Supplement to "Large Multiple Graphical Model Inference via Bootstrap"

Yongli Zhang Xiaotong Shen Shaoli Wang

April 30, 2019

Proof of Theorem 1

The following notations are used in the proof.

Let $\mathbf{\Omega}$ be the true concentration matrix and suppose $F_\lambda^v(.)$ is a convex penalty on a positive definite and symmetric matrix $\mathbf{G}$. Let

$$\begin{aligned}
\widetilde{\mathbf{\Omega}} &= \text{argmin}_{\mathbf{G}} \big(\text{tr}(\mathbf{\Sigma G}) - \log\det(\mathbf{G}) + F_{\lambda,v}(\mathbf{G}) \big) \\
\widehat{\mathbf{\Omega}} &= \text{argmin}_{\mathbf{G}} \big(\text{tr}(\mathbf{SG}) - \log\det(\mathbf{G}) + F_{\lambda,v}(\mathbf{G}) \big) \\
\widehat{\mathbf{\Omega}}^* &= \text{argmin}_{\mathbf{G}} \big(\text{tr}(\mathbf{S}^*\mathbf{G}) - \log\det(\mathbf{G}) + F_\lambda^v(\mathbf{G}) \big).
\end{aligned}$$

Lemma 1 *Let $\mathbf{Q}$ be a real matrix and $\|.\|$ be the operator norm of a matrix. Let $\mathbf{A}$ and $\mathbf{B}$*

be positive definite and symmetric matrices. Then

$$\left|\operatorname{tr}(\mathbf{QA})\right| \leqslant \|\mathbf{Q}\| \operatorname{tr}(\mathbf{A}); \tag{1}$$

$$\log\det(\mathbf{AB}) \leqslant \operatorname{tr}(\mathbf{AB}) - p; \tag{2}$$

$$\operatorname{tr}(\mathbf{AB}) \operatorname{tr}\left[(\mathbf{B}^{-1} - \mathbf{A}^{-1})(\mathbf{A} - \mathbf{B})\right] \geqslant \|\mathbf{A} - \mathbf{B}\|_F^2. \tag{3}$$

Proof: First, $\mathbf{A} = \sum_{j=1}^{p} \gamma_j \epsilon_j \epsilon_j'$. Then using the Cauchy-Schwarz Inequality,

$$\left|\operatorname{tr}(\mathbf{QA})\right| \leqslant \sum_{j=1}^{p} \gamma_j \left|\epsilon_j' \mathbf{Q} \epsilon_j\right| \leqslant \sum_{j=1}^{p} \gamma_j \|\mathbf{Q}\epsilon_j\|_2 \leqslant \|\mathbf{Q}\| \operatorname{tr}(\mathbf{A}).$$

Second,

$$\log\det(\mathbf{AB}) = \log\det(\mathbf{B}^{1/2}\mathbf{AB}^{1/2}) \leqslant \operatorname{tr}(\mathbf{B}^{1/2}\mathbf{AB}^{1/2}) - p = \operatorname{tr}(\mathbf{AB}) - p.$$

Finally, $\operatorname{tr}(\mathbf{B}^{-1}\mathbf{A}+\mathbf{BA}^{-1}-2\mathbf{I}) = \operatorname{tr}(\mathbf{B}^{-1/2}\mathbf{AB}^{-1/2}+\mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}-2\mathbf{I})$. Let $\mathbf{C} = \mathbf{B}^{-1/2}\mathbf{AB}^{-1/2}$, then $\mathbf{C}$ is positive definite and $\mathbf{C}^{-1} = \mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}$. Since $\mathbf{C} + \mathbf{C}^{-1} - 2\mathbf{I}$ is positive definite [**?**], we conclude $\operatorname{tr}(\mathbf{B}^{-1}\mathbf{A} + \mathbf{A}^{-1}\mathbf{B} - 2\mathbf{I}) \geqslant 0$. According to Ex 12.14 [**?**] (page 329)

$$\operatorname{tr}(\mathbf{A}) \operatorname{tr}(\mathbf{B}) \geqslant \operatorname{tr}(\mathbf{AB}) \geqslant 0 \text{ and } \operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{B}^{1/2}\mathbf{AB}^{1/2}) \tag{4}$$

