}^{-1}\boldsymbol{\Delta}\boldsymbol{C}^{-1}) + o(\|\boldsymbol{C}^{-1/2}\boldsymbol{\Delta}\boldsymbol{C}^{-1/2}\|_F^2)\\
= \;&\mathrm{vec}(\boldsymbol{C}^{-1})^\top \mathrm{vec}(\boldsymbol{\Delta}) - \frac{1}{2}\mathrm{vec}(\boldsymbol{\Delta})^\top \left(\boldsymbol{C}^{-1}\otimes_s \boldsymbol{C}^{-1}\right)\mathrm{vec}(\boldsymbol{\Delta}) + o(\|\boldsymbol{C}^{-1/2}\boldsymbol{\Delta}\boldsymbol{C}^{-1/2}\|_F^2)\,,
\end{aligned}$$

where the definition of $\otimes_s$ and (B.1) have been used. This yields (B.2).

For (B.3), the log-likelihood for $\boldsymbol{X}\sim N(0,\boldsymbol{\Sigma}^0)$ is $-\frac{1}{2}\mathrm{vec}(\boldsymbol{\Omega}^0)^\top\mathrm{vec}(\boldsymbol{X}\boldsymbol{X}^\top)+\frac{1}{2}\log\det(\boldsymbol{\Omega}^0)$. Using properties of the exponential family [2], $\mathrm{Var}\left(\frac{1}{2}\mathrm{vec}(\boldsymbol{X}\boldsymbol{X}^\top)\right) = \nabla^2\left(-\frac{1}{2}\log\det\boldsymbol{\Omega}^0\right) = \boldsymbol{I}$, implying (B.3). Finally, for any symmetric matrix $\boldsymbol{C}$, note that

$$\begin{aligned}
\mathrm{vec}(\boldsymbol{C})^\top\boldsymbol{I}\,\mathrm{vec}(\boldsymbol{C}) &= \frac{1}{2}\mathrm{vec}(\boldsymbol{C})^\top\left(\boldsymbol{\Sigma}^0\otimes_s\boldsymbol{\Sigma}^0\right)\mathrm{vec}(\boldsymbol{C})\\
&= \frac{1}{2}\mathrm{vec}(\boldsymbol{C})^\top\mathrm{vec}(\boldsymbol{\Sigma}^0\boldsymbol{C}\boldsymbol{\Sigma}^0) = \frac{1}{2}\mathrm{tr}(\boldsymbol{C}\boldsymbol{\Sigma}^0\boldsymbol{C}\boldsymbol{\Sigma}^0),
\end{aligned}$$

leading to (B.4). This completes the proof.

Lemma 3 *For any symmetric matrix* $\boldsymbol{T}$ *and* $\nu>0$

$$\mathbb{P}\left(|\,\mathrm{tr}\left((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right)|\ge\nu\right)\le 2\exp\left(-n\tfrac{\nu^2}{9\|\boldsymbol{T}\|^2+8\nu\|\boldsymbol{T}\|}\right)\,, \tag{B.5}$$

where $\|\boldsymbol{T}\|^2=\frac{n}{2}\mathrm{Var}\left(\mathrm{tr}\left((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right)\right)$. *Furthermore, for* $\boldsymbol{T}_1,\cdots,\boldsymbol{T}_K$ *such that* $\|\boldsymbol{T}_k\|\le c_0; k=$

$1, \cdots, K$ *with* $c_0 > 0$ *and any* $\nu > 0$, *we have that*

$$\mathbb{P}\left(\max_{1\le k\le K}\left|\mathrm{tr}((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}_k)\right| \ge \nu\right) \le 2\exp\left(-n\frac{\nu^2}{9c_0^2+8c_0\nu}+\log K\right), \qquad \text{(B.6)}$$

which implies that $\max_{1\le k\le K}|\mathrm{tr}((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}_k)| = O_p\left(c_0\sqrt{\frac{\log K}{n}}\right)$. *Particularly, for any* $\nu > 0$ *and any index set* B,

$$\mathbb{P}\left(\|\operatorname{vec}_B(\boldsymbol{S}-\boldsymbol{\Sigma}^0)\|_\infty \ge \nu\right) \le 2\exp\left(-n\frac{\nu^2}{9\lambda_{max}^2(\boldsymbol{\Sigma}^0)+8\nu\lambda_{max}(\boldsymbol{\Sigma}^0)}+\log|B|\right), \qquad \text{(B.7)}$$

implying that $\|\operatorname{vec}_B(\boldsymbol{S}-\boldsymbol{\Sigma}^0)\|_\infty = O_p\left(\lambda_{max}(\boldsymbol{\Sigma}^0)\sqrt{\frac{\log|B|}{n}}\right)$.

Proof of Lemma 3: By Markov's inequality, for any $\nu > 0$,

$$\begin{aligned} P\left(\mathrm{tr}\left((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right) \ge \nu\right) &\le \exp\left(-\frac{\gamma\sqrt{n}\nu}{2}\right)\mathbb{E}\exp\left(\frac{\gamma\sqrt{n}}{2}\,\mathrm{tr}\left((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right)\right) \\ &\le \exp\Bigg(\underbrace{\log\mathbb{E}\exp\left(\frac{\gamma\sqrt{n}}{2}tr\left((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right)\right)-\frac{\gamma\sqrt{n}\nu}{2}}_{I_1}\Bigg), \end{aligned}$$

where γ is chosen such that $\gamma \in \left[0, \frac{M_0\sqrt{n}}{\|\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{T}\sqrt{\boldsymbol{\Sigma}^0}\|_F}\right]$ for some constant $0 < M_0 < 1$, which is to be determined later. Moreover, after some calculations, we have that

$$\begin{aligned} \mathbb{E}\exp\left(\frac{\gamma\sqrt{n}}{2}\,\mathrm{tr}\left((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right)\right) &= \left(\mathbb{E}\exp\left(\frac{\gamma\sqrt{n}}{2}\,\mathrm{tr}\left((\boldsymbol{X}\boldsymbol{X}^T-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right)\right)\right)^n \\ &= \exp\left(-\frac{\gamma\sqrt{n}}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}^0\boldsymbol{T})\right)\det\left(\boldsymbol{I}-\frac{\gamma}{\sqrt{n}}\boldsymbol{\Sigma}^0\boldsymbol{T}\right)^{-n/2}, \end{aligned} \qquad \text{(B.8)}$$

where $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^0)$ and the last equality requires that $\sqrt{n}\boldsymbol{\Omega}^0 \succeq \gamma\boldsymbol{T}$, which is ensured by the fact that $\gamma \le \frac{M_0\sqrt{n}}{\|\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{T}\sqrt{\boldsymbol{\Sigma}^0}\|_F} < \frac{\sqrt{n}}{\|\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{T}\sqrt{\boldsymbol{\Sigma}^0}\|_F}$. Consequently,

$$\log\mathbb{E}\exp\left(\frac{\gamma\sqrt{n}}{2}\,\mathrm{tr}\left((\boldsymbol{S}-\boldsymbol{\Sigma}^0)\boldsymbol{T}\right)\right) = \log\det\left(\boldsymbol{I}-\frac{\gamma}{\sqrt{n}}\boldsymbol{\Sigma}^0\boldsymbol{T}\right)^{-n/2} - \frac{\gamma\sqrt{n}}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}^0\boldsymbol{T}). \qquad \text{(B.9)}$$

An expansion of the log det function gives

$$\begin{aligned} & \log\det(\boldsymbol{I} - \frac{\gamma}{\sqrt{n}}\boldsymbol{\Sigma}^0\boldsymbol{T})^{-n/2} \\ = \quad & \frac{\gamma\sqrt{n}}{2}\operatorname{tr}(\boldsymbol{\Sigma}^0\boldsymbol{T}) + \frac{\gamma^2}{4}\operatorname{tr}((\boldsymbol{\Sigma}^0\boldsymbol{T})^2) + \underbrace{\frac{n}{2}\sum_{l=3}^{\infty} l^{-1}\operatorname{tr}\big((\frac{\gamma\boldsymbol{\Sigma}^0\boldsymbol{T}}{\sqrt{n}})^l\big)}_{I_2}. \end{aligned} \tag{B.10}$$

For I_2, note that $I_2 \le \frac{n}{2}\sum_{l=3}^{\infty} l^{-1}\left(\frac{\gamma\|\boldsymbol{T}\|}{\sqrt{n}}\right)^l \le \gamma^2\|\boldsymbol{T}\|^2\frac{3-M_0}{12(1-M_0)}$. Similarly, $I_1 \le \frac{M_1+1}{4}\gamma^2\|\boldsymbol{T}\|^2 - \frac{\gamma\sqrt{n}\nu}{2}$, where $M_1 = \frac{3-M_0}{3(1-M_0)}$. Minimizing this upper bound of I_1 as a function of γ over the interval $\left[0, \frac{M_0\sqrt{n}}{\|\boldsymbol{T}\|}\right]$, we obtain that

$$\begin{array}{ll} I_1 \le -\frac{n\nu^2}{4(1+M_1)\|\boldsymbol{T}\|^2} & \text{if } \nu \le M_0(1+M_1)\|\boldsymbol{T}\| \\ I_1 \le -\frac{nM_0}{2\|\boldsymbol{T}\|}\left(\nu - \frac{M_0(1+M_1)}{2}\|\boldsymbol{T}\|\right) & \text{otherwise.} \end{array}$$

A combination of these two cases yields that $I_1 \le -\frac{nM_0\nu^2}{4M_0(M_1+1)\|\boldsymbol{T}\|^2+2\nu\|\boldsymbol{T}\|}$. Set $M_0 = 4^{-1}$, and then $M_1 = 11/9$, we obtain the desired results

$$P\Big(\operatorname{tr}\big((\boldsymbol{S} - \boldsymbol{\Sigma}^0)\boldsymbol{T}\big) \ge \nu\Big) \le \exp\Big(-n\frac{\nu^2}{9\|\boldsymbol{T}\|^2 + 8\nu\|\boldsymbol{T}\|}\Big),$$

for any $\nu > 0$. The other direction follows exactly the same argument, and thus is omitted.

Finally, (B.7) follows by letting $\{\boldsymbol{T}_1, \cdots, \boldsymbol{T}_k\} = \big\{(\boldsymbol{e}_i^\top\boldsymbol{e}_j + \boldsymbol{e}_j^\top\boldsymbol{e}_i)/2\big\}_{(i,j)\in B}$ then applying an inequality $\|\sqrt{\boldsymbol{\Sigma}^0}(\boldsymbol{e}_i^\top\boldsymbol{e}_j + \boldsymbol{e}_j^\top\boldsymbol{e}_i)\sqrt{\boldsymbol{\Sigma}^0}/2\|_F^2 \le \lambda_{\max}(\boldsymbol{\Sigma}^0)$ and a union bound. This completes the proof.

Lemma 4 *(The Kullback-Leibler divergence and Fisher-norm) For a positive definite matrix*

$\boldsymbol{\Omega}$ *the following connection holds:*

$$K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega}) \;\geq\; \min\left(\frac{1}{16\sqrt{2}}, \frac{\sqrt{K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega})}}{2\sqrt{6}}\right) \|\boldsymbol{\Omega} - \boldsymbol{\Omega}^0\|, \tag{B.11}$$

$$K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega}) \;\geq\; \min\left(\frac{1}{16\sqrt{2}}, \frac{\|\boldsymbol{\Omega} - \boldsymbol{\Omega}^0\|}{24}\right) \|\boldsymbol{\Omega} - \boldsymbol{\Omega}^0\|. \tag{B.12}$$

Proof of Lemma 4: Let $\boldsymbol{\Delta} = \boldsymbol{\Omega} - \boldsymbol{\Omega}^0$ and $\lambda_1, \cdots, \lambda_p$ be the eigenvalues of $\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{\Delta}\sqrt{\boldsymbol{\Sigma}^0}$. Then $\lambda_j > -1$; $j = 1, \cdots, p$, because $\boldsymbol{I}_{p\times p} + \sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{\Delta}\sqrt{\boldsymbol{\Sigma}^0} = \sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{\Omega}\sqrt{\boldsymbol{\Sigma}^0}$ is positive definite. Moreover, let $B_1 = \sum_{i=1}^p \lambda_i^2 \mathbb{I}(\lambda_i \leq 1/3)$, $B_2 = \sum_{i=1}^p \lambda_i^2 \mathbb{I}(\lambda_i > 1/3)$, and $B_3 = \sum_{i=1}^p \lambda_i \mathbb{I}(\lambda_i > 1/3)$. Easily, $\|\boldsymbol{\Omega} - \boldsymbol{\Omega}^0\| = \sqrt{B_1 + B_2}$. Using the inequality $x - \log(1+x) \geq 6^{-1}x^2\mathbb{I}(x \leq 1/3) + 8^{-1}x\mathbb{I}(x > 1/3)$ for $x > -1$, we have that

$$\begin{aligned} K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega}) &= \frac{1}{2}\left(\mathrm{tr}(\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{\Delta}\sqrt{\boldsymbol{\Sigma}^0}) - \log\det(\boldsymbol{I}_{p\times p} + \sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{\Delta}\sqrt{\boldsymbol{\Sigma}^0})\right) \\ &= \frac{1}{2}\sum_{i=1}^p \lambda_i - \frac{1}{2}\sum_{i=1}^p \log(1+\lambda_i) \\ &\geq 12^{-1}\sum_{i=1}^p \lambda_i^2 \mathbb{I}(\lambda_i \leq 1/3) + 16^{-1}\sum_{i=1}^p \lambda_i \mathbb{I}(\lambda_i > 1/3) = 12^{-1}B_1 + 16^{-1}B_3\,. \end{aligned}$$

Next we examine two cases. First, if $B_1 < B_2$, then $\frac{K(\boldsymbol{\Omega}^0,\boldsymbol{\Omega})}{\|\boldsymbol{\Omega}-\boldsymbol{\Omega}^0\|} \geq \frac{12^{-1}B_1+16^{-1}B_3}{\sqrt{B_1+B_2}} \geq \frac{B_3}{16\sqrt{2B_2}} \geq \frac{1}{16\sqrt{2}}$ because $B_3^2 \geq B_2$. If $B_1 \geq B_2$, then

$$\frac{K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega})}{\|\boldsymbol{\Omega} - \boldsymbol{\Omega}^0\|} \geq \frac{12^{-1}B_1 + 16^{-1}B_3}{\sqrt{B_1+B_2}} \geq \frac{B_1}{12\sqrt{B_1+B_2}} \geq \frac{B_1+B_2}{24\sqrt{B_1+B_2}} \geq \frac{\sqrt{B_1+B_2}}{24} = \frac{\|\boldsymbol{\Omega} - \boldsymbol{\Omega}^0\|}{24}\,.$$

Similarly,

$$\frac{K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega})}{\|\boldsymbol{\Omega} - \boldsymbol{\Omega}^0\|} \geq \sqrt{K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega})}\frac{\sqrt{12^{-1}B_1 + 16^{-1}B_3}}{\sqrt{B_1+B_2}} \geq \sqrt{K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega})}\frac{\sqrt{24^{-1}(B_1+B_2)}}{\sqrt{B_1+B_2}} = \frac{\sqrt{K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega})}}{2\sqrt{6}}\,.$$

This leads to (B.12) and (B.11).

