SUPPLEMENT TO "GEMINI: GRAPH ESTIMATION WITH MATRIX VARIATE NORMAL INSTANCES"

By Shuheng Zhou

9. Notation and an outline. Suppose that we have n i.i.d. random matrices $X^n = (X(1), X(2), \ldots, X(n))$ where $X(i) \sim \mathcal{N}_{f,m}(0, A_0 \otimes B_0)$ for all i, and $A_0 = (a_{jk})$ and $B_0 = (b_{jk})$ are positive definite. Let $Y(t) = X(t)^T$, $\widehat{S}_n = \frac{1}{n}\sum_{t=1}^n \mathrm{vec}\{\mathrm{X(t)}\}\mathrm{vec}\{\mathrm{X(t)}\}^{\mathrm{T}}$, and $\widetilde{\mathrm{S}}_{\mathrm{n}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{t=1}}^{\mathrm{n}} \mathrm{vec}\{\mathrm{Y(t)}\}\mathrm{vec}\{\mathrm{Y(t)}\}^{\mathrm{T}}$. Denote the ℓ, k^{th} block of size $f \times f$ in $\widehat{S}_n$ by $\widehat{S}_n^{\ell k} = (\widehat{S}_{ij}^{\ell k})$ and that of size $m \times m$ in $\widetilde{S}_n$ by $\widetilde{S}_n^{\ell k} = (\widetilde{S}_{ij}^{\ell k})$. Let $x(t)^1, \ldots, x(t)^m \in \mathbf{R}^f$ be column vectors and $y(t)^1, \ldots, y(t)^f \in \mathbf{R}^m$ be row vectors of matrix $X(t)$. Then

$$\begin{aligned}
\tfrac{1}{n}\textstyle\sum_{t=1}^n x(t)^\ell \otimes x(t)^k &= \widehat{S}_n^{\ell k} \quad \text{where} \quad \mathbb{E}\left[x(t)^\ell \otimes x(t)^k\right] = a_{\ell k} B_0, \\
\tfrac{1}{n}\textstyle\sum_{t=1}^n y(t)^\ell \otimes y(t)^k &= \widetilde{S}_n^{\ell k} \quad \text{where} \quad \mathbb{E}\left[y(t)^\ell \otimes y(t)^k\right] = b_{\ell k} A_0.
\end{aligned}$$

We now state the following shorthand notation. Let $a_{\min} = \min_i A_{0,ii}$ and $b_{\min} = \min_i B_{0,ii}$. By positive-definiteness of A_0 and B_0, we have $a_{\min} > 0$ and $b_{\min} > 0$. Let $a_{\max} = \max_i A_{0,ii}$ and $b_{\max} = \max_i B_{0,ii}$. Then $0 < a_{\min} \leq a_{\max} \leq \|A_0\|_2 < +\infty$ and $0 < b_{\min} \leq b_{\max} \leq \|B_0\|_2 < +\infty$ on (A2). Recall

$$A_* = mA_0/\mathrm{tr}(A_0) \quad \text{and} \quad B_* = B_0 \mathrm{tr}(A_0)/m. \tag{43}$$

For two matrices M, A, let $\langle A, M \rangle$ denote $\mathrm{tr}(AM)$.

We state the rate of convergence for the Gemini estimators in Section 10. We prove Theorems 3.1 and 3.2 in Section 11, where we state the absolute error bounds in the operator and the Frobenius norm for estimating both $A_0 \otimes B_0$ and its inverse in Theorems 11.1 and 11.2. The proofs of Theorems 11.1 and 11.2 appear in Section 15. We defer the proof of Theorem 3.3 to Section 16. We prove Theorem 4.1 and Lemma 4.3 in Section 12. In Section 13, we prove a large deviation inequality in Theorem 13.1 for the weighted sum of submatrices $\widehat{S}_n^{\ell k}, \ell, k = 1, \ldots, m$, where the weights correspond to entries in an $m \times m$ matrix M. Lemma 4.3 is a special case of Theorem 13.1 when we set $M = I$. Proofs of Theorem 4.5 and Corollary 14.1 are given in Section 14. Proofs for the Non-iterative Penalized Flip-Flop algorithm appear in Section 17.

10. Rates of convergence for Covariance matrices. Corollary 10.1 provides rates of convergence in the operator and the Frobenius norm for estimating A_*, B_* (and their inverses), with

$$\widehat{A}_* = m\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1 / (\tfrac{1}{n}\textstyle\sum_{i=1}^n \|X(i)\|_F^2) \quad \text{and} \quad \widehat{B}_* = \widehat{W}_2 \widehat{B}_\rho \widehat{W}_2 / m \tag{44}$$

Clearly $\widehat{A}_* \otimes \widehat{B}_* = \widehat{A \otimes B}$.