where $\mathbf{B}^{1/2}\mathbf{AB}^{1/2}$ is positive definite. Furthermore, according to the Cauchy-Schwarz Inequality (trace version) [?] (page 325)

$$\operatorname{tr}(\mathbf{A}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}) = \operatorname{tr}\left[(\mathbf{A}^{1/2})'(\mathbf{A}^{1/2})(\mathbf{A}^{-1/2}\mathbf{B})'(\mathbf{A}^{-1/2}\mathbf{B})\right) \geqslant \operatorname{tr}\left[((\mathbf{A}^{1/2})'\mathbf{A}^{-1/2}\mathbf{B})^2\right) = \operatorname{tr}(\mathbf{B}^2). \quad (5)$$

Hence,

$$\begin{aligned}
& \operatorname{tr}(\mathbf{A}\mathbf{B})\operatorname{tr}(\mathbf{B}^{-1}\mathbf{A} + \mathbf{A}^{-1}\mathbf{B} - 2\mathbf{I}) \text{ (Note:}\mathbf{AB}\text{ is not necessarily positive definite)} \\
= \ & \operatorname{tr}(\mathbf{B}^{1/2}\mathbf{A}\mathbf{B}^{1/2})\operatorname{tr}(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} + \mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2} - 2\mathbf{I}) \\
\geqslant \ & \operatorname{tr}(\mathbf{B}^{1/2}\mathbf{A}\mathbf{B}^{1/2}\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} + \mathbf{B}^{1/2}\mathbf{A}\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2} - 2\mathbf{B}^{1/2}\mathbf{A}\mathbf{B}^{1/2}) \\
= \ & \operatorname{tr}(\mathbf{B}^{1/2}\mathbf{A}^2\mathbf{B}^{-1/2} + \mathbf{B}^{1/2}\mathbf{A}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{1/2} - \mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A}) = \operatorname{tr}(\mathbf{A}^2 + \mathbf{A}\mathbf{B}\mathbf{A}^{-1}\mathbf{B} - \mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A}) \\
\geqslant \ & \operatorname{tr}\left[(\mathbf{A} - \mathbf{B})^2\right] = \|\mathbf{A} - \mathbf{B}\|_F^2.
\end{aligned}$$

The proof is complete.

Lemma 2 *Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two positive semidefinite $p \times p$ matrices, and*

$$\widehat{\mathbf{\Omega}}_1 = \operatorname{argmin}_{\mathbf{G}}\left(\mathcal{L}(\mathbf{S}_1; \mathbf{G}) + F_\lambda^\upsilon(\mathbf{G})\right); \quad \widehat{\mathbf{\Omega}}_2 = \operatorname{argmin}_{\mathbf{\Omega}}\left(\mathcal{L}(\mathbf{S}_2; \mathbf{G}) + F_\lambda^\upsilon(\mathbf{G})\right).$$

Then

$$\operatorname{tr}\left[(\mathbf{S}_2 - \mathbf{S}_1)(\widehat{\mathbf{\Omega}}_1 - \widehat{\mathbf{\Omega}}_2)\right] \geqslant \operatorname{tr}\left[\left(\widehat{\mathbf{\Omega}}_2^{-1} - \widehat{\mathbf{\Omega}}_1^{-1}\right)\left(\widehat{\mathbf{\Omega}}_1 - \widehat{\mathbf{\Omega}}_2\right)\right] \geqslant 0. \quad (6)$$

Proof: We establish

$$\mathbf{S}_1 - \widehat{\boldsymbol{\Omega}}_1^{-1} + \widehat{\mathbf{H}}_1 = 0; \quad \mathbf{S}_2 - \widehat{\boldsymbol{\Omega}}_2^{-1} + \widehat{\mathbf{H}}_2 = 0 \tag{7}$$

where $\widehat{\mathbf{H}}_1$ and $\widehat{\mathbf{H}}_2$ are $p \times p$ matrices of sub-differentials. Thus

$$\widehat{\mathbf{H}}_1 - \widehat{\mathbf{H}}_2 = \big[\mathbf{S}_2 - \mathbf{S}_1\big] + \big[\widehat{\boldsymbol{\Omega}}_1^{-1} - \widehat{\boldsymbol{\Omega}}_2^{-1}\big]. \tag{8}$$