Lemma 5 *(Rate of convergence of constrained MLE) Let $\tilde{A} \supseteq A^0$ be an index set. For $\widehat{\boldsymbol{\Omega}}_{\tilde{A}}$, we have that*

$$\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\| \leq 12\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2\,. \tag{B.13}$$

on the event that $\{\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2 < \frac{1}{8\sqrt{2}}\}$. Moreover, if $\frac{|\tilde{A}|\log p}{n} \to 0$, then

$$\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\| = O_p\left(\sqrt{\frac{|\tilde{A}|\log p}{n}}\right)\,. \tag{B.14}$$

Proof of Lemma 5: By definition of the CMLE, $L_n(\widehat{\boldsymbol{\Omega}}_{\tilde{A}}) - L_n(\boldsymbol{\Omega}^0) \geq 0$, or $-\log\det\widehat{\boldsymbol{\Omega}}_{\tilde{A}} + \log\det\boldsymbol{\Omega}^0 \leq -\operatorname{tr}((\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0)\boldsymbol{S})$. By the Cauchy-Schwarz inequality, this inequality becomes

$$\begin{aligned} 2K(\boldsymbol{\Omega}^0, \widehat{\boldsymbol{\Omega}}_{\tilde{A}}) &\leq \operatorname{tr}((\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0)(\boldsymbol{\Sigma}^0 - \boldsymbol{S})) \leq \|\sqrt{\boldsymbol{\Sigma}^0}(\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0)\sqrt{\boldsymbol{\Sigma}^0}\|_F\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2 \\ &= \|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\|\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2 \end{aligned} \tag{B.15}$$

On the other hand, by (B.12) $\frac{K(\boldsymbol{\Omega}^0, \widehat{\boldsymbol{\Omega}}_{\tilde{A}})}{\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\|} \geq \min\left(\frac{1}{16\sqrt{2}}, \frac{\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\|}{24}\right)$, which, together with (B.15), implies that $\min\left(\frac{1}{8\sqrt{2}}, \frac{\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\|}{12}\right) \leq \|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2$. If $\frac{\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\|}{12} \leq \frac{1}{8\sqrt{2}}$, then it follows immediately that $\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\| \leq 12\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2$. If $\frac{\|\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0\|}{12} > \frac{1}{8\sqrt{2}}$, then $\frac{1}{8\sqrt{2}} \leq \|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2$, which does not happen on the event $\{\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2 < \frac{1}{8\sqrt{2}}\}$.

Moreover, by property of exponential family [2], $\operatorname{Var}(\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})) = 4n^{-1}\boldsymbol{I}_{\tilde{A},\tilde{A}}$. Thus, $\operatorname{Var}(\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})) = 4n^{-1}\boldsymbol{I}_{|\tilde{A}|\times|\tilde{A}|}$. This, combined with Lemma 3, implies that

$$\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2 \leq \sqrt{|\tilde{A}|}\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_\infty = O_p\left(\sqrt{\frac{|\tilde{A}|\log p}{n}}\right) \tag{B.16}$$

on the event that $\{\|\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1/2}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Sigma}^0 - \boldsymbol{S})\|_2 < \frac{1}{8\sqrt{2}}\}$. This event, on the other hand, happens with probability tending to 1 by the assumption that $\frac{|\tilde{A}|\log p}{n} \to 0$. This completes the proof.

Lemma 6 *(Selection consistency) If* $K = |A^0|$, $\tau \le \frac{\bar{\lambda}_{min}\min(\sqrt{C_{min}}, C_{min}^2)}{12|A^0|}$, *then*

$$\begin{aligned}
& \max\left(P\left(\widehat{\boldsymbol{\Omega}}^{(0)} \ne \widehat{\boldsymbol{\Omega}}_{A^0}\right), P\left(\widehat{\boldsymbol{\Omega}}^{(1)} \ne \widehat{\boldsymbol{\Omega}}_{A^0\cup B}\right)\right) \\
\le\ & 2\exp\left(\frac{-nC_{min}}{2560} + 2\log p\right) + \exp\left(\frac{-n}{2560\times 512} + |A^0|\log p\right) \\
& +2\exp\left(-n\frac{\min\left(\sqrt{\frac{\min(C_{min}/512, 3/32)}{48\lambda^2_{max}(|A^0|+|B|)}}, \lambda_{max}(\boldsymbol{\Sigma}^0)\right)^2}{18\lambda^2_{max}(\boldsymbol{\Sigma}^0)} + 2\log p\right) \longrightarrow 0 \qquad \text{(B.17)}
\end{aligned}$$

as $n \to \infty$ *under Assumptions 1-2, where* $\widehat{\boldsymbol{\Omega}}^{(0)}$, $\widehat{\boldsymbol{\Omega}}^{(1)}$, *and* C_{min} *are as defined in (1)–(3).*

Proof of Lemma 6: Let $\hat{A} = \{(i,j) : |\widehat{\omega}^{(1)}_{ij}| \ge \tau, (i,j) \notin B\}$. By definition, $|\hat{A}| \le |A^0|$, $\hat{A} \cap B = \emptyset$ and $\sum_{(i,j)\notin \hat{A}\cup B} |\widehat{\omega}^{(1)}_{ij}| \le \tau(|A^0| - |\hat{A}|)$. Hence, if $\hat{A} = A^0$, then $\widehat{\boldsymbol{\Omega}}^{(1)} = \widehat{\boldsymbol{\Omega}}_{A^0\cup B}$. Suppose $\hat{A} \ne A^0$. On event $\{\hat{A} = A\}$; with fixed $A \ne A^0$, $|A| \le |A^0|$, and $A \cap B = \emptyset$, we bound the Fisher-norm between $\widehat{\boldsymbol{\Omega}}^{(1)}_{A\cup B}$ and an approximating point of $\boldsymbol{\Omega}^0$, $\bar{\boldsymbol{\Omega}}^0_{A\cup B} = \operatorname{argmin}_{\boldsymbol{\Omega}:\boldsymbol{\Omega}_{(A\cup B)^c}=\mathbf{0}} K(\boldsymbol{\Omega}^0, \boldsymbol{\Omega})$. Let $\bar{\boldsymbol{\Sigma}}^0_{A\cup B} = (\bar{\boldsymbol{\Omega}}^0_{A\cup B})^{-1}$. By the Karush-Kuhn-Tucker conditions, $\mathrm{vec}_{A\cup B}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B}) = \mathrm{vec}_{A\cup B}(\boldsymbol{\Sigma}^0)$. Moreover, let $\bar{\lambda}_{\max} = \max_{A:|A|\le K, A\cap B=\emptyset} \lambda_{\max}(\bar{\boldsymbol{\Omega}}^0_{A\cup B})$ and $\bar{\lambda}_{\min} = \min_{A:|A|\le K, A\cap B=\emptyset} \lambda_{\min}(\bar{\boldsymbol{\Omega}}^0_{A\cup B})$. We also define

$$\mathcal{G} = \left\{\|\boldsymbol{S} - \boldsymbol{\Sigma}^0\|_\infty \le \min\left(\frac{1}{16\sqrt{2}\bar{\lambda}_{\max}\sqrt{|A^0|+|B|}}, \sqrt{\frac{\tilde{C}_{\min}}{48\bar{\lambda}^2_{\max}|A^0\cup B|}}, \lambda_{\max}(\boldsymbol{\Sigma}^0)\right)\right\},$$

where

$$\tilde{C}_{\min} = \min_{A:A\ne A^0, |A|=|A^0|, A\cap B=\emptyset} \min\left(\frac{\max(K(\boldsymbol{\Omega}^0, \bar{\boldsymbol{\Omega}}^0_{A\cup B}), K^2(\boldsymbol{\Omega}^0, \bar{\boldsymbol{\Omega}}^0_{A\cup B}))}{|A^0\setminus A|}, 1\right). \qquad \text{(B.18)}$$

By definition of the CMLE, $L_n(\widehat{\boldsymbol{\Omega}}^{(1)}) - L_n(\bar{\boldsymbol{\Omega}}^0_{A\cup B}) \ge 0$, or $-\log\det\widehat{\boldsymbol{\Omega}}^{(1)} + \log\det\bar{\boldsymbol{\Omega}}^0_{A\cup B} \le -\mathrm{tr}((\widehat{\boldsymbol{\Omega}}^{(1)} - \bar{\boldsymbol{\Omega}}^0_{A\cup B})\boldsymbol{S})$. Now let $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Omega}}^{(1)}_{A\cup B} - \bar{\boldsymbol{\Omega}}^0_{A\cup B}$ and $\boldsymbol{\Phi} = \widehat{\boldsymbol{\Omega}}^{(1)} - \widehat{\boldsymbol{\Omega}}^{(1)}_{A\cup B}$, where $\|\boldsymbol{\Phi}\|_1 = \sum_{(i,j)\notin \hat{A}\cup B} |\widehat{\omega}^{(1)}_{ij}| \le (|A^0| - |A|)\tau$. By the Cauchy-Schwarz inequality, the forgoing inequality becomes

$$
\begin{aligned}
& -\log\det(\boldsymbol{I}_{p\times p} + \sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}(\widehat{\boldsymbol{\Delta}} + \boldsymbol{\Phi})\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}) + \mathrm{tr}(\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}(\widehat{\boldsymbol{\Delta}} + \boldsymbol{\Phi})\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}) \\
\leq\ & \mathrm{tr}((\widehat{\boldsymbol{\Delta}} + \boldsymbol{\Phi})(\bar{\boldsymbol{\Sigma}}^0_{A\cup B} - \boldsymbol{S})) = \mathrm{vec}_A(\widehat{\boldsymbol{\Delta}})^\top \mathrm{vec}_A(\bar{\boldsymbol{\Sigma}}^0_{A\cup B} - \boldsymbol{S}) + \mathrm{tr}(\boldsymbol{\Phi}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B} - \boldsymbol{S})) \\
=\ & (\bar{\boldsymbol{I}}^{1/2}_{A\cup B, A\cup B}\,\mathrm{vec}_{A\cup B}(\widehat{\boldsymbol{\Delta}}))^\top \bar{\boldsymbol{I}}^{-1/2}_{A\cup B, A\cup B}\,\mathrm{vec}_{A\cup B}(\boldsymbol{\Sigma}^0_{A\cup B} - \boldsymbol{S}) + \mathrm{tr}(\boldsymbol{\Phi}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B} - \boldsymbol{S})) \\
\leq\ & \left\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\widehat{\boldsymbol{\Delta}}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\right\|_F \left\|\bar{\boldsymbol{I}}^{-1/2}_{A\cup B, A\cup B}\,\mathrm{vec}_{A\cup B}(\boldsymbol{\Sigma}^0_{A\cup B} - \boldsymbol{S})\right\|_2 + \tau(|A^0| - |A|)\|\bar{\boldsymbol{\Sigma}}^0_{A\cup B} - \boldsymbol{S}\|_\infty \\
\leq\ & \left\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\widehat{\boldsymbol{\Delta}}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\right\|_F \lambda_{\max}(\bar{\boldsymbol{\Omega}}^0_{A\cup B})\sqrt{|A\cup B|}\|\boldsymbol{\Sigma}^0 - \boldsymbol{S}\|_\infty \\
& \quad + (2\lambda_{\max}(\boldsymbol{\Sigma}^0) + \lambda_{\max}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B}))\tau K \\
\leq\ & \bar{\lambda}_{\max}\sqrt{|A^0\cup B|}\left\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\widehat{\boldsymbol{\Delta}}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\right\|_F \|\boldsymbol{\Sigma}^0 - \boldsymbol{S}\|_\infty + 3\bar{\lambda}^{-1}_{\min}\tau K
\end{aligned} \tag{B.19}
$$

on the event $\mathcal{G}$, where $\bar{\boldsymbol{I}}_{A\cup B, A\cup B} = \left[\bar{\boldsymbol{\Sigma}}^0_{A\cup B, A\cup B} \otimes_s \bar{\boldsymbol{\Sigma}}^0_{A\cup B, A\cup B}\right]_{A\cup B, A\cup B}$. On the other hand, by Lemma 4,

$$
\begin{aligned}
& -\log\det(\boldsymbol{I}_{p\times p} + \sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}(\widehat{\boldsymbol{\Delta}} + \boldsymbol{\Phi})\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}) + \mathrm{tr}(\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}(\widehat{\boldsymbol{\Delta}} + \boldsymbol{\Phi})\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}) \\
\geq\ & \min\left(\frac{\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}(\widehat{\boldsymbol{\Delta}} + \boldsymbol{\Phi})\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\|_F}{8\sqrt{2}}, \frac{\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}(\widehat{\boldsymbol{\Delta}} + \boldsymbol{\Phi})\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\|_F^2}{12}\right) \\
\geq\ & \min\left(\frac{\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\widehat{\boldsymbol{\Delta}}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\|_F}{8\sqrt{2}}, \frac{\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\widehat{\boldsymbol{\Delta}}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\|_F^2}{24}\right) \\
& - \max\left(\frac{(|A^0| - |A|)\lambda_{\max}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B})\tau}{8\sqrt{2}}, \frac{(|A^0| - |A|)^2\lambda^2_{\max}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B})\tau^2}{12}\right) \\
\geq\ & \min\left(\frac{\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\widehat{\boldsymbol{\Delta}}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\|_F}{8\sqrt{2}}, \frac{\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\widehat{\boldsymbol{\Delta}}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\|_F^2}{24}\right) - \frac{\lambda_{\max}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B})K\tau}{8}
\end{aligned}
$$

where the last two inequalities use that $\|\boldsymbol{M}_1 + \boldsymbol{M}_2\|_F^2 \geq 2^{-1}\|\boldsymbol{M}_1\|_F^2 - \|\boldsymbol{M}_2\|_F^2$, $\|\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\boldsymbol{\Phi}\sqrt{\bar{\boldsymbol{\Sigma}}^0_{A\cup B}}\|_F^2 \leq \lambda^2_{\max}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B})\|\boldsymbol{\Phi}\|_F^2 \leq \lambda^2_{\max}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B})\|\boldsymbol{\Phi}\|_1^2 \leq \lambda^2_{\max}(\bar{\boldsymbol{\Sigma}}^0_{A\cup B})(|A^0| - |A|)^2\tau^2$, and $\min(a-b, c-d) \geq$