COROLLARY 10.1. *Let A_*, B_* be as defined in (43). Let $a_{*,\min} = \min_i A_{*,ii}$ and $a_{*,\max} = \max_i A_{*,ii}$. Let $b_{*,\min} = \min_i B_{*,ii}$ and $b_{*,\max} = \max_i B_{*,ii}$. Let $\widehat{A}_*, \widehat{B}_*$ be as defined in (44). Suppose (A1) and (A2) hold. Suppose event $\mathcal{X}_0$ holds. For some absolute constants $39/2 < C, C' < 39$,*

$$\begin{aligned}
\left\|\widehat{A}_* - A_*\right\|_2 &\leq 2C\lambda_{B_0} a_{*,\max} \kappa\left(\rho(A_0)\right)^2 \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}, \\
\left\|\widehat{B}_* - B_*\right\|_2 &\leq C'\lambda_{A_0} b_{*,\max} \kappa\left(\rho(B_0)\right)^2 \sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1}.
\end{aligned}$$

For the inverses, we have for $11 < C, C' < 19$,

$$\begin{aligned}
\left\|\widehat{A}_*^{-1} - A_*^{-1}\right\|_2 &\leq 2C\lambda_{B_0} \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1} / (a_{*,\min} \varphi^2_{\min}(\rho(A_0))), \\
\textit{and} \quad \left\|\widehat{B}_*^{-1} - B_*^{-1}\right\|_2 &\leq C'\lambda_{A_0} \sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1} / (b_{*,\min} \varphi^2_{\min}(\rho(B_0))).
\end{aligned}$$

For the Frobenius norm, we have for $39/2 < C, C' < 39$,

$$\begin{aligned}
\left\|\widehat{A}_* - A_*\right\|_F &\leq 2C a_{*,\max} \kappa\left(\rho(A_0)\right)^2 \lambda_{B_0} \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}, \\
\left\|\widehat{B}_* - B_*\right\|_F &\leq C' b_{*,\max} \kappa\left(\rho(B_0)\right)^2 \lambda_{A_0} \sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f},
\end{aligned}$$

and for $11 < C, C' < 19$,

$$\begin{aligned}
\left\|\widehat{A}_*^{-1} - A_*^{-1}\right\|_F &\leq 2C\lambda_{B_0} \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m} / (a_{*,\min} \varphi^2_{\min}(\rho(A_0))), \\
\textit{and} \quad \left\|\widehat{B}_*^{-1} - B_*^{-1}\right\|_F &\leq C'\lambda_{A_0} \sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f} / (b_{*,\min} \varphi^2_{\min}(\rho(B_0))).
\end{aligned}$$

We prove Corollary 10.1 in Section 15.2.

11. Proofs of Theorems 3.1 and 3.2. Theorem 11.1 and 11.2 show the absolute error bounds in the operator and the Frobenius norm for estimating both $A_0 \otimes B_0$ and its inverse, which will depend on n through λ_{A_0} and λ_{B_0}, which in turn depend on the rates of convergence α_n and β_n in entry-wise max norm in estimating $\rho(B_0)$ and $\rho(A_0)$. We show an outline for proving Theorems 11.1 and 11.2 in Section 15, with actual proof in Section 15.3 and 15.4.

THEOREM 11.1. *Let* $\lambda_{A_0} \wedge \lambda_{B_0} := \min\{\lambda_{A_0}, \lambda_{B_0}\}$. *Suppose (A1) and (A2) hold. Suppose that* $\lambda_{A_0}, \lambda_{B_0} < 1$, *and for some* $0 < \varepsilon_1, \varepsilon_2 < 1$, *we set*

$$\lambda_{A_0} := 3\alpha_n/\varepsilon_1 \quad \text{and} \quad \lambda_{B_0} = 3\beta_n/\varepsilon_2,$$

where α_n *and* β_n *are as defined in Theorem 4.1. Then on event* $\mathcal{X}_0$, *where* $\mathbb{P}(\mathcal{X}_0) \geq 1 - \frac{3}{(m \vee f)^2}$, *we have for* $18 < C, C' < 36$,