Moreover,

$$\widehat{\boldsymbol{\Omega}}_1(h - h_2) = \widehat{\boldsymbol{\Omega}}_1\mathbf{S}_2 - \widehat{\boldsymbol{\Omega}}_1(\widehat{\boldsymbol{\Omega}}_2)^{-1} - \widehat{\boldsymbol{\Omega}}_1\mathbf{S}_1 + \widehat{\boldsymbol{\Omega}}_1(\widehat{\boldsymbol{\Omega}}_1)^{-1} \tag{9}$$

$$(-\widehat{\boldsymbol{\Omega}}_2)(h - h_2) = (-\widehat{\boldsymbol{\Omega}}_2)\mathbf{S}_2 - (-\widehat{\boldsymbol{\Omega}}_2)(\widehat{\boldsymbol{\Omega}}_2)^{-1} - (-\widehat{\boldsymbol{\Omega}}_2)\mathbf{S}_1 + (-\widehat{\boldsymbol{\Omega}}_2)(\widehat{\boldsymbol{\Omega}}_1)^{-1}. \tag{10}$$

Using the monotonicity property of convex functions, we establish $\operatorname{tr}\big[(\widehat{\mathbf{H}}_1 - \widehat{\mathbf{H}}_2)(\widehat{\boldsymbol{\Omega}}_1 - \widehat{\boldsymbol{\Omega}}_2)\big] \geqslant 0$, which implies

$$\operatorname{tr}\big[(\mathbf{S}_2 - \mathbf{S}_1)(\widehat{\boldsymbol{\Omega}}_1 - \widehat{\boldsymbol{\Omega}}_2)\big] \geqslant \operatorname{tr}\big[\big(\widehat{\boldsymbol{\Omega}}_2^{-1} - \widehat{\boldsymbol{\Omega}}_1^{-1}\big)\big(\widehat{\boldsymbol{\Omega}}_1 - \widehat{\boldsymbol{\Omega}}_2\big)\big] \geqslant 0. \tag{11}$$

The proof is complete.

(Note: The inequality $0 \leqslant \operatorname{tr}\big[(\boldsymbol{\Sigma} - \mathbf{S})(\widehat{\boldsymbol{\Omega}} - \widetilde{\boldsymbol{\Omega}})\big]$ is automatically implied by the above result.)

Lemma 3 *Let $\widetilde{\mathbf{H}}$ be the matrix of subdifferentials of $F_\lambda^v(.)$ evaluated at $\widetilde{\mathbf{\Omega}}$. Then*

$$0 \leqslant \mathcal{L}(\mathbf{\Sigma};\widetilde{\mathbf{\Omega}}) - \mathcal{L}(\mathbf{\Sigma};\mathbf{\Omega}) \leqslant F_{\lambda,v}(\mathbf{\Omega}) - F_{\lambda,v}(\widetilde{\mathbf{\Omega}}) \tag{12}$$

$$\textit{For } L_1 \textit{ penalty: } F_\lambda^1(\widetilde{\mathbf{\Omega}}) = \mathrm{tr}\left[\widetilde{\mathbf{H}}\widetilde{\mathbf{\Omega}}\right] = \lambda\|\widetilde{\mathbf{\Omega}}\|_1 \leqslant p; \quad \mathrm{tr}\left[\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}\right] + \lambda\|\widetilde{\mathbf{\Omega}}\|_1 = p. \tag{13}$$

Proof: Obviously,

$$\mathrm{tr}(\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}) - \log\det(\widetilde{\mathbf{\Omega}}) + F_{\lambda,v}(\widetilde{\mathbf{\Omega}}) \;\leqslant\; \mathrm{tr}(\mathbf{\Sigma}\mathbf{\Omega}) - \log\det(\mathbf{\Omega}) + F_{\lambda,v}(\mathbf{\Omega}).$$

Thus we conclude $0 \leqslant \mathcal{L}(\mathbf{\Sigma};\widetilde{\mathbf{\Omega}}) - \mathcal{L}(\mathbf{\Sigma};\mathbf{\Omega}) \leqslant F_{\lambda,v}(\mathbf{\Omega}) - F_{\lambda,v}(\widetilde{\mathbf{\Omega}})$.