$\min(a,c)-\max(b,d)$. Combining this with (B.19), we obtain

$$\begin{aligned}
&\bar{\lambda}_{\max}\sqrt{|A^0\cup B|}\left\|\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\widehat{\mathbf{\Delta}}\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\right\|_F\|\mathbf{\Sigma}^0-\boldsymbol{S}\|_\infty+4\bar{\lambda}^{-1}_{\min}\tau K\\
\geq\;& \min\left(\frac{\|\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\widehat{\mathbf{\Delta}}\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\|_F}{8\sqrt{2}},\frac{\|\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\widehat{\mathbf{\Delta}}\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\|_F^2}{24}\right),
\end{aligned}$$

which implies that

$$\left\|\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\widehat{\mathbf{\Delta}}\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\right\|_F\leq 24\bar{\lambda}_{\max}\sqrt{|A^0\cup B|}\|\boldsymbol{S}-\mathbf{\Sigma}^0\|_\infty+4\sqrt{6\bar{\lambda}^{-1}_{\min}\tau K},$$

on the event $\{\hat{A}=A\}\cap\mathcal{G}$. Next, note that

$$\begin{aligned}
&\frac{2}{n}\left(L_n(\widehat{\mathbf{\Omega}}^{(1)})-L_n(\mathbf{\Omega}^0)\right)+2\left(L(\mathbf{\Omega}^0)-L(\bar{\mathbf{\Omega}}^0_{A\cup B})\right)\\
=\;&\frac{2}{n}\left(L_n(\widehat{\mathbf{\Omega}}^{(1)})-L_n(\bar{\mathbf{\Omega}}^0_{A\cup B})\right)+\mathrm{tr}\left((\mathbf{\Omega}^0-\bar{\mathbf{\Omega}}^0_{A\cup B})(\boldsymbol{S}-\mathbf{\Sigma}^0)\right)\\
=\;&2\left(L(\widehat{\mathbf{\Omega}}^{(1)})-L(\bar{\mathbf{\Omega}}^0_{A\cup B})\right)+\mathrm{tr}((\boldsymbol{S}-\bar{\mathbf{\Sigma}}^0_{A\cup B})(\widehat{\mathbf{\Omega}}^{(1)}-\bar{\mathbf{\Omega}}^0_{A\cup B}))+\mathrm{tr}\left((\mathbf{\Omega}^0-\bar{\mathbf{\Omega}}^0_{A\cup B})(\boldsymbol{S}-\mathbf{\Sigma}^0)\right)\\
\leq\;&\mathrm{tr}((\boldsymbol{S}-\bar{\mathbf{\Sigma}}^0_{A\cup B})(\widehat{\mathbf{\Omega}}^{(1)}-\widehat{\mathbf{\Omega}}^{(1)}_{A\cup B}))+\mathrm{tr}((\boldsymbol{S}-\mathbf{\Sigma}^0)(\widehat{\mathbf{\Omega}}^{(1)}_{A\cup B}-\bar{\mathbf{\Omega}}^0_{A\cup B}))\\
&+\mathrm{tr}\left((\mathbf{\Omega}^0-\bar{\mathbf{\Omega}}^0_{A\cup B})(\boldsymbol{S}-\mathbf{\Sigma}^0)\right) \qquad \text{(B.20)}
\end{aligned}$$

For the first two terms, using $\tau\leq\frac{\bar{\lambda}_{\min}\min(\sqrt{\tilde{C}_{\min}},\tilde{C}^2_{\min})}{12|A^0|}$ and $\|\boldsymbol{S}-\mathbf{\Sigma}^0\|_\infty\leq\sqrt{\frac{\tilde{C}_{\min}}{48\bar{\lambda}^2_{\max}(|A^0|+|B|)}}$, we have that on the event $\mathcal{G}$

$$\begin{aligned}
&\mathrm{tr}((\boldsymbol{S}-\mathbf{\Sigma}^0)(\widehat{\mathbf{\Omega}}^{(1)}_{A\cup B}-\bar{\mathbf{\Omega}}^0_{A\cup B}))+\mathrm{tr}((\boldsymbol{S}-\bar{\mathbf{\Sigma}}^0_{A\cup B})(\widehat{\mathbf{\Omega}}^{(1)}-\widehat{\mathbf{\Omega}}^{(1)}_{A\cup B})\\
\leq\;&\left\|\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\widehat{\mathbf{\Delta}}\sqrt{\bar{\mathbf{\Sigma}}^0_{A\cup B}}\right\|_F\left\|\bar{\boldsymbol{I}}^{-1/2}_{A\cup B,A\cup B}\,\mathrm{vec}_{A\cup B}(\boldsymbol{S}-\mathbf{\Sigma}^0)\right\|_2+\tau K\|\boldsymbol{S}-\bar{\mathbf{\Sigma}}^0_{A\cup B}\|_\infty\\
\leq\;&24\min\left(\bar{\lambda}^2_{\max}|A^0\cup B|\|\boldsymbol{S}-\mathbf{\Sigma}^0\|^2_\infty,\frac{\bar{\lambda}_{\max}\sqrt{|A^0\cup B|}\|\boldsymbol{S}-\mathbf{\Sigma}^0\|_\infty}{16\sqrt{2}}\right)\\
&+\frac{\sqrt{3\bar{\lambda}^{-1}_{\min}\tau K}}{4}+3\bar{\lambda}^{-1}_{\min}\tau K\\
\leq\;&2^{-1}K(\mathbf{\Omega}^0,\bar{\mathbf{\Omega}}^0_{A\cup B})+2^{-1}K(\mathbf{\Omega}^0,\bar{\mathbf{\Omega}}^0_{A\cup B}))=L(\mathbf{\Omega}^0)-L(\bar{\mathbf{\Omega}}^0_{A\cup B}),
\end{aligned}$$

which, together with (B.20), implies that for any $A \neq A^0, |A| \leq K, A \cap B = \emptyset$, we have that

$$\left\{ L_n(\widehat{\mathbf{\Omega}}^{(1)}) - L_n(\mathbf{\Omega}^0) \geq 0; \hat{A} = A; \mathcal{G} \right\} \subseteq \left\{ \operatorname{tr}\left((\mathbf{\Omega}^0 - \bar{\mathbf{\Omega}}^0_{A \cup B})(\boldsymbol{S} - \mathbf{\Sigma}^0) \right) \geq L(\mathbf{\Omega}^0) - L(\bar{\mathbf{\Omega}}^0_{A \cup B}) \right\}$$

Hence,

$$\begin{aligned}
\mathbb{P}\left(\widehat{\mathbf{\Omega}}^{(1)} \neq \widehat{\mathbf{\Omega}}_{A^0 \cup B} \right) &\leq \sum_{A: A \neq A^0, |A| \leq K, A \cap B = \emptyset} \mathbb{P}\left(L_n(\widehat{\mathbf{\Omega}}^{(1)}) - L_n(\mathbf{\Omega}^0) \geq 0; \hat{A} = A; \mathcal{G} \right) + \mathbb{P}(\mathcal{G}^c) \\
\leq & \sum_{A: A \neq A^0, |A| \leq K, A \cap B = \emptyset} \mathbb{P}\left(\operatorname{tr}\left((\mathbf{\Omega}^0 - \bar{\mathbf{\Omega}}^0_{A \cup B})(\boldsymbol{S} - \mathbf{\Sigma}^0) \right) \geq L(\mathbf{\Omega}^0) - L(\bar{\mathbf{\Omega}}^0_{A \cup B}) \right) + \mathbb{P}(\mathcal{G}^c),
\end{aligned}$$

where the first probability can be further bounded by applying Lemmas 3 and 4.

$$\begin{aligned}
& \sum_{A: A \neq A^0, |A| \leq K, A \cap B = \emptyset} \mathbb{P}\left(\operatorname{tr}\left((\mathbf{\Omega}^0 - \bar{\mathbf{\Omega}}^0_{A \cup B})(\boldsymbol{S} - \mathbf{\Sigma}^0) \right) \geq L(\mathbf{\Omega}^0) - L(\bar{\mathbf{\Omega}}^0_{A \cup B}) \right) \\
\leq & \sum_{A: A \neq A^0, |A| \leq K, A \cap B = \emptyset} \exp\left(\frac{-n 10^{-1} K^2(\mathbf{\Omega}^0, \bar{\mathbf{\Omega}}^0_{A \cup B})}{\|\bar{\mathbf{\Omega}}^0_{A \cup B} - \mathbf{\Omega}^0\|^2 + K(\mathbf{\Omega}^0, \bar{\mathbf{\Omega}}^0_{A \cup B}) \|\bar{\mathbf{\Omega}}^0_{A \cup B} - \mathbf{\Omega}^0\|} \right) \\
\leq & \sum_{A: A \neq A^0, |A| \leq K, A \cap B = \emptyset} \exp\left(\frac{-n \min\left(128^{-1}, K(\mathbf{\Omega}^0, \bar{\mathbf{\Omega}}^0_{A \cup B}) \right)}{20} \right) \\
\leq & \sum_{A: A \neq A^0, |A| \leq K, A \cap B = \emptyset, K(\mathbf{\Omega}^0, \bar{\mathbf{\Omega}}^0_{A \cup B}) \leq 1} \exp\left(\frac{-n K(\mathbf{\Omega}^0, \bar{\mathbf{\Omega}}^0_{A \cup B})}{2560} \right) \\
& + \sum_{A: A \neq A^0, |A| \leq K, A \cap B = \emptyset, K(\mathbf{\Omega}^0, \bar{\mathbf{\Omega}}^0_{A \cup B}) > 1} \exp\left(\frac{-n}{2560} \right) \\
\leq & \sum_{j=1}^{|A^0|} \sum_{i=1}^{|A^0|-j} \binom{|A^0|}{j} \binom{p - |A^0|}{i} \exp\left(\frac{-n j \tilde{C}_{\min}}{2560} \right) + \exp\left(\frac{-n}{2560} + |A^0| \log p \right) \\
\leq & \sum_{j=1}^{|A^0|} \exp\left(\frac{-n j \tilde{C}_{\min}}{2560} + 2j \log p \right) + \exp\left(\frac{-n}{2560} + |A^0| \log p \right) \\
\leq & \ 2 \exp\left(\frac{-n \tilde{C}_{\min}}{2560} + 2 \log p \right) + \exp\left(\frac{-n}{2560} + |A^0| \log p \right) \longrightarrow 0
\end{aligned}$$

as $n \to \infty$, provided that $\frac{|A^0| \log p}{n} \leq 3000^{-1}$ and $\tilde{C}_{\min} \geq 3000 \frac{\log p}{n}$.