$$\begin{aligned}\left\|\widehat{A \otimes B} - A_0 \otimes B_0\right\|_2 \leq \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2} \|A_0\|_2 \|B_0\|_2 + C\lambda_{B_0} a_{\max} \|B_0\|_2 \kappa(\rho(A_0))^2 \times \\ \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1} + C'\lambda_{A_0} b_{\max} \|A_0\|_2 \kappa(\rho(B_0))^2 \sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1} + \\ 2CC'\lambda_{A_0}\lambda_{B_0} a_{\max} b_{\max} \kappa(\rho(A_0))^2 \kappa(\rho(B_0))^2 \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1},\end{aligned}$$

and for $10 < C, C' < 19$,

$$\begin{aligned}\left\|\widehat{A \otimes B}^{-1} - A_0^{-1} \otimes B_0^{-1}\right\|_2 &\leq \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{3} \|A_0^{-1}\|_2 \|B_0^{-1}\|_2 \\ &+ \lambda_{B_0} \|B_0^{-1}\|_2 \frac{C\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}}{a_{\min}\varphi^2_{\min}(\rho(A_0))} + \lambda_{A_0} \|A_0^{-1}\|_2 \frac{C'\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1}}{b_{\min}\varphi^2_{\min}(\rho(B_0))} \\ &+ \frac{3CC'\lambda_{A_0}\lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1}}{2a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))}.\end{aligned}$$

THEOREM 11.2. *Suppose all conditions in Theorem 11.1 hold. Suppose (A1) and (A2) hold. Then on event* $\mathcal{X}_0$, *we have for* $18 < C, C' < 36$,

$$\begin{aligned}\left\|\widehat{A \otimes B} - A_0 \otimes B_0\right\|_F \leq \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2} \|A_0 \otimes B_0\|_F + C\lambda_{B_0} a_{\max} \|B_0\|_F \kappa(\rho(A_0))^2 \times \\ \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m} + C'\lambda_{A_0} b_{\max} \|A_0\|_F \kappa(\rho(B_0))^2 \sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f} \\ + \quad 2CC'\lambda_{A_0}\lambda_{B_0} a_{\max} b_{\max} \kappa(\rho(A_0))^2 \kappa(\rho(B_0))^2 \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f},\end{aligned}$$

and for $10 < C, C' < 19$,

$$\begin{aligned}\left\|\widehat{A \otimes B}^{-1} - A_0^{-1} \otimes B_0^{-1}\right\|_F &\leq \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{3} \|A_0^{-1}\|_F \|B_0^{-1}\|_F \\ &+ \lambda_{B_0} \|B_0^{-1}\|_F \frac{C\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}}{a_{\min}\varphi^2_{\min}(\rho(A_0))} + \lambda_{A_0} \|A_0^{-1}\|_F \frac{C'\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f}}{b_{\min}\varphi^2_{\min}(\rho(B_0))} \\ &+ \frac{7CC'\lambda_{A_0}\lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f}}{5a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))}.\end{aligned}$$

Remark 11.3. *For the two cases considered in Theorem 3.2, namely, when either (i)* $1 \le n \le \log(m \vee f)$ *Or (ii)* $\left|A_0^{-1}\right|_{0,\text{off}} = O(m)$ *and* $\left|B_0^{-1}\right|_{0,\text{off}} = O(f)$*, we have under (A1) and (A2) and for* $\tau_0 \asymp \log^{1/2}(m \vee f)/\sqrt{n}$

$$\left\|\widehat{A \otimes B} - A_0 \otimes B_0\right\|_F \le \delta \left\|A_0\right\|_F \left\|B_0\right\|_F = O\left(\tau_0(\sqrt{m} + \sqrt{f})\right).$$

Suppose that $n > \log(m \vee f)$ *is moderately large. Suppose that* $\left|A_0^{-1}\right|_{0,\text{off}} \ge m$ *and* $\left|B_0^{-1}\right|_{0,\text{off}} \ge f$*. Then under (A1) and (A2),*

$$\delta = O\left(\lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}}/m} + \lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}/f}\right) = o\left(\frac{1}{\sqrt{f}} + \frac{1}{\sqrt{m}}\right) \tag{45}$$

and

$$\left\|\widehat{A \otimes B} - A_0 \otimes B_0\right\|_F \le \delta \left\|A_0\right\|_F \left\|B_0\right\|_F \asymp o\left(\sqrt{m} + \sqrt{f}\right)$$