Furthermore, we establish

$$\mathbf{\Sigma} - \widetilde{\mathbf{\Omega}}^{-1} + \widetilde{\mathbf{H}} = 0; \quad \mathbf{\Sigma}\widetilde{\mathbf{\Omega}} - \mathbf{I}_p + \widetilde{\mathbf{H}}\widetilde{\mathbf{\Omega}} = 0; \quad \mathrm{tr}\left[\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}\right] + \mathrm{tr}\left[\widetilde{\mathbf{H}}\widetilde{\mathbf{\Omega}}\right] = p \tag{14}$$

where $\mathrm{tr}\left[\widetilde{\mathbf{H}}\widetilde{\mathbf{\Omega}}\right] = \lambda\|\widetilde{\mathbf{\Omega}}\|_1$ for L_1 penalty.

The proof is complete.

Lemma 4 *Let $\widehat{\mathbf{H}}$ be the matrix of subdifferentials of $F_\lambda^v(.)$ evaluated at $\widehat{\mathbf{\Omega}}$. Then for L_1 penalty:*

$$\mathrm{tr}(\mathbf{S}\widehat{\mathbf{\Omega}}) + \lambda\|\widehat{\mathbf{\Omega}}\|_1 = p; \tag{15}$$

$$-F_\lambda^v(\widetilde{\mathbf{\Omega}}) \leqslant \mathcal{L}(\mathbf{S};\widetilde{\mathbf{\Omega}}) - \mathcal{L}(\mathbf{S};\widehat{\mathbf{\Omega}}) \leqslant p\|\mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2} - \mathbf{I}_p\| + \mathrm{tr}(\mathbf{\Sigma}) + \lambda p + p\log(1/\lambda). \tag{16}$$

Proof: We establish

$$\mathbf{S} - \widehat{\mathbf{\Omega}}^{-1} + \widehat{\mathbf{H}} = 0; \quad \mathbf{S}\widehat{\mathbf{\Omega}} - \mathbf{I}_p + \widehat{\mathbf{H}}\widehat{\mathbf{\Omega}} = 0; \quad \mathrm{tr}\left[\mathbf{S}\widehat{\mathbf{\Omega}}\right] + \mathrm{tr}\left[\widehat{\mathbf{H}}\widehat{\mathbf{\Omega}}\right] = p \tag{17}$$

where $F_\lambda^1(\widehat{\mathbf{\Omega}}) = \lambda\|\widehat{\mathbf{\Omega}}\|_1 = \mathrm{tr}\left[\widehat{\mathbf{H}}\widehat{\mathbf{\Omega}}\right] \leqslant p$ for the Lasso penalty.

Since $\mathrm{tr}(\mathbf{S}\widetilde{\mathbf{\Omega}}) - \log\det(\widetilde{\mathbf{\Omega}}) + F_\lambda^v(\widetilde{\mathbf{\Omega}}) \geqslant \mathrm{tr}(\mathbf{S}\widehat{\mathbf{\Omega}}) - \log\det(\widehat{\mathbf{\Omega}}) + F_\lambda^v(\widehat{\mathbf{\Omega}})$,

$$\mathcal{L}(\mathbf{S};\widetilde{\mathbf{\Omega}}) - \mathcal{L}(\mathbf{S};\widehat{\mathbf{\Omega}}) \geqslant F_\lambda^v(\widehat{\mathbf{\Omega}}) - F_\lambda^v(\widetilde{\mathbf{\Omega}}) \geqslant -F_\lambda^v(\widetilde{\mathbf{\Omega}}). \tag{18}$$