To bound $\mathbb{P}(\mathcal{G}^c)$, we apply Lemma 3 with $\nu = \min\left(\frac{1}{16\sqrt{2}\bar{\lambda}_{\max}\sqrt{|A^0|+|B|}}, \sqrt{\frac{\tilde{C}_{\min}}{48\bar{\lambda}^2_{\max}|A^0\cup B|}}, \lambda_{\max}(\boldsymbol{\Sigma}^0)\right)$ and get

$$\begin{aligned}\mathbb{P}(\mathcal{G}^c) &\leq \mathbb{P}\left(\|\boldsymbol{S}-\boldsymbol{\Sigma}^0\|_\infty \geq \nu\right) \leq 2\exp\left(-n\frac{\nu^2}{9\lambda^2_{\max}(\boldsymbol{\Sigma}^0)+8\nu\lambda_{\max}(\boldsymbol{\Sigma}^0)} + 2\log p\right) \\ &\leq 2\exp\left(-n\frac{\nu^2}{18\lambda^2_{\max}(\boldsymbol{\Sigma}^0)} + 2\log p\right) \longrightarrow 0\,,\end{aligned}$$

provided that $\tilde{C}_{\min} \geq 2000\frac{\bar{\lambda}^2_{\max}}{\lambda^2_{\min}(\boldsymbol{\Omega}^0)}\frac{(|A^0|+|B|)\log p}{n}$ and $\frac{\bar{\lambda}^2_{\max}}{\lambda^2_{\min}(\boldsymbol{\Omega}^0)}\frac{(|A^0|+|B|)\log p}{n} \leq 18000$. Combining, we obtain

$$\begin{aligned}P\left(\widehat{\boldsymbol{\Omega}}^{(1)} \neq \widehat{\boldsymbol{\Omega}}_{A^0\cup B}\right) &\leq \exp\left(\frac{-n\tilde{C}_{\min}}{2560} + 2\log p\right) + \exp\left(\frac{-n}{2560} + |A^0|\log p\right) \\ &+ \exp\left(-n\frac{\min\left(\sqrt{\frac{\min\left(\tilde{C}_{\min}, 3/32\right)}{48\bar{\lambda}^2_{\max}(|A^0|+|B|)}}, \lambda_{\max}(\boldsymbol{\Sigma}^0)\right)^2}{18\lambda^2_{\max}(\boldsymbol{\Sigma}^0)} + 2\log p\right)\end{aligned}$$

For $\mathbb{P}\left(\widehat{\boldsymbol{\Omega}}^{(0)} \neq \widehat{\boldsymbol{\Omega}}_{A^0}\right)$, we let $B = \emptyset$ and a similar bound can be established. Moreover, by Lemma 4, it is easy to see that $\max(K(\boldsymbol{\Omega}^0,\boldsymbol{\Omega}), K^2(\boldsymbol{\Omega}^0,\boldsymbol{\Omega})) \geq \frac{\|\boldsymbol{\Omega}^0-\boldsymbol{\Omega}\|^2}{512}$ for any $\boldsymbol{\Omega}$. Consequently, $\tilde{C}_{\min} \geq \frac{C_{\min}}{512}$. Thus, the bound in (B.17) is established. This completes the proof.

Lemma 7 *Let* $\boldsymbol{\Gamma}_k = (\gamma_{k1}, \cdots \gamma_{km}) \in \mathbb{R}^m$*;* $k = 1, \cdots, n$ *be iid random vectors with* $\mathrm{Var}(\boldsymbol{\gamma}_1) = \boldsymbol{I}_{m\times m}$*. If* m *is fixed, then*

$$n^{-1}\|\sum_{k=1}^n \boldsymbol{\gamma}_k\|_2^2 \xrightarrow{d} \chi^2_m, \text{ as } n \to \infty\,. \tag{B.21}$$

Otherwise, if $\max\left(m, m_2m/n, m_3/n, m_3m^{3/2}/n^2\right) \to 0$*, where* $m_j = \max_{1\leq i\leq m}\mathbb{E}\gamma_{1i}^{2j}$*;* $j = 2, 3$*, then*

$$\frac{\|\sum_{k=1}^n \boldsymbol{\gamma}_k\|_2^2 - nm}{n\sqrt{2m}} \xrightarrow{d} N(0,1), \text{ as } n \to \infty\,. \tag{B.22}$$

Proof of Lemma 7: If m is fixed, then (B.21) follows from the central limit theorem and the continuous mapping theorem.

For (B.22), let $\boldsymbol{\Gamma}_k = \sum_{j=1}^{k} \boldsymbol{\gamma}_j$; $k = 1, \cdots, n$ be a partial sum of k iid m-dimensional vectors $\boldsymbol{\gamma}_j$'s. Next we apply Theorem 18.1 of [1] to show that $\frac{\|\boldsymbol{\Gamma}_n\|_2^2 - nm}{n\sqrt{2m}} \to N(0,1)$ for triangular arrays of martingale differences $\{\eta_{n,k} = \frac{\|\boldsymbol{\Gamma}_k\|_2^2 - \|\boldsymbol{\Gamma}_{k-1}\|_2^2 - m}{n\sqrt{2m}} = \frac{\|\boldsymbol{\gamma}_k\|_2^2 - m + 2\boldsymbol{\gamma}_k^\top \boldsymbol{\Gamma}_{k-1}}{n\sqrt{2m}}\}$. Towards this end, we verify that

$$\sum_{k=1}^{n} \mathbb{E}\big(\eta_{n,k}^2 \,|\, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{k-1}\big) \xrightarrow{P} 1, \quad \sum_{k=1}^{n} \mathbb{E}|\eta_{n,k}|^3 \to 0. \tag{B.23}$$

For the first condition of (B.23), we compute $\mathbb{E}$ and Var of $\mathbb{E}\big(\eta_{n,k}^2 \,|\, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{k-1}\big)$. Note that $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_m$ are iid vectors with $\mathrm{Var}(\boldsymbol{\gamma}_m) = \boldsymbol{I}_{m\times m}$, $\mathbb{E}\boldsymbol{\Gamma}_{k-1} = 0$, and $\mathbb{E}\|\boldsymbol{\Gamma}_{k-1}\|_2^2 = (k-1)m$. Then, for each $k = 1, \cdots, n$, $\mathbb{E}\mathbb{E}\big(\eta_{n,k}^2 \,|\, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{k-1}\big)$ becomes

$$\begin{aligned}
&(2mn^2)^{-1}\Big(\mathbb{E}\big(\|\boldsymbol{\gamma}_k\|_2^2 - m\big)^2 + 4\mathbb{E}\big((\|\boldsymbol{\gamma}_k\|_2^2 - m)\boldsymbol{\gamma}_k\big)^\top \mathbb{E}\boldsymbol{\Gamma}_{k-1} + 4\mathbb{E}\mathbb{E}\big((\boldsymbol{\gamma}_k^\top \boldsymbol{\Gamma}_{k-1})^2 \,|\, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{k-1}\big)\Big) \\
&= (2mn^2)^{-1}\Big(\mathrm{Var}(\|\boldsymbol{\gamma}_k\|_2^2) + 4\mathbb{E}\|\boldsymbol{\Gamma}_{k-1}\|_2^2\Big) = (2mn^2)^{-1}\Big(\mathrm{Var}(\|\boldsymbol{\gamma}_k\|_2^2) + 4(k-1)m\Big),
\end{aligned}$$

which, after summing over $k = 1, \cdots, n$, leads to

$$\sum_{k=1}^{n} \frac{2(k-1)}{n^2} \le \mathbb{E}\Big(\sum_{k=1}^{n} \mathbb{E}\big(\eta_{n,k}^2 \,|\, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{k-1}\big)\Big) \le \frac{mm_2}{2n} + \sum_{k=1}^{n} \frac{2(k-1)}{n^2},$$

where $\mathrm{Var}(\|\boldsymbol{\gamma}_k\|_2) \le m^2 m_2$; $k = 1, \cdots, n$. Consequently, $\Big|\mathbb{E}\Big(\sum_{k=1}^{n} \mathbb{E}\big(\eta_{n,k}^2 \,|\, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{k-1}\big)\Big) - 1\Big| \le \frac{2}{n} + \frac{mm_2}{2n}$. Let $\boldsymbol{a} = \mathbb{E}\big((\|\boldsymbol{\gamma}_1\|_2^2 - m)\boldsymbol{\gamma}_1\big)$. Similarly, using an inequality $(a_1 + a_2 + a_3)^2 \le$

$3(a_1^2+a_2^2+a_3^2)$ for real numbers a_j; $j=1,\cdots,3$.

$$\begin{aligned}
&\mathrm{Var}\Big(\sum_{k=1}^n \mathbb{E}\big(\eta_{n,k}^2 \,|\, \boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_{k-1}\big)\Big) = \frac{4}{m^2n^4}\,\mathrm{Var}\left(\sum_{k=1}^n \big(\boldsymbol{a}^\top \boldsymbol{\Gamma}_{k-1} + \|\boldsymbol{\Gamma}_{k-1}\|_2^2\big)\right)\\
&= \frac{4}{m^2n^4}\,\mathrm{Var}\left(\sum_{k=1}^n (n-k)\Big(\boldsymbol{a}^\top\boldsymbol{\gamma}_k + \|\boldsymbol{\gamma}_k\|_2^2\Big) + 2\sum_{k<k'}(n-(k\vee k'))\boldsymbol{\gamma}_k^\top\boldsymbol{\gamma}_{k'}\right)\\
&\le \frac{12}{m^2n^4}\left[\mathrm{Var}\left(\sum_{k=1}^n (n-k)\boldsymbol{a}^\top\boldsymbol{\gamma}_k\right) + \mathrm{Var}\left(\sum_{k=1}^n (n-k)\|\boldsymbol{\gamma}_k\|_2^2\right)\right.\\
&\quad \left. + \mathrm{Var}\left(\sum_{k<k'}(n-(k\vee k'))\boldsymbol{\gamma}_k^\top\boldsymbol{\gamma}_{k'}\right)\right] \equiv \frac{12}{m^2n^4}\Big[T_1+T_2+T_3\Big]. \qquad \text{(B.24)}
\end{aligned}$$

For T_1, note that $\|\boldsymbol{a}\|_2^2 \le \sum_{k=1}^m \mathbb{E}^2\big((\|\boldsymbol{\gamma}_1\|_2^2 - m)\gamma_{1k}\big) \le \sum_{k=1}^m \mathbb{E}\big((\|\boldsymbol{\gamma}_1\|_2^2-m)^2\big)\mathbb{E}\gamma_{1k}^2 \le m^3m_2$. Then

$$\begin{aligned}
\mathrm{Var}\left(\sum_{k=1}^n (n-k)\boldsymbol{a}^\top\boldsymbol{\gamma}_k\right) &= \sum_{k=1}^n (n-k)^2\mathbb{E}\left(\boldsymbol{a}^\top\boldsymbol{\gamma}_k\right)^2 = \sum_{k=1}^n (n-k)^2\sum_{j=1}^m a_j^2\mathbb{E}\gamma_{kj}^2\\
&= \frac{\|\boldsymbol{a}\|_2^2}{6}(n-1)n(2n-1) \le n^3m^3m_2\,.
\end{aligned}$$

For T_2, note that $\mathrm{Var}\,(\sum_{k=1}^n (n-k)\|\boldsymbol{\gamma}_k\|_2^2) \le \sum_{k=1}^n (n-k)^2m^2m_2 = \frac{1}{6}(n-1)n(2n-1)m^2m_2$. To bound T_3, note that, for $k\neq k'$ and $j\neq j'$, $\mathbb{E}\big(\boldsymbol{\gamma}_k^\top\boldsymbol{\gamma}_{k'}\boldsymbol{\gamma}_j^\top\boldsymbol{\gamma}_{j'}\big) = \mathbb{I}(\{j,j'\}=\{k,k'\})\mathbb{E}\left(\boldsymbol{\gamma}_k^\top\boldsymbol{\gamma}_{k'}\right)^2 = \mathbb{I}(\{j,j'\}=\{k,k'\})m$, yielding that

$$\mathrm{Var}\left(\sum_{k<k'}(n-(k\vee k'))\boldsymbol{\gamma}_k^\top\boldsymbol{\gamma}_{k'}\right) = \sum_{k<k'}(n-(k\vee k'))^2\mathbb{E}\left(\boldsymbol{\gamma}_k^\top\boldsymbol{\gamma}_{k'}\right)^2 \le n^4m\,.$$

Combining (B.24) with the bounds of $T_1 - T_3$, we obtain

$$\mathrm{Var}\Big(\sum_{k=1}^n \mathbb{E}\big(\eta_{n,k}^2 \,|\, \boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_{k-1}\big)\Big) \le \frac{12\,(n^3m^3m_2 + n^3m^2m_2 + n^4m)}{m^2n^4}.$$

Hence the first condition of (B.23) is implied by the assumption that $mm_2/n \to 0$ and

$m \to \infty$.

For the second condition of (B.23), note that $\mathbb{E}|\eta_{n,k}|^3 = \mathbb{E}\left(\left|\|\boldsymbol{\gamma}_k\|_2^2 - m + 2\boldsymbol{\gamma}_k^\top \boldsymbol{\Gamma}_{k-1}\right|^3\right)$ is bounded by

$$
\begin{aligned}
& 4\mathbb{E}\left(\left|\|\boldsymbol{\gamma}_k\|_2^2 - m\right|^3\right) + 16\mathbb{E}\left(\left|\boldsymbol{\gamma}_k^\top \boldsymbol{\Gamma}_{k-1}\right|^3\right) \le \mathbb{E}\left(\|\boldsymbol{\gamma}_k\|_2^6\right) + \sqrt{\mathbb{E}\big((\boldsymbol{\gamma}_k^\top \boldsymbol{\Gamma}_{k-1})^6\big)} \\
\le \;& m^3 m_3 + \sqrt{(k-1)^3 m^3 m_3 + (k-1)^2 m^3 m_2 m_3 + (k-1) m^3 m_3^2} \\
\le \;& m^3 m_3 + k^{3/2} m^{3/2} m_3^{1/2} + k m^{3/2} m_2^{1/2} m_3^{1/2} + k^{1/2} m^{3/2} m_3.
\end{aligned}
$$

Summing over k, $\frac{\sum_{k=1}^n \mathbb{E}\left(\left|\|\boldsymbol{\gamma}_k\|_2^2 - m + 2\boldsymbol{\gamma}_k^\top \boldsymbol{\Gamma}_{k-1}\right|^3\right)}{n^3 m^{3/2}}$ is upper bounded by

$$
\begin{aligned}
& \frac{\left(n m^3 m_3 + n^{5/2} m^{3/2} m_3^{1/2} + n^2 m^{3/2} m_2^{1/2} m_3^{1/2} + n^{3/2} m^{3/2} m_3\right)}{n^3 m^{3/2}} \\
= \;& \frac{m^{3/2} m_3}{n^2} + \frac{m_3^{1/2}}{n^{1/2}} + \frac{m_2^{1/2} m_3^{1/2}}{n} + \frac{m_3}{n^{3/2}} \to 0\,,
\end{aligned}
$$

provided that $\max\left(m_2 m/n, m_3/n, m_3 m^{3/2}/n^2\right) \to 0$. Thus the second condition in (B.23) is met. As a consequence of Theorem 18.1 of [1], the desired asymptotic normality is established. This completes the proof.