On the other hand, since the sample covariance matrix $\mathbf{S}$ may be singular,

$$\begin{aligned}
& \mathcal{L}(\mathbf{S};\widetilde{\mathbf{\Omega}}) - \mathcal{L}(\mathbf{S};\widehat{\mathbf{\Omega}}) = \mathrm{tr}\left[\mathbf{S}(\widetilde{\mathbf{\Omega}} - \widehat{\mathbf{\Omega}})\right] - \log\det(\widetilde{\mathbf{\Omega}}) + \log\det(\widehat{\mathbf{\Omega}}) \\
= & \ \mathrm{tr}\left[(\mathbf{S} - \mathbf{\Sigma})\widetilde{\mathbf{\Omega}}\right] + \mathrm{tr}\left[\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}\right] + \log\det(\widetilde{\mathbf{\Omega}}^{-1}) - \mathrm{tr}\left[\mathbf{S}\widehat{\mathbf{\Omega}}\right] + \log\det((\mathbf{S} + \lambda\mathbf{I}_p)^{-1}) \\
& + \log\det((\mathbf{S} + \lambda\mathbf{I}_p)\widehat{\mathbf{\Omega}}) \\
\leqslant & \ \|\mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2} - \mathbf{I}_p\|\,\mathrm{tr}(\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}) + \mathrm{tr}(\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}) + \mathrm{tr}\,\mathbf{\Sigma} + \lambda p - p + p\log(1/\lambda) + \mathrm{tr}(\mathbf{S}\widehat{\mathbf{\Omega}}) + \lambda\|\widehat{\mathbf{\Omega}}\|_1 - p \\
\leqslant & \ p\|\mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2} - \mathbf{I}_p\| + \mathrm{tr}(\mathbf{\Sigma}) + \lambda p + p\log(1/\lambda).
\end{aligned}$$

The proof is complete.

LEMMA 5 *Let $\widehat{\mathbf{H}}^*$ be the matrix of subdifferentials of $F_\lambda^v(.)$ evaluated at $\widehat{\mathbf{\Omega}}^*$. Then for L_1*

penalty:

$$\mathrm{tr}(\widehat{\mathbf{\Omega}}^*\widehat{\mathbf{H}}^*) + \lambda\|\widehat{\mathbf{\Omega}}^*\|_1 = p; \tag{19}$$

$$-F_\lambda^v(\widetilde{\mathbf{\Omega}}) \leqslant \mathcal{L}(\mathbf{S}^*;\widetilde{\mathbf{\Omega}}) - \mathcal{L}(\mathbf{S}^*;\widehat{\mathbf{\Omega}}^*) \leqslant p\|\mathbf{\Omega}^{1/2}\mathbf{S}^*\mathbf{\Omega}^{1/2} - \mathbf{I}_p\| + \mathrm{tr}(\mathbf{\Sigma}) + p\lambda + p\log(1/\lambda). \tag{20}$$

Proof: We establish

$$\mathbf{S}^* - (\widehat{\mathbf{\Omega}}^*)^{-1} + \widehat{\mathbf{H}}^* = 0; \quad \mathbf{S}^*\widehat{\mathbf{\Omega}}^* - \mathbf{I}_p + \widehat{\mathbf{H}}^*\widehat{\mathbf{\Omega}}^* = 0; \quad \mathrm{tr}\left[\mathbf{S}^*\widehat{\mathbf{\Omega}}^*\right] + \mathrm{tr}\left[\widehat{\mathbf{H}}^*\widehat{\mathbf{\Omega}}\right] = p \tag{21}$$

where $F_\lambda^1(\widehat{\mathbf{\Omega}}^*) = \lambda\|\widehat{\mathbf{\Omega}}^*\|_1 = \mathrm{tr}\left[\widehat{\mathbf{H}}^*\widehat{\mathbf{\Omega}}^*\right] \leqslant p$ for the Lasso penalty.

Since $\mathrm{tr}(\mathbf{S}^*\widetilde{\mathbf{\Omega}}) - \log\det(\widetilde{\mathbf{\Omega}}) + F_\lambda^v(\widetilde{\mathbf{\Omega}}) \geqslant \mathrm{tr}(\mathbf{S}^*\widehat{\mathbf{\Omega}}^*) - \log\det(\widehat{\mathbf{\Omega}}^*) + F_\lambda^v(\widehat{\mathbf{\Omega}}^*)$,

$$\mathcal{L}(\mathbf{S}^*;\widetilde{\mathbf{\Omega}}) - \mathcal{L}(\mathbf{S}^*;\widehat{\mathbf{\Omega}}^*) \geqslant F_\lambda^v(\widehat{\mathbf{\Omega}}^*) - F_\lambda^v(\widetilde{\mathbf{\Omega}}) \geqslant -F_\lambda^v(\widetilde{\mathbf{\Omega}}). \tag{22}$$

On the other hand, since the sample covariance matrix $\mathbf{S}^*$ may be singular,

$$\begin{aligned}
& \mathrm{tr}\left[\mathbf{S}^*(\widetilde{\mathbf{\Omega}} - \widehat{\mathbf{\Omega}}^*)\right] - \log\det(\widetilde{\mathbf{\Omega}}) + \log\det(\widehat{\mathbf{\Omega}}^*) \\
= \; & \mathrm{tr}\left[(\mathbf{S}^* - \mathbf{\Sigma})\widetilde{\mathbf{\Omega}}\right] + \mathrm{tr}(\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}) + \log\det(\widetilde{\mathbf{\Omega}}^{-1}) - \mathrm{tr}\left[\mathbf{S}^*\widehat{\mathbf{\Omega}}^*\right] + \log\det((\mathbf{S}^* + \lambda\mathbf{I}_p)^{-1}) \\
& + \log\det((\mathbf{S}^* + \lambda\mathbf{I}_p)\widehat{\mathbf{\Omega}}^*) \\
\leqslant \; & \mathrm{tr}\left[(\mathbf{S}^* - \mathbf{\Sigma})\widetilde{\mathbf{\Omega}}\right] + \mathrm{tr}(\mathbf{\Sigma}\widetilde{\mathbf{\Omega}}) + \mathrm{tr}(\mathbf{\Sigma}) + p\lambda - p + p\log(1/\lambda) + \mathrm{tr}(\mathbf{S}^*\widehat{\mathbf{\Omega}}^*) + \lambda\|\widehat{\mathbf{\Omega}}^*\|_1 - p \\
\leqslant \; & \mathrm{tr}\left[(\mathbf{S}^* - \mathbf{\Sigma})\widetilde{\mathbf{\Omega}}\right] + \mathrm{tr}(\mathbf{\Sigma}) + p\lambda + p\log(1/\lambda) \\
\leqslant \; & p\|\mathbf{\Omega}^{1/2}\mathbf{S}^*\mathbf{\Omega}^{1/2} - \mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2}\| + p\|\mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2} - \mathbf{I}_p\| + \mathrm{tr}(\mathbf{\Sigma}) + p\lambda + p\log(1/\lambda).
\end{aligned}$$

The proof is complete.

Proof of Theorem 1: The operator norm $\|\mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2} - \mathbf{I}_p\|$ is bounded by

$$E\|\mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2} - \mathbf{I}_p\| \leqslant CK^2\Big(\sqrt{\frac{p}{n}} + \frac{p}{n}\Big)$$
$$\|\mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2} - \mathbf{I}_p\| \leqslant CK^2\Big(\sqrt{\frac{p+u}{n}} + \frac{p+u}{n}\Big) \text{ with probability at least } 1 - 2\exp(u)$$

using a result of [?] (page 99-100), and the constants K and C are defined as there.

According to [?] (page 129-130),

$$E\Big(\|\mathbf{\Omega}^{1/2}\mathbf{S}^*\mathbf{\Omega}^{1/2} - \mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2}\|\Big|\boldsymbol{Y}\Big) \leqslant C\Big(\sqrt{\frac{\widehat{K}^2 p \log p}{n}} + \frac{\widehat{K}^2 p \log p}{n}\Big);$$
$$\|\mathbf{\Omega}^{1/2}\mathbf{S}^*\mathbf{\Omega}^{1/2} - \mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2}\| \leqslant C\Big(\sqrt{\frac{\widehat{K}^2 p(\log p + u)}{n}} + \frac{\widehat{K}^2 p(\log p + u)}{n}\Big)$$

with probability at least $1 - 2\exp(u)$, where C is a absolute constant and

$$\widehat{K} = \frac{\max_{1\leqslant i\leqslant n}\{\|\mathbf{Z}_1\|_2^2, \cdots, \|\mathbf{Z}_n\|_2^2\}}{(1/n)\sum_{i=1}^n \|\mathbf{Z}_i\|_2^2}. \tag{23}$$