Lemma 8 *Let $\boldsymbol{X} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^0)$ and $\gamma = \operatorname{tr}(\boldsymbol{X}\boldsymbol{X}^\top - \boldsymbol{\Sigma}^0)\boldsymbol{T})$ with $\boldsymbol{T}$ a symmetric matrix. Then*

$$
\mathbb{E}(\gamma^{2m}) \le (2m-1)!\, 2^{m-1} \left(\mathbb{E}(\gamma^2)\right)^m \text{ for any integer } m \ge 1\,. \tag{B.25}
$$

Proof of Lemma 8: As in (B.8) and (B.10), we expand the moment generating function of γ: $M_\gamma(\lambda) = \mathbb{E}\exp(\lambda\gamma) = \lambda^2 \|\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{T}\sqrt{\boldsymbol{\Sigma}^0}\|_F^2 + (1/2)\sum_{l=3}^\infty l^{-1}\lambda^l \operatorname{tr}\left[(2\boldsymbol{T}\boldsymbol{\Sigma}^0)^l\right]$ for any $|\lambda| < \|\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{T}\sqrt{\boldsymbol{\Sigma}^0}\|_F/2$. Direct computation of high-order derivatives of $M_\gamma(\lambda)$ in λ yields that $\mathbb{E}(\gamma^{2m}) = (2m-1)!\, 2^{2m-1} \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{\Sigma}^0)^{2m}\right)$ for any integer $m \ge 1$. An application of $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{\Sigma}^0)^{2m}\right) \le \|\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{T}\sqrt{\boldsymbol{\Sigma}^0}\|_F^{2m}$ yields that $\mathbb{E}(\gamma^{2m}) \le (2m-1)!\, 2^{2m-1}\|\sqrt{\boldsymbol{\Sigma}^0}\boldsymbol{T}\sqrt{\boldsymbol{\Sigma}^0}\|_F^{2m} = (2m-1)!\, 2^{m-1}\left(\mathbb{E}(\gamma^2)\right)^m$. This completes the proof.

Proof of Lemma 9: Let $\widehat{\mathbf{\Delta}}_{\tilde{A}} = \widehat{\mathbf{\Omega}}_{\tilde{A}} - \mathbf{\Omega}^0$ for any $\tilde{A} \supseteq A^0$. Applying Lemma 5 to $\widehat{\mathbf{\Delta}}_{\tilde{A}}$ and $\widehat{\mathbf{\Delta}}_{A^0}$, we have that both $\|\widehat{\mathbf{\Delta}}_{\tilde{A}}\|$ and $\|\widehat{\mathbf{\Delta}}_{A^0}\|$ tend to zero in probability as n goes to infinity. Hence, we could assume throughout the proof that $\max\left(\|\widehat{\mathbf{\Delta}}_{\tilde{A}}\|, \|\widehat{\mathbf{\Delta}}_{A^0}\|\right) \leq 1/2$ holds with probability tending to one. Note that $\mathbf{\Omega}^0 = (\mathbf{\Sigma}^0)^{-1}$, and $\log\det(\widehat{\mathbf{\Omega}}_{\tilde{A}}) = \log\det(\boldsymbol{I}_{p\times p} + \widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0) + \log\det(\mathbf{\Omega}^0)$. Then

$$\begin{aligned}
& \log\det(\boldsymbol{I}_{p\times p} + \widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0) \\
= \ & \log\det(\boldsymbol{I}_{p\times p} + [\mathbf{\Sigma}^0]^{1/2}\widehat{\mathbf{\Delta}}_{\tilde{A}}[\mathbf{\Sigma}^0]^{1/2}) = \operatorname{tr}(\log(\boldsymbol{I}_{p\times p} + [\mathbf{\Sigma}^0]^{1/2}\widehat{\mathbf{\Delta}}_{\tilde{A}}[\mathbf{\Sigma}^0]^{1/2})) \\
= \ & \operatorname{tr}\left(\sum_{i=1}^{\infty}(-1)^{i+1}\frac{\left([\mathbf{\Sigma}^0]^{1/2}\widehat{\mathbf{\Delta}}_{\tilde{A}}[\mathbf{\Sigma}^0]^{1/2}\right)^i}{i}\right), \\
= \ & \operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right) - \frac{1}{2}\operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right) + R_1(\widehat{\mathbf{\Delta}}_{\tilde{A}}),
\end{aligned} \tag{B.26}$$

where $R_1(\widehat{\mathbf{\Delta}}_{\tilde{A}}) = \sum_{i=3}^{\infty}\frac{(-1)^{i+1}}{i}\operatorname{tr}\left(\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right)^i\right)$ and the expansion is valid since $\|\widehat{\mathbf{\Delta}}_{\tilde{A}}\| \leq 1/2 < 1$. As a result,

$$\begin{aligned}
& n^{-1}\left(L_n(\widehat{\mathbf{\Omega}}_{\tilde{A}}) - L_n(\mathbf{\Omega}^0)\right) \\
= \ & \frac{1}{2}\operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right) - \frac{1}{4}\operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right) - \frac{1}{2}\operatorname{tr}(\widehat{\mathbf{\Delta}}_{\tilde{A}}\boldsymbol{S}) + \frac{1}{2}R_1(\widehat{\mathbf{\Delta}}_{\tilde{A}}) \\
= \ & \frac{1}{2}\operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}(\mathbf{\Sigma}^0 - \boldsymbol{S})\right) - \frac{1}{4}\|\widehat{\mathbf{\Delta}}_{\tilde{A}}\|^2 + \frac{1}{2}R_1(\widehat{\mathbf{\Delta}}_{\tilde{A}}).
\end{aligned} \tag{B.27}$$

Moreover, using the property of the CMLE, $\widehat{\mathbf{\Delta}}_{\tilde{A}}$ satisfies a score equation: $[-(\widehat{\mathbf{\Delta}}_{\tilde{A}} + \mathbf{\Omega}^0)^{-1} + \boldsymbol{S}]_{\tilde{A}} = 0$. This, in turn, yields that

$$\left[\mathbf{\Sigma}^0\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right]_{\tilde{A}} = \left[R_2(\widehat{\mathbf{\Delta}}_{\tilde{A}}) + \mathbf{\Sigma}^0 - \boldsymbol{S}\right]_{\tilde{A}}, \tag{B.28}$$

where $(\widehat{\mathbf{\Delta}}_{\tilde{A}} + \mathbf{\Omega}^0)^{-1} = \mathbf{\Sigma}^0 - \mathbf{\Sigma}^0\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0 + R_2(\widehat{\mathbf{\Delta}}_{\tilde{A}})$ is used, and $R_2(\widehat{\mathbf{\Delta}}_{\tilde{A}}) = \mathbf{\Sigma}^0\sum_{i=2}^{\infty}(-1)^i\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right)^i$. By the definition of $\otimes$ and (B.2), (B.28) can be rewritten in a vector form as

$$2\boldsymbol{I}_{\tilde{A},\tilde{A}}\operatorname{vec}_{\tilde{A}}(\widehat{\mathbf{\Delta}}_{\tilde{A}}) = \operatorname{vec}\left(R_2(\widehat{\mathbf{\Delta}}_{\tilde{A}}) + \mathbf{\Sigma}^0 - \boldsymbol{S}\right). \tag{B.29}$$

Moreover, after taking the inner product with $\widehat{\mathbf{\Delta}}_{\tilde{A}}$ for both sides of (B.28), we obtain

$$\operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\widehat{\mathbf{\Delta}}_{\tilde{A}}\mathbf{\Sigma}^0\right) = \operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}R_2(\widehat{\mathbf{\Delta}}_{\tilde{A}})\right) + \operatorname{tr}\left(\widehat{\mathbf{\Delta}}_{\tilde{A}}(\mathbf{\Lambda})\right), \tag{B.30}$$

where $\mathbf{\Lambda} = \mathbf{\Sigma}^0 - \boldsymbol{S}$. Hence, combining (B.29) and (B.30) with (B.27) yields that

$$2n^{-1}\left(L_n(\widehat{\boldsymbol{\Omega}}_{\tilde{A}}) - L_n(\boldsymbol{\Omega}^0)\right) = \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\boldsymbol{\Lambda}\right) - \frac{1}{2}\operatorname{tr}\left(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) + R_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})$$
$$= \frac{1}{2}\left(\operatorname{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Delta}})\right)^\top \operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Lambda} - R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) + R_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})$$
$$= \frac{1}{4}\operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)^\top \boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}} \operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Lambda} - R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) + R_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})$$
$$= \frac{1}{4}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Lambda})^\top \boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}} \operatorname{vec}_{\tilde{A}}(\boldsymbol{\Lambda}) - \frac{1}{4}\operatorname{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)^\top \boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}} \operatorname{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) + R_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\,.$$

Similarly,

$$2n^{-1}\left(L_n(\widehat{\boldsymbol{\Omega}}_{A^0}) - L_n(\boldsymbol{\Omega}^0)\right)$$
$$= \frac{1}{4}\operatorname{vec}_{A^0}(\boldsymbol{\Lambda})^\top \boldsymbol{I}^{-1}_{A^0,A^0} \operatorname{vec}_{A^0}(\boldsymbol{\Lambda}) - \frac{1}{4}\operatorname{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right) \boldsymbol{I}^{-1}_{A^0,A^0} \operatorname{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right) + R_1(\widehat{\boldsymbol{\Delta}}_{A^0}).$$

Combining, we obtain that

$$2\left(L_n(\widehat{\boldsymbol{\Omega}}_{\tilde{A}}) - L_n(\widehat{\boldsymbol{\Omega}}_{A^0})\right) = \frac{n}{4}\operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Lambda}\right)^\top \boldsymbol{I}^{-1}_{B,B}\operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Lambda}\right) - \frac{n}{4}\operatorname{vec}_{A^0}\left(\boldsymbol{\Lambda}\right)^\top \boldsymbol{I}^{-1}_{A^0,A^0}\operatorname{vec}_{A^0}\left(\boldsymbol{\Lambda}\right) + R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0}) \quad \text{(B.31)}$$

where

$$R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0}) = nR_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - \frac{n}{4}\operatorname{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)^\top \boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}}\operatorname{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) - nR_1(\widehat{\boldsymbol{\Delta}}_{A^0}) + \frac{n}{4}\operatorname{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right)^\top \boldsymbol{I}^{-1}_{A^0,A^0}\operatorname{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right) \quad \text{(B.32)}$$

is the remainder to be bounded subsequently. For now, we focus on the leading term in the likelihood ratio expansion. Let $\boldsymbol{\lambda} = \sqrt{n}\operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Sigma}^0 - \boldsymbol{S}\right)$. Now write $\boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}}$ as

$$\boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}} = \begin{pmatrix} \boldsymbol{J}_{A^0,A^0} & \boldsymbol{J}_{A^0,B} \\ \boldsymbol{J}_{B,A^0} & \boldsymbol{J}_{B,B} \end{pmatrix}. \quad \text{(B.33)}$$

Note that $\boldsymbol{I}_{A^0,A^0} = [\boldsymbol{J}^{-1}]_{A^0,A^0} = \left(\boldsymbol{J}_{A^0,A^0} - \boldsymbol{J}_{A^0,B}\boldsymbol{J}^{-1}_{B,B}\boldsymbol{J}_{B,A^0}\right)^{-1}$. Thus,

$$\frac{n}{4}\operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Lambda}\right)^\top \boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}}\operatorname{vec}_{\tilde{A}}\left(\boldsymbol{\Lambda}\right) - \frac{n}{4}\operatorname{vec}_{A^0}\left(\boldsymbol{\Lambda}\right)^\top \boldsymbol{I}^{-1}_{A^0,A^0}\operatorname{vec}_{A^0}\left(\boldsymbol{\Lambda}\right)$$
$$= \frac{1}{4}\boldsymbol{\lambda}^\top_{\tilde{A}}\boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}}\boldsymbol{\lambda}_{\tilde{A}} - \frac{1}{4}\boldsymbol{\lambda}^\top_{A^0}\boldsymbol{I}^{-1}_{A^0,A^0}\boldsymbol{\lambda}_{A^0}$$
$$= \frac{1}{4}\boldsymbol{\lambda}^\top_{\tilde{A}}\boldsymbol{J}\boldsymbol{\lambda}_{\tilde{A}} - \frac{1}{4}\boldsymbol{\lambda}^\top_{A^0}\left(\boldsymbol{J}_{A^0,A^0} - \boldsymbol{J}_{A^0,B}\boldsymbol{J}^{-1}_{B,B}\boldsymbol{J}_{B,A^0}\right)\boldsymbol{\lambda}_{A^0}$$
$$= \frac{1}{4}\left(\boldsymbol{J}_{B,A^0}\boldsymbol{\lambda}_{A^0} + \boldsymbol{J}_{\tilde{A}\setminus A^0,B\setminus A^0}\boldsymbol{\lambda}_B\right)^\top \boldsymbol{J}^{-1}_{A\setminus A^0,B}\left(\boldsymbol{J}_{\tilde{A}\setminus A^0,A^0}\boldsymbol{\lambda}_{A^0} + \boldsymbol{J}_{B\setminus A^0,B}\boldsymbol{\lambda}_B\right)$$
$$= \frac{1}{4}\boldsymbol{\lambda}^\top_{\tilde{A}}\boldsymbol{J}_{\tilde{A},B}\boldsymbol{J}^{-1}_{B,B}\boldsymbol{J}_{\tilde{A}\setminus A^0,A}\boldsymbol{\lambda}_{\tilde{A}} = \left\|\frac{1}{2}\boldsymbol{J}^{-1/2}_{B,B}\boldsymbol{J}_{B,\tilde{A}}\sqrt{n}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Lambda})\right\|_2^2. \quad \text{(B.34)}$$

This, together with (B.31), implies that

$$2\left(L_n(\widehat{\boldsymbol{\Omega}}_{\tilde{A}}) - L_n(\widehat{\boldsymbol{\Omega}}_{A^0})\right) = \left\|\frac{1}{2}\boldsymbol{J}^{-1/2}_{B,B}\boldsymbol{J}_{B,\tilde{A}}\sqrt{n}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Lambda})\right\|_2^2 + R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0}), \quad \text{(B.35)}$$

Recall from (B.47) that $\mathrm{Var}\left(\frac{1}{2}\boldsymbol{J}_{B,B}^{-1/2}\boldsymbol{J}_{B,\tilde{A}}\sqrt{n}\,\mathrm{vec}_A(\boldsymbol{\Lambda})\right) = \boldsymbol{I}_{|B|\times|B|}$, thus by Lemma 7 and Lemma 8, if $|B|$ is a fixed constant, $2\Big(L_n(\widehat{\boldsymbol{\Omega}}_{\tilde{A}}) - L_n(\widehat{\boldsymbol{\Omega}}_{A^0})\Big) \xrightarrow{P_0} W_{|\tilde{A}\setminus A^0|}$ provided that $R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0}) = o_p(1)$; if $|\tilde{A}\setminus A^0| \to \infty$, $(2|\tilde{A}\setminus A^0|)^{-1/2}\Big(2\big(L_n(\widehat{\boldsymbol{\Omega}}_{\tilde{A}}) - L_n(\widehat{\boldsymbol{\Omega}}_{A^0})\big) - |\tilde{A}\setminus A^0|\Big) \xrightarrow{P_0} N(0,1)$ provided that $R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0})/\sqrt{|B|} = o_p(1)$. Next it remains to prove that the remainder term $R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0})$ satisfies the aforementioned conditions. Toward this end, we bound $R_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_1(\widehat{\boldsymbol{\Delta}}_{A^0})$ and $\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1}\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) - \mathrm{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right)\boldsymbol{I}_{A^0,A^0}^{-1}\mathrm{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right)$ respectively.