As for the Mallow's metric $d_1(D_t^*, D_t)$,

$$\frac{1}{p^2} d_1\big(\mathcal{L}(\mathbf{S}^*; \widehat{\mathbf{\Omega}}^*), \mathcal{L}(\mathbf{S}; \widehat{\mathbf{\Omega}})\big) \leqslant A_n + B_n + C_n \tag{24}$$

where

$$A_n = \frac{1}{p^2} d_1\big(\mathcal{L}(\mathbf{S}^*;\widehat{\mathbf{\Omega}}^*), \mathcal{L}(\mathbf{S}^*;\widetilde{\mathbf{\Omega}})\big); \tag{25}$$

$$B_n = \frac{1}{p^2} d_1\big(\mathcal{L}(\mathbf{S}^*;\widetilde{\mathbf{\Omega}}), \mathcal{L}(\mathbf{S};\widetilde{\mathbf{\Omega}})\big); \tag{26}$$

$$C_n = \frac{1}{p^2} d_1\big(\mathcal{L}(\mathbf{S};\widetilde{\mathbf{\Omega}}), \mathcal{L}(\mathbf{S};\widehat{\mathbf{\Omega}})\big). \tag{27}$$

Considering the independence of $\mathbf{Y}_i^*\widetilde{\mathbf{\Omega}}\mathbf{Y}_i^*$ $(i = 1, \cdots, n)$ and the independence of $\mathbf{Y}_i'\widetilde{\mathbf{\Omega}}\mathbf{Y}_i$ $(i = 1, \cdots, n)$, we establish

$$B_n \leqslant \frac{1}{p^2} d_1\big(\operatorname{tr}(\mathbf{S}^*\widetilde{\mathbf{\Omega}}), \operatorname{tr}(\mathbf{S}\widetilde{\mathbf{\Omega}})\big) \leqslant \frac{1}{np^2} d_1\Big(\sum_{i=1}^n (\mathbf{Y}_i^*)'\widetilde{\mathbf{\Omega}}\mathbf{Y}_i^*, \sum_{i=1}^n \mathbf{Y}_i'\widetilde{\mathbf{\Omega}}\mathbf{Y}_i\Big) \leqslant \frac{1}{p^2} d_1\big((\mathbf{Y}_1^*)'\widetilde{\mathbf{\Omega}}\mathbf{Y}_1^*, \mathbf{Y}_1'\widetilde{\mathbf{\Omega}}\mathbf{Y}_1\big).$$

Let $\mathbf{Y} \sim MVN(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{Z} = \mathbf{\Omega}^{1/2}\mathbf{Y} \sim MVN(\mathbf{0}, \mathbf{I}_p)$. Then $Var(\mathbf{Y}'\mathbf{G}\mathbf{Y}) = 2\operatorname{tr}(\mathbf{G\Sigma G\Sigma})$ where

$$\operatorname{tr}(\mathbf{G\Sigma G\Sigma}) = \operatorname{tr}(\mathbf{G^{1/2}\Sigma G^{1/2} G^{1/2}\Sigma G^{1/2}}) \leqslant \big(\operatorname{tr}(\mathbf{G^{1/2}\Sigma G^{1/2}})\big)^2 = \big(\operatorname{tr}(\mathbf{G\Sigma})\big)^2.$$

Let $\mathbf{S}_0 = \mathbf{\Omega}^{1/2}\mathbf{S}\mathbf{\Omega}^{1/2}$. Then $E(\mathbf{Z}^*(\mathbf{Z}^*)'|\boldsymbol{Y}) = \mathbf{S}_0$.