For $\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1}\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)$, recursively applying $\|C_1C_2\|_F \le \|C_1\|_F\|C_2\|_F$ and using the fact that $\|C_1C_2\|_F \le \lambda_{\max}(C_2)\|C_1\|_F$ and $\|C_1C_2\|_F \le \lambda_{\max}(C_1)\|C_2\|_F$, we obtain

$$\begin{aligned}\left\|\mathrm{vec}_{\tilde{A}}\left(\boldsymbol{\Sigma}^0\left(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\boldsymbol{\Sigma}^0\right)^i\right)\right\|_2 &\le \left\|\sqrt{\boldsymbol{\Sigma}^0}\left(\sqrt{\boldsymbol{\Sigma}^0}\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\sqrt{\boldsymbol{\Sigma}^0}\right)^i\sqrt{\boldsymbol{\Sigma}^0}\right\|_F \\ &\le \lambda_{\max}(\boldsymbol{\Sigma}^0)\left\|\sqrt{\boldsymbol{\Sigma}^0}\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\sqrt{\boldsymbol{\Sigma}^0}\right\|_F^i = \lambda_{\max}(\boldsymbol{\Sigma}^0)\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^i \qquad \text{(B.36)}\end{aligned}$$

Summing over i yields that

$$\begin{aligned}\left\|\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)\right\|_2 &\le \sum_{i=2}^{\infty}\left\|\mathrm{vec}_{\tilde{A}}\left(\boldsymbol{\Sigma}^0\left(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\boldsymbol{\Sigma}^0\right)^i\right)\right\|_2 \\ &\le \lambda_{\max}(\boldsymbol{\Sigma}^0)\sum_{i=2}^{\infty}\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^i \le 2\lambda_{\max}(\boldsymbol{\Sigma}^0)\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^2. \qquad \text{(B.37)}\end{aligned}$$

Consequently,

$$\begin{aligned}&\mathrm{vec}_B\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)\boldsymbol{I}_{B,B}^{-1}\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) \le \left\|\boldsymbol{I}_{B,B}^{-1}\right\|_{\mathrm{opt}}\left\|\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)\right\|_2^2 \\ \le\ & \lambda_{\min}^{-2}(\boldsymbol{\Sigma}^0)\left\|\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)\right\|_2^2 \le 4\kappa_0^2\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^4\,. \qquad \text{(B.38)}\end{aligned}$$

Similarly, $\mathrm{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right)\boldsymbol{I}_{A^0,A^0}^{-1}\mathrm{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right) \le 4\kappa_0^2\|\widehat{\boldsymbol{\Delta}}_{A^0}\|^4$. Hence,

$$\begin{aligned}&\frac{1}{4}\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right)\boldsymbol{I}_{\tilde{A},\tilde{A}}^{-1}\mathrm{vec}_{\tilde{A}}\left(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})\right) - \frac{1}{4}\mathrm{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right)\boldsymbol{I}_{A^0,A^0}^{-1}\mathrm{vec}_{A^0}\left(R_2(\widehat{\boldsymbol{\Delta}}_{A^0})\right) \\ \le\ & \kappa_0^2\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^4 + \kappa_0^2\|\widehat{\boldsymbol{\Delta}}_{A^0}\|^4 \qquad \text{(B.39)}\end{aligned}$$

For $R_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_1(\widehat{\boldsymbol{\Delta}}_{A^0})$, by Cauchy-Schwartz inequality, we have that $\mathrm{tr}(\big(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\boldsymbol{\Sigma}^0\big)^i) \le \|\sqrt{\boldsymbol{\Sigma}^0}\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\sqrt{\boldsymbol{\Sigma}^0}\|_F\left\|(\sqrt{\boldsymbol{\Sigma}^0}\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\sqrt{\boldsymbol{\Sigma}^0})^{i-1}\right\|_F \le \|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^i$; $i = 2, \cdots ,$. Hence,

$$\Big|\sum_{i=4}^{\infty}\frac{(-1)^{i+1}}{i}\mathrm{tr}(\big(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\boldsymbol{\Sigma}^0\big)^i)\Big| \le \sum_{i=4}^{\infty}i^{-1}\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^i \le \frac{\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^4}{4(1-\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|)} \le \frac{1}{2}\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^4\,. \qquad \text{(B.40)}$$

Similarly, $\left|\sum_{i=4}^{\infty}\frac{(-1)^{i+1}}{i}\mathrm{tr}(\big(\widehat{\boldsymbol{\Delta}}_{A^0}\boldsymbol{\Sigma}^0\big)^i)\right| \le \frac{1}{2}\|\widehat{\boldsymbol{\Delta}}_{A^0}\|^4$. Combining, we have that

$$\left|R_1(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_1(\widehat{\boldsymbol{\Delta}}_{A^0})\right| \leq \frac{\left|\mathrm{tr}\left((\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\boldsymbol{\Sigma}^0)^3\right) - \mathrm{tr}\left((\widehat{\boldsymbol{\Delta}}_{A^0}\boldsymbol{\Sigma}^0)^3\right)\right|}{3} + \frac{\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^4 + \|\widehat{\boldsymbol{\Delta}}_{A^0}\|^4}{2} \quad \text{(B.41)}$$

Let $f_{\tilde{A}}(\mathrm{vec}_{\tilde{A}}(\boldsymbol{\Delta})) = \mathrm{tr}\left((\boldsymbol{\Delta}\boldsymbol{\Sigma}^0)^3\right)$ with $\mathrm{vec}_{A^c}(\boldsymbol{\Delta}) = \mathbf{0}$. A Taylor expansion of $f_{\tilde{A}}(\mathrm{vec}_{\tilde{A}}(\boldsymbol{\Delta}))$ at $\mathrm{vec}_{A^0}(\boldsymbol{\Delta}))$ yields that

$$\begin{aligned}
\frac{1}{3}\left|\mathrm{tr}\left((\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\boldsymbol{\Sigma}^0)^3\right) - \mathrm{tr}\left((\widehat{\boldsymbol{\Delta}}_{A^0}\boldsymbol{\Sigma}^0)^3\right)\right| &= \frac{1}{3}\left(\mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - \mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Delta}}_{A^0})\right)^\top \nabla f(\mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Delta}}^*)) \\
&= \left(\mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0})\right)^\top \mathrm{vec}_{\tilde{A}}\left(\boldsymbol{\Sigma}^0(\widehat{\boldsymbol{\Delta}}^*\boldsymbol{\Sigma}^0)^2\right) = \mathrm{tr}\left(\boldsymbol{\Sigma}^0(\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0})(\boldsymbol{\Sigma}^0\widehat{\boldsymbol{\Delta}}^*)^2\right) \\
&\leq 2\left\|\sqrt{\boldsymbol{\Sigma}^0}(\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0})\sqrt{\boldsymbol{\Sigma}^0}\right\|_F \max\left(\|\sqrt{\boldsymbol{\Sigma}^0}\widehat{\boldsymbol{\Delta}}_{A^0}\sqrt{\boldsymbol{\Sigma}^0}\|_F^2, \|\sqrt{\boldsymbol{\Sigma}^0}\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\sqrt{\boldsymbol{\Sigma}^0}\|_F^2\right) \quad \text{(B.42)}
\end{aligned}$$

where $\widehat{\boldsymbol{\Delta}}^*$ is some convex combination of $\widehat{\boldsymbol{\Delta}}_{\tilde{A}}$ and $\widehat{\boldsymbol{\Delta}}_{A^0}$ and the last equality uses (B.36). Lastly, we bound $\left\|\sqrt{\boldsymbol{\Sigma}^0}(\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0})\sqrt{\boldsymbol{\Sigma}^0}\right\|_F = \left\|\boldsymbol{I}^{1/2}_{\tilde{A},\tilde{A}}\mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0})\right\|_2$. By (B.29), we have that

$$\begin{aligned}
\boldsymbol{I}^{1/2}_{\tilde{A},\tilde{A}}\mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0}) &= \boldsymbol{I}^{1/2}_{\tilde{A},\tilde{A}}\left(\mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Omega}}_{\tilde{A}} - \boldsymbol{\Omega}^0) - \mathrm{vec}_{\tilde{A}}(\widehat{\boldsymbol{\Omega}}_{A^0} - \boldsymbol{\Omega}^0)\right) \\
&= \frac{1}{2}\boldsymbol{I}^{-1/2}_{\tilde{A},\tilde{A}}\mathrm{vec}_{\tilde{A}}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})) - \frac{1}{2}\boldsymbol{I}^{1/2}_{\tilde{A},\tilde{A}}\begin{bmatrix}\boldsymbol{I}^{-1}_{A^0,A^0}\mathrm{vec}_{A^0}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{A^0})) \\ \mathbf{0}\end{bmatrix} \\
&= \frac{1}{2}\boldsymbol{I}^{-1/2}_{\tilde{A},\tilde{A}}\left(\mathrm{vec}_{\tilde{A}}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})) - \begin{bmatrix}\mathrm{vec}_{A^0}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{A^0})) \\ \boldsymbol{I}_{B,A^0}\boldsymbol{I}^{-1}_{A^0,A^0}\mathrm{vec}_{A^0}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{A^0}))\end{bmatrix}\right) \\
&= \frac{1}{2}\boldsymbol{I}^{-1/2}_{\tilde{A},\tilde{A}}\begin{bmatrix}\mathrm{vec}_{A^0}(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_2(\widehat{\boldsymbol{\Delta}}_{A^0})) \\ \mathrm{vec}_B(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})) - \boldsymbol{I}_{B,A^0}\boldsymbol{I}^{-1}_{A^0,A^0}\mathrm{vec}_{A^0}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{A^0}))\end{bmatrix}, \quad \text{(B.43)}
\end{aligned}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^0 - \boldsymbol{S}$. Let $\boldsymbol{J} = \boldsymbol{I}^{-1}_{\tilde{A},\tilde{A}}$. An application of an inequality $\|\boldsymbol{I}^{-1/2}_{\tilde{A},\tilde{A}}\boldsymbol{x}\|_2^2 = \boldsymbol{x}^\top\boldsymbol{J}\boldsymbol{x} \leq 2\boldsymbol{x}^\top_{A^0}\boldsymbol{J}_{A^0,A^0}\boldsymbol{x}_{A^0} + 2\boldsymbol{x}^\top_B\boldsymbol{J}_{B,B}\boldsymbol{x}_B$ yields that

$$\begin{aligned}
&\left\|\boldsymbol{I}^{-1/2}_{\tilde{A},\tilde{A}}\begin{bmatrix}\mathrm{vec}_{A^0}(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_2(\widehat{\boldsymbol{\Delta}}_{A^0})) \\ \mathrm{vec}_B(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})) - \boldsymbol{I}_{B,A^0}\boldsymbol{I}^{-1}_{A^0,A^0}\mathrm{vec}_{A^0}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{A^0}))\end{bmatrix}\right\|_F^2 \\
\leq\ & 2\left\|\boldsymbol{J}^{1/2}_{B,B}\left(\mathrm{vec}_{B\setminus A^0}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}})) - \boldsymbol{I}_{B,A^0}\boldsymbol{I}^{-1}_{A^0,A^0}\mathrm{vec}_{A^0}(\boldsymbol{\Lambda} + R_2(\widehat{\boldsymbol{\Delta}}_{A^0}))\right)\right\|_2^2 \\
& +2\left\|\boldsymbol{J}^{1/2}_{A^0,A^0}\mathrm{vec}_{A^0}(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_2(\widehat{\boldsymbol{\Delta}}_{A^0}))\right\|_2^2. \quad \text{(B.44)}
\end{aligned}$$