Consider

$$Var((\mathbf{Y}^*)'\mathbf{G}\mathbf{Y}^*|\boldsymbol{Y}) = E\big(\operatorname{tr}^2\big[\mathbf{\Omega}^{1/2}\mathbf{Y}^*(\mathbf{Y}^*)'\mathbf{\Omega}^{1/2} - \mathbf{S}_0)(\mathbf{\Sigma}^{1/2}\mathbf{G}\mathbf{\Sigma}^{1/2})\big]|\boldsymbol{Y}\big)$$

$$\leqslant \operatorname{tr}^2(\mathbf{\Sigma}^{1/2}\mathbf{G}\mathbf{\Sigma}^{1/2})E\big(\|\mathbf{\Omega}^{1/2}\mathbf{Y}^*(\mathbf{Y}^*)'\mathbf{\Omega}^{1/2} - \mathbf{S}_0)\|^2|\boldsymbol{Y}\big) \leqslant \operatorname{tr}^2(\mathbf{G\Sigma})E\big(\|\mathbf{Z}^*(\mathbf{Z}^*)' - \mathbf{S}_0)\|^2|\boldsymbol{Y}\big).$$

Obviously, $\{\mathbf{Z}^*(\mathbf{Z}^*)' - \mathbf{S}_0)\}^2$ is a symmetric and positive definite matrix and

$$\begin{aligned} & \|\mathbf{Z}^*(\mathbf{Z}^*)' - \mathbf{S}_0)\|^2 \leqslant \|\{\mathbf{Z}^*(\mathbf{Z}^*)' - \mathbf{S}_0\}^2\| \leqslant \operatorname{tr}\left[\{\mathbf{Z}^*(\mathbf{Z}^*)' - \mathbf{S}_0\}^2\right] \\ = \; & \operatorname{tr}\left[\{\mathbf{Z}^*(\mathbf{Z}^*)'\}^2\right] + \operatorname{tr}\left[\{\mathbf{S}_0\}^2\right] - 2\operatorname{tr}\left[\mathbf{Z}^*(\mathbf{Z}^*)'\mathbf{S}_0\right]. \end{aligned}$$

Therefore,

$$\begin{aligned} E\big(\|\mathbf{Z}^*(\mathbf{Z}^*)' - \mathbf{S}_0)\|^2|\boldsymbol{Y}\big) &= E\big(\operatorname{tr}\left[\{\mathbf{Z}^*(\mathbf{Z}^*)'\}^2\right]|\boldsymbol{Y}\big) - \operatorname{tr}\left[\mathbf{S}_0^2\right] = E\big(\|\mathbf{Z}^*\|_2^4|\boldsymbol{Y}\big) - \operatorname{tr}\left[\mathbf{S}_0^2\right] \\ &= Var\big(\|\mathbf{Z}^*\|_2^2|\boldsymbol{Y}\big) \leqslant p\sum Var\big((\mathbf{Z}_{[i]})^2|\boldsymbol{Y}\big). \end{aligned}$$

Therefore,

$$\begin{aligned} & Var(\mathbf{Y}'\widetilde{\boldsymbol{\Omega}}\mathbf{Y}) \leqslant 2\operatorname{tr}(\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Sigma}) \leqslant 2\big(\operatorname{tr}(\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Sigma})\big)^2 \leqslant 2p^2; \\ & Var((\mathbf{Y}^*)'\mathbf{G}\mathbf{Y}^*|\boldsymbol{Y}) \leqslant p^3\sum Var\big((\mathbf{Z}_{[i]})^2|\boldsymbol{Y}\big). \end{aligned}$$

Furthermore, since $\mathbf{Y}_i^{*'}\widetilde{\boldsymbol{\Omega}}\mathbf{Y}_i^*$ $(i = 1, \cdots, n)$ are a bootstrapping sample from $\mathbf{Y}_i'\widetilde{\boldsymbol{\Omega}}\mathbf{Y}_i$ $(i = 1, \cdots, n)$,

$$B_n \leqslant \frac{1}{p^2} d_1\big(\mathbf{Y}_1^{*'}\widetilde{\boldsymbol{\Omega}}\mathbf{Y}_1^*, \mathbf{Y}_1'\widetilde{\boldsymbol{\Omega}}\mathbf{Y}_1\big) \to 0 \text{ almost surely as } n \to \infty.$$

Both A_n and C_n converge to 0 almost surely by Lemmas 4 and 5.

The proof is complete.