Moreover, $\boldsymbol{J}^{-1}_{B,B}\boldsymbol{J}_{B,A^0} + \boldsymbol{I}_{B,A^0}\boldsymbol{I}^{-1}_{A^0,A^0} = \mathbf{0}$. Using this, we have that

$$\begin{aligned}
&\left\| \boldsymbol{J}_{B,B}^{1/2} \left(\mathrm{vec}_B(\boldsymbol{\Lambda}) - \boldsymbol{I}_{B,A^0} \boldsymbol{I}_{A^0,A^0}^{-1} \mathrm{vec}_{A^0}(\boldsymbol{\Lambda}) \right) \right\|_2^2 \\
= &\left\| \boldsymbol{J}_{B,B}^{-1/2} \left(\boldsymbol{J}_{B,B} \mathrm{vec}_B(\boldsymbol{\Lambda}) + \boldsymbol{J}_{B,A^0} \mathrm{vec}_{A^0}(\boldsymbol{\Lambda}) \right) \right\|_2^2 = \left\| \boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,\tilde{A}} \mathrm{vec}_{\tilde{A}}(\boldsymbol{\Lambda}) \right\|_2^2 . \quad \text{(B.45)}
\end{aligned}$$

This, together with (B.43) and (B.44), implies that

$$\begin{aligned}
&\left\| \sqrt{\boldsymbol{\Sigma}^0} (\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0}) \sqrt{\boldsymbol{\Sigma}^0} \right\|_F^2 \\
\leq &\frac{1}{2} \left\| \boldsymbol{J}_{A^0,A^0}^{1/2} \mathrm{vec}_{A^0}(R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_2(\widehat{\boldsymbol{\Delta}}_{A^0})) \right\|_2^2 + \frac{1}{2} \left\| \boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,B} \mathrm{vec}_B(\boldsymbol{\Lambda}) \right\|_2^2 . \quad \text{(B.46)}
\end{aligned}$$

By (B.3), the covariance matrix of $\boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,B} \mathrm{vec}_{\tilde{A}}(\boldsymbol{\Lambda})$ is

$$\begin{aligned}
&\mathrm{Var} \left(\boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,B} \mathrm{vec}_B(\boldsymbol{\Lambda}) \right) = n^{-1} \boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,B} \mathrm{Var} \left(\sqrt{n} \mathrm{vec}_{\tilde{A}}(\boldsymbol{\Lambda}) \right) \boldsymbol{J}_{\tilde{A},B} \boldsymbol{J}_{B,B}^{-1/2} \\
= &n^{-1} \boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,\tilde{A}} \left(4\boldsymbol{J}^{-1} \right) \boldsymbol{J}_{\tilde{A},B} \boldsymbol{J}_{B,B}^{-1/2} = 4n^{-1} \boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,B} \boldsymbol{J}_{B,B}^{-1/2} = 4n^{-1} \boldsymbol{I}_{|B|\times|B|} , \quad \text{(B.47)}
\end{aligned}$$

By Lemma 3, $\left\| \boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,A} \mathrm{vec}_{\tilde{A}}(\boldsymbol{\Lambda}) \right\|_2^2 \leq |B| \left\| \boldsymbol{J}_{B,B}^{-1/2} \boldsymbol{J}_{B,\tilde{A}} \mathrm{vec}_A(\boldsymbol{\Lambda}) \right\|_\infty^2 = O_p\left(\frac{|B| \log |B|}{n} \right)$. Using this and (B.37), we bound (B.46) as follows:

$$\begin{aligned}
&\left\| \sqrt{\boldsymbol{\Sigma}^0} (\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0}) \sqrt{\boldsymbol{\Sigma}^0} \right\|_F^2 \leq 2^{-1} \lambda_{\min}^{-2}(\boldsymbol{\Sigma}^0) \left\| R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) - R_2(\widehat{\boldsymbol{\Delta}}_{A^0}) \right\|_F^2 + O_p\left(\frac{|B| \log |B|}{n} \right) \\
\leq &\; 2\lambda_{\min}^{-2}(\boldsymbol{\Sigma}^0) \max \left(\left\| R_2(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}) \right\|_F^2 , \left\| R_2(\widehat{\boldsymbol{\Delta}}_{A^0}) \right\|_F^2 \right) + O_p\left(\frac{|B| \log |B|}{n} \right) \\
\leq &\; 8\kappa_0^2 \max \left(\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^4, \|\widehat{\boldsymbol{\Delta}}_{A^0}\|^4 \right) + O_p\left(\frac{|B| \log |B|}{n} \right).
\end{aligned}$$

Let $\Delta = \max \left(\|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|, \|\widehat{\boldsymbol{\Delta}}_{A^0}\| \right)$. Then combining the above bound with (B.42), we obtain

$$\begin{aligned}
&\frac{1}{3} \left| \mathrm{tr} \left((\widehat{\boldsymbol{\Delta}}_{\tilde{A}} \boldsymbol{\Sigma}^0)^3 \right) - \mathrm{tr} \left((\widehat{\boldsymbol{\Delta}}_{A^0} \boldsymbol{\Sigma}^0)^3 \right) \right| \\
\leq &\; 2 \left\| \sqrt{\boldsymbol{\Sigma}^0} (\widehat{\boldsymbol{\Delta}}_{\tilde{A}} - \widehat{\boldsymbol{\Delta}}_{A^0}) \sqrt{\boldsymbol{\Sigma}^0} \right\|_F \max \left(\|\widehat{\boldsymbol{\Delta}}_{A^0}\|^2, \|\widehat{\boldsymbol{\Delta}}_{\tilde{A}}\|^2 \right) \\
\leq &\; 4\Delta^2 \max \left(3\kappa_0 \Delta^2 , O_p\left(\sqrt{\frac{|B| \log |B|}{n}} \right) \right).
\end{aligned}$$

This together with (B.39) and (B.41) implies that the remainder term $R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0})$ defined in (B.32) is bounded by $n\Delta^2 \max \left(\kappa_0^2 \Delta^2 , O_p\left(\sqrt{\frac{|B| \log |B|}{n}} \right) \right)$ up to some positive constants. By Lemma 5, we have that $\Delta^2 = O_p\left(\frac{|\tilde{A}| \log p}{n} \right)$. This together with (B.39) and (B.41) yields that

$$R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0}) = O_p\left(\max \left(\frac{\kappa_0^2 |\tilde{A}|^2 \log^2(p+1)}{n}, |\tilde{A}| \log(p+1) \sqrt{\frac{|B| \log |B|}{n}} \right) \right)$$

Hence, if $|B|$ is fixed, $R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0}) = o_p(1)$, provided that $\frac{\kappa_0^2 |A|^2 \log^2 p}{n} \to 0$; and if $|\tilde{A} \backslash A^0| \to \infty$,

$R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}}, \widehat{\boldsymbol{\Delta}}_{A^0})/\sqrt{|B|} = o_p(1)$, provided that $\frac{\kappa_0^2 |\tilde{A}|^2 \log^2 p \log(|B|)}{n} \to 0$. This completes the proof.

C Proofs of Theorem 3 and 4

Proof of Theorem 3. Let $\Lambda_n(B)$ be the likelihood ratio test statistic defined in Theorem 1. A measure change from $\mathbb{P}_{\boldsymbol{\theta}^n}$ to $\mathbb{P}_{\boldsymbol{\theta}^0}$ yields that for any $u \geq 0$,

$$
\begin{aligned}
& \mathbb{P}_{\boldsymbol{\theta}^n}(\Lambda_n(B) \geq u) = \mathbb{E}_{\boldsymbol{\theta}^n}\mathbb{I}(\Lambda_n(B) \geq u) \\
= \;& \mathbb{E}_{\boldsymbol{\theta}^0}\left(\mathbb{I}(\Lambda_n(B) > u)\exp(\sqrt{n}\operatorname{vec}_B(\delta_n)^\top Z_n - \frac{n\operatorname{vec}_B(\delta_n)^\top \boldsymbol{I}_{B,B}\operatorname{vec}_B(\delta_n)}{2} + R_n(\boldsymbol{\theta}^0, \delta_n))\right),
\end{aligned}
$$

where $\mathbb{P}_{\boldsymbol{\theta}^n}$ is the probability measure under H_a, $Z_n = n^{-1/2}\frac{\partial L_n(\boldsymbol{\theta}^0)}{\partial \theta_B}$, $\boldsymbol{I}$ is the Fisher information matrix, and $R_n(\boldsymbol{\theta}^0, \delta_n)) = L_n(\boldsymbol{\theta}^n) - L_n(\boldsymbol{\theta}^0) - \sqrt{n}\operatorname{vec}_B(\delta_n)^\top Z_n + \frac{n\operatorname{vec}_B(\delta_n)^\top \boldsymbol{I}_{B,B}\operatorname{vec}_B(\delta_n)}{2}$. We will verify later that

$$
R_n(\boldsymbol{\theta}^0, \delta_n)) \xrightarrow{\mathbb{P}_{\theta^0}} 0 \tag{C.1}
$$

in the Gaussian graphical model and linear regression model.

For the Gaussian graphical model, we first verify (C.1). Now let $\boldsymbol{h}_n = \sqrt{n}\operatorname{vec}_B(\delta_n)$ with $\|\boldsymbol{h}_n\|_2 = h$. Then $Z_n = n^{-1/2}\frac{\partial L_n(\Omega)}{\partial \Omega_B} = \sqrt{n}\operatorname{vec}_B((\boldsymbol{\Omega}^0)^{-1} - \boldsymbol{S}) = \sqrt{n}\operatorname{vec}_B(\boldsymbol{\Lambda})$. It follows from the Taylor expansion of $\log\det(\cdot)$ that

$$
\begin{aligned}
& L_n(\boldsymbol{\theta}^n) - L_n(\boldsymbol{\theta}^0) = n\left(\log\det(\boldsymbol{\Omega}^n) - \operatorname{tr}(\boldsymbol{\Omega}^n\boldsymbol{S}) - \log\det(\boldsymbol{\Omega}^0) + \operatorname{tr}(\boldsymbol{\Omega}^0\boldsymbol{S})\right) \\
= \;& \boldsymbol{h}_n^\top\sqrt{n}\operatorname{vec}_B((\boldsymbol{\Omega}^0)^{-1} - \boldsymbol{S}) - \sqrt{n}\boldsymbol{h}_n^\top\operatorname{vec}_B((\boldsymbol{\Omega}^0)^{-1}) + n(\log\det(\boldsymbol{\Omega}^n) - \log\det(\boldsymbol{\Omega}^0)) \\
= \;& \boldsymbol{h}_n^\top Z_n - \frac{1}{2}\boldsymbol{h}_n^\top\boldsymbol{I}_{B,B}\boldsymbol{h}_n + r(\boldsymbol{\Omega}^n),
\end{aligned}
$$

where we have used (B.26) and

$$r(\boldsymbol{\Omega}^n) = n\sum_{i=3}^{\infty}(-1)^{i+1}\frac{\operatorname{tr}\left[(\sqrt{\Sigma^0}(\boldsymbol{\Omega}^n-\boldsymbol{\Omega}^0)\sqrt{\Sigma^0})^i\right]}{i} \tag{C.2}$$

By similar calculations as in (B.40), we have that

$$|r(\boldsymbol{\Omega}^n)| \leq \begin{cases} \frac{n}{3}\sum_{i=3}^{n}(\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n)^{i/2}\left(\frac{|B|^{1/4}}{\sqrt{n}}\right)^i & \text{if } |B|\to\infty \\ \frac{n}{3}\sum_{i=3}^{n}(\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n)^{i/2}\left(\frac{1}{\sqrt{n}}\right)^i & \text{if } |B| \text{ is fixed.}\end{cases} \tag{C.3}$$

Hence, when $|B|$ is fixed and n is large enough, we have that $|r(\boldsymbol{\Omega}^n)| \leq (\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n)^{3/2}n^{-1/2} \to 0$. When $|B|\to\infty$ but $|B|^{3/2}/n \to 0$, we have that $|r(\boldsymbol{\Omega}^n)| \leq (\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n)^{3/2}\frac{|B|^{3/4}}{n^{1/2}} \to 0$. Therefore,

$$R_n(\boldsymbol{\theta}^0,\delta_n) = L_n(\boldsymbol{\theta}^n) - L_n(\boldsymbol{\theta}^0) - \boldsymbol{h}_n^\top Z_n + \frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n = r(\boldsymbol{\Omega}^n) \to 0\,. \tag{C.4}$$

By (B.35), we have that, with probability tending to 1 under $P_{\boldsymbol{\theta}^0}$,

$$\Lambda_n(B) = \left\|\frac{1}{2}\boldsymbol{J}_{B,B}^{-1/2}\boldsymbol{J}_{B,\tilde{A}}\sqrt{n}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Lambda})\right\|_2^2 + R(\widehat{\boldsymbol{\Delta}}_{\tilde{A}},\widehat{\boldsymbol{\Delta}}_{A^0}). \tag{C.5}$$

Note that $\operatorname{Var}(\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Lambda})) = 4\boldsymbol{I}$. Hence, by Lemmas 7 and 8,

$$\left(\frac{1}{2}\boldsymbol{J}_{B,B}^{-1/2}\boldsymbol{J}_{B,\tilde{A}}\sqrt{n}\operatorname{vec}_{\tilde{A}}(\boldsymbol{\Lambda}), \frac{1}{2}\sqrt{n}\operatorname{vec}_B(\boldsymbol{\Lambda})\right) \xrightarrow{d} (Z_1,Z_2) \sim N\left(\boldsymbol{0}, \begin{pmatrix}\boldsymbol{I}_{|B|\times|B|} & \boldsymbol{J}_{B,B}^{-1/2} \\ \boldsymbol{J}_{B,B}^{-1/2} & \boldsymbol{I}_{B,B}\end{pmatrix}\right), \tag{C.6}$$

where $\boldsymbol{J} = \boldsymbol{I}^{-1}$. Therefore,

$$Z_1 \sim N(0, I_{|B|\times|B|}) \text{ and } Z_2 \mid Z_1 = z_1 \sim N\left(\boldsymbol{J}_{B,B}^{-1/2}z_1\,, \boldsymbol{I}_{B,A^0}\boldsymbol{I}_{A^0,A^0}^{-1}\boldsymbol{I}_{A^0,B}\right) \tag{C.7}$$

where the fact that $\boldsymbol{J}_{B,B} = (\boldsymbol{I}_{B,B} - \boldsymbol{I}_{B,A^0}\boldsymbol{I}^{-1}_{A^0,A^0}\boldsymbol{I}_{A^0,B})^{-1}$ is used. Hence, for any θ_j; $j \in B^c$,

$$
\begin{aligned}
& P_{H_a}(\Lambda_n(B) \geq u) \to \mathbb{E}\left(\mathbb{I}(\|Z_1\|_2^2 \geq u)\exp(\boldsymbol{h}_n^\top Z_2 - \frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n)\right) \\
= & \exp\left(-\frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n\right)\mathbb{E}_{Z_1}\left[\mathbb{I}(\|Z_1\|_2^2 \geq u)\mathbb{E}_{Z_2|Z_1}\left(\exp(\boldsymbol{h}_n^\top Z_2)\right]\right. \\
= & \exp\left(-\frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{J}^{-1}_{B,B}\boldsymbol{h}_n\right)\mathbb{E}_{Z_1}\left[\mathbb{I}(\|Z_1\|_2^2 \geq u)\exp\left(Z_1^\top \boldsymbol{J}^{-1/2}_{B,B}\boldsymbol{h}_n\right)\right] \\
= & \mathbb{E}_{Z_1}\mathbb{I}(\|Z_1 + \boldsymbol{J}^{-1/2}_{B,B}\boldsymbol{h}_n\|_2^2 \geq u) = \mathbb{P}\left(\|Z_1 + \boldsymbol{J}^{-1/2}_{B,B}\boldsymbol{h}_n\|_2^2 \geq u\right)
\end{aligned}
$$

where we have used the fact that $\boldsymbol{J}^{-1}_{B,B} = \boldsymbol{I}_{B,B} - \boldsymbol{I}_{B,A^0}\boldsymbol{I}^{-1}_{A^0,A^0}\boldsymbol{I}_{A^0,B}$. Hence, we must have $\Lambda_n(B) \xrightarrow{d} \|Z_1 + bmJ^{-1/2}_{B,B}\boldsymbol{h}_n\|_2^2$ with $Z_1 \sim N(\mathbf{0}, I_{|B|\times|B|})$ when $|B|$ is fixed. When $|B| \to \infty$, for any vector v with $\|v\|_2 = c|B|^{1/4}$ for some constant c, we have that

$$
\frac{\|Z+v\|_2^2 - |B|}{\sqrt{2|B|}} = \frac{\|Z\|_2^2 - |B|}{\sqrt{2|B|}} + \frac{\|v\|_2}{\sqrt{2}|B|^{1/4}}\left(\frac{2v^\top Z}{\|v\|_2|B|^{1/4}} + \frac{\|v\|_2}{|B|^{1/4}}\right) \xrightarrow{d} N\left(\frac{c^2}{\sqrt{2}}, 1\right), \quad \text{(C.8)}
$$

because the first term converges to $N(0,1)$ by CLT, and the second term converges $c^2/\sqrt{2}$ to since $\frac{2v^\top Z}{\|v\|_2|B|^{1/4}} \to 0$ in probability.

Consequently, the *local limiting power functions* for the proposed CMLR test is

$$
\pi_{LR}(h, \theta_{B^c}) \quad = \quad \begin{cases} \mathbb{P}\left(\|\boldsymbol{Z} + \boldsymbol{J}^{-1/2}_{B,B}\boldsymbol{h}_n\|_2^2 \geq \chi^2_{\alpha,|B|}\right) & \text{when } |B| \text{ is fixed,} \\ \mathbb{P}\left(Z + \frac{\boldsymbol{h}_n^\top \boldsymbol{J}^{-1}_{B,B}\boldsymbol{h}_n}{\sqrt{2|B|}} \geq z_\alpha\right) & \text{when } |B| \to \infty, \end{cases}
$$

where $\alpha > 0$ is the level of significance, $\boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{I}_{|B|\times|B|})$ is a multivariate normal random variable, and $\boldsymbol{J}_{B,B}$ is the asymptotic variance of $\mathrm{vec}_B(\widehat{\boldsymbol{\Omega}}^{(1)})$.

To make a comparison between the debiased lasso test proposed in [3], we consider the case when $|B| = 1$. Assume that $B = \{(i,j)\}$. In this case, the *local limiting power functions* for the proposed method is

$$\pi_{LR}(h,\theta_{B^c}) = \mathbb{P}\left(\left(Z + \frac{|h|}{\sigma_{\mathrm{LR}}}\right)^2 > \chi^2_\alpha\right) = \mathbb{P}\left(\left|Z + \frac{|h|}{\sigma}\right| > z_{\alpha/2}\right) \tag{C.9}$$

where σ^2_{LR} is the asymptotic variance of $\hat{\omega}^{(1)}_{ij}$. In contrast, The *local limiting power functions* for the debiased lasso test proposed in [3] is

$$\pi_{debias}(h,\theta_{B^c}) = \mathbb{P}\left(\left|Z + \frac{|h|}{\sqrt{\omega^2_{ij} + \omega_{ii}\omega_{jj}}}\right| > z_{\alpha/2}\right) \tag{C.10}$$

where $Z \sim N(0,1)$ is a standard normal random variable. By applying Corollary 1, we have that $\sigma^2_{\mathrm{LR}} < \omega^2_{ij} + \omega_{ii}\omega_{jj}$, which implies that our $\pi_{LR}(h,\theta_{B^c}) \geq \pi_{debias}(h,\theta_{B^c})$. This completes the proof.

Proof of Theorem 4. The proof is similar to that of Theorem 3. Again, we first verify that (C.1) is satisfied for linear regression. Toward that end, let $\boldsymbol{h}_n = \sqrt{n}\,\mathrm{vec}_B(\delta_n)$ with $\|\boldsymbol{h}_n\|_2 = h$. Notice that $L_n(\theta) = L_n(\beta,\sigma) = n\log(1/\sqrt{2\pi}\sigma) - (2\sigma^2)^{-1}\|y - X\beta\|_2^2$.

$$Z_n = n^{-1/2}\frac{\partial L_n(\beta^0)}{\partial \beta^0_B} = n^{-1/2}\sigma^{-2}\,\mathrm{vec}_B\left(X^\top(y - X\beta^0)\right) = n^{-1/2}\sigma^{-2}\,\mathrm{vec}_B(X^\top\boldsymbol{\epsilon}), \tag{C.11}$$

where $\boldsymbol{\epsilon} \sim N(0,\sigma^2 I_{n\times n})$. Moreover, we have that

$$\begin{aligned}
&L_n(\boldsymbol{\theta}^n) - L_n(\boldsymbol{\theta}^0) = (2\sigma^2)^{-1}\left(\|y - X\beta^0\|_2^2 - \|y - X(\beta^0+\delta_n)\|_2^2\right)\\
=\;& \sqrt{n}\,\mathrm{vec}_B(\delta_n)^\top n^{-1/2}\sigma^{-2}\,\mathrm{vec}_B(X^\top(y - X\beta^0)) - (2\sigma^2)^{-1}\,\mathrm{vec}_B(\delta_n)^\top(X^\top X)_{B,B}\,\mathrm{vec}_B(\delta_n)\\
=\;& \boldsymbol{h}_n^\top Z_n - \frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{I}_{B,B}\boldsymbol{h}_n
\end{aligned}$$

where $\boldsymbol{I} = (n\sigma^2)^{-1}X^\top X$. Hence (C.1) is satisfied with the remaining term to be exactly 0.

By similar arguments used in Theorem 2 and the fact that $\|\epsilon\|_2^2/n \xrightarrow{\mathbb{P}_{\beta^0}} 0$, we have that the likelihood ratio test statistic is

$$\Lambda_n(B) = \boldsymbol{\epsilon}^\top(\boldsymbol{P}_{A^0 \cup B} - \boldsymbol{P}_{A^0})\boldsymbol{\epsilon} + R(\boldsymbol{\epsilon}) \tag{C.12}$$

where $R(\boldsymbol{\epsilon}) \xrightarrow{\mathbb{P}_{\beta^0}} 0$. Moreover, since the matrix $\boldsymbol{P}_{A^0 \cup B} - \boldsymbol{P}_{A^0}$ is idempotent and has rank $|B|$, there must exist $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{|B|}$ such that $\boldsymbol{P}_{A^0 \cup B} - \boldsymbol{P}_{A^0} = \sum_{k=1}^{|B|} \boldsymbol{a}_k \boldsymbol{a}_k^\top$ and

$$\Lambda_n(B) = \sum_{k=1}^{|B|} (\boldsymbol{a}_k^\top \boldsymbol{\epsilon})^2 + R(\boldsymbol{\epsilon}) \tag{C.13}$$

Note that, under $\mathbb{P}_{\beta^0}$, we have that

$$\big((\boldsymbol{a}_1^\top \boldsymbol{\epsilon}, \ldots \boldsymbol{a}_{|B|}^\top \boldsymbol{\epsilon}), \mathrm{vec}_B(X^\top \boldsymbol{\epsilon})\big) = (Z_1, Z_2) \sim N\left(\mathbf{0}, \begin{pmatrix} I_{|B|\times|B|} & \boldsymbol{A}\boldsymbol{X}_B \\ \boldsymbol{X}_B^\top \boldsymbol{A}^\top & \boldsymbol{X}_B^\top \boldsymbol{X}_B \end{pmatrix}\right) \tag{C.14}$$

where $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{|B|})^\top \in \mathbb{R}^{|B|\times n}$.

Therefore,

$$Z_1 \sim N(0, I_{|B|\times|B|}) \text{ and } Z_2 \mid Z_1 = z_1 \sim N\left(\boldsymbol{X}_B^\top \boldsymbol{A}^\top z_1\,, \boldsymbol{X}_B^\top (I_{n\times n} - \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{X}_B\right) \tag{C.15}$$

Hence, for any β_j; $j \in B^c$ and any $u \geq 0$,

$$
\begin{aligned}
& P_{H_a}(\Lambda_n(B) \geq u) \\
\to \quad & \mathbb{E}\left(\mathbb{I}(\|Z_1\|_2^2 \geq u) \exp(\boldsymbol{h}_n^\top Z_2 - \frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{X}_B^\top \boldsymbol{X}_B \boldsymbol{h}_n\right) \\
= \quad & \exp\left(-\frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{X}_B^\top \boldsymbol{X}_B \boldsymbol{h}_n)\right) \mathbb{E}_{Z_1}\left[\mathbb{I}(\|Z_1\|_2^2 \geq u)\mathbb{E}_{Z_2|Z_1}\left(\exp(\boldsymbol{h}_n^\top Z_2)\right]\right. \\
= \quad & \exp\left(-\frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{X}_B^\top \boldsymbol{X}_B \boldsymbol{h}_n\right) \mathbb{E}_{Z_1}\left[\mathbb{I}(\|Z_1\|_2^2 \geq u) \exp\left(Z_1^\top \boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n + \frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{X}_B^\top (I_{n\times n} - \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{X}_B \boldsymbol{h}_n\right)\right] \\
= \quad & \exp\left(-\frac{1}{2}\boldsymbol{h}_n^\top \boldsymbol{X}_B^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n\right) \mathbb{E}_{Z_1}\left[\mathbb{I}(\|Z_1\|_2^2 \geq u) \exp\left(Z_1^\top \boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n\right)\right] \\
= \quad & \mathbb{E}_{Z_1}\mathbb{I}(\|Z_1 + \boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n\|_2^2 \geq u) = \mathbb{P}\left(\|Z_1 + \boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n\|_2^2 \geq u\right)
\end{aligned}
$$

Hence, we must have $\Lambda_n(B) \xrightarrow{d} \|Z + \boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n\|_2^2$ with $Z \sim N(\mathbf{0}, I_{|B|\times|B|})$ when $|B|$ is fixed. When $|B| \to \infty$, a similar argument used in Theorem 3 can be applied.

Consequently, the *local limiting power functions* for the proposed CMLR test is

$$
\pi_{LR}(h, \beta_{B^c}) = \begin{cases} \mathbb{P}\left(\|Z + \boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n\|_2^2 \geq \chi^2_{\alpha,|B|}\right) & \text{if } |B| \text{ is fixed,} \\ \mathbb{P}\left(Z_1 + \frac{\|\boldsymbol{A}\boldsymbol{X}_B \boldsymbol{h}_n\|_2^2}{\sqrt{2|B|}} \geq z_\alpha\right) & \text{if } |B| \to \infty \quad . \end{cases} \tag{C.16}
$$

where $\alpha > 0$ is the level of significance, $Z \sim N(\mathbf{0}, \boldsymbol{I}_{|B|\times|B|})$ is a multivariate normal random variable, and $Z_1 \sim N(0, 1)$ is a standard normal random variable.

Since $\boldsymbol{A}\boldsymbol{X}_B$ has full rank $|B|$, it is easy to see that when $\|\boldsymbol{h}_n\|_2 \to \infty$ and $|B|$ is finite, then $\pi_{LR}(h, \beta_{B^c}) \to 1$; and when $\|\boldsymbol{h}_n\|_2^2/\sqrt{|B|} \to \infty$ and $|B| \to \infty$, then $\pi_{LR}(h, \beta_{B^c}) \to 1$. This completes the proof.

References

[1] P. Billingsley. *Convergence of probability measures*, volume 493. John Wiley & Sons, 2009.

[2] L. D. Brown. Fundamentals of statistical exponential families with applications in statistical decision theory. *Lecture Notes-monograph series*, pages 1–279, 1986.

[3] J. Janková and S. Van de Geer. Honest confidence regions and optimality in high-dimensional precision matrix estimation. *TEST*, pages 1–20, 2016.