where we assume that both $\left|A_0^{-1}\right|_{0,\text{off}}$ *and* $\left|B_0^{-1}\right|_{0,\text{off}}$ *are allowed to grow linearly with* n *in the worst case, and hence the bound on* δ *becomes independent of* n*. We note that for the rate we calculated in (45), we assume that* n *is not too large; otherwise, one can refine the calculations a bit further. For example, suppose that* $n > \log(m \vee f)\left(\frac{f^2}{m} \vee \frac{m^2}{f}\right)$*. Then (A1) becomes vacuous and trivially, under (A2)* $\delta = O\left(\lambda_{B_0}\sqrt{m} + \lambda_{A_0}\sqrt{f}\right) = O\left(\tau_0(m+f)/\sqrt{fm}\right)$*. Hence when* $n \asymp \log(m \vee f)\left(\frac{f^2}{m} \vee \frac{m^2}{f}\right)$*, we have* $\delta = o\left(\frac{1}{\sqrt{m}} \wedge \frac{1}{\sqrt{f}}\right)$ *and*

$$\left\|\widehat{A \otimes B} - A_0 \otimes B_0\right\|_F \le \delta \left\|A_0\right\|_F \left\|B_0\right\|_F \asymp o(\sqrt{f} \wedge \sqrt{m}).$$

We do not pursue such refinements in this work.

To understand the rates for $n > 1$*, build the connection between the one-matrix and the multiple-matrix cases, we first imagine stacking matrices* $X(1), \ldots, X(n)$ *on top of each other to form a single* $nf \times m$ *matrix* X'*. This way, we can then imagine being in the situation of one-matrix case: where* $X' \sim \mathcal{N}_{nf,m}(0, A_0 \otimes B')$*, where the dimension of* B' *is* $nf \times nf$ *respectively. Similar to the general case with one data matrix, we will use a sample of size* nf *to estimate the structure and parameters for* A_0*; however, unlike the one matrix case we have focused on so far,* B' *has an additional structural property other than the assumed sparsity on its inverse which allows for a faster rate of convergence, namely, matrix* B' *is block diagonal with* n *identical submatrices* B_0 *along its diagonal. Hence the estimation step for* B' *takes advantage of this knowledge by stacking* $Y(1), \ldots, Y(n)$ *on top of each other, where* $Y(i) = X(i)^T$*, to form a* $nm \times f$ *matrix* $Y' \sim \mathcal{N}_{nm,f}(0, B_0 \otimes A')$*, where matrix* A' *is in turn block diagonal with* n *identical submatrices* A_0

staying along its diagonal. We treat Y' as having nm samples, all of which subject to normalization come from the same multivariate normal distribution $\mathcal{N}_f(0, B_0)$ as shown in Figure 1. This enables faster convergence rates for $n > 1$ as shown in all theorems in Section 3.

We now state some more shorthand notation and convenient bounds.

$$r_a := a_{\max}/a_{\min} \quad \text{and} \quad r_b := b_{\max}/b_{\min} \tag{46}$$

$$1/\varphi_{\min}(A_0) = \left\|A_0^{-1}\right\|_2 \le \left\|\rho(A_0)^{-1}\right\|_2/a_{\min} = \frac{1}{a_{\min}\varphi_{\min}(\rho(A_0))}, \tag{47}$$

$$1/\varphi_{\min}(B_0) = \left\|B_0^{-1}\right\|_2 \le \left\|\rho(B_0)^{-1}\right\|_2/b_{\min} = \frac{1}{b_{\min}\varphi_{\min}(\rho(B_0))}, \tag{48}$$

$$1/\varphi_{\min}(\rho(A_0)) = \left\|\rho(A_0)^{-1}\right\|_2 \le a_{\max}\left\|A_0^{-1}\right\|_2, \tag{49}$$

$$1/\varphi_{\min}(\rho(B_0)) = \left\|\rho(B_0)^{-1}\right\|_2 \le b_{\max}\left\|B_0^{-1}\right\|_2, \tag{50}$$

$$\|A_0\|_2 \le a_{\max}\|\rho(A_0)\|_2, \quad \|B_0\|_2 \le b_{\max}\|\rho(B_0)\|_2, \tag{51}$$

$$\|\rho(A_0)\|_2 \le \|A_0\|_2/a_{\min}, \quad \text{and} \quad \|\rho(B_0)\|_2 \le \|B_0\|_2/b_{\min}. \tag{52}$$

Proof of Theorem 3.1. Theorem 3.1 is a simple corollary of Theorem 11.1. We now insert the bounds as in (51) in Theorem 11.1 to obtain $\left\|\widehat{A \otimes B} - A_0 \otimes B_0\right\|_2 \le \|A_0\|_2\|B_0\|_2\,\delta$ where

$$\begin{aligned}
\delta = {} & (\lambda_{A_0} \wedge \lambda_{B_0})/2 + \\
& \frac{C r_a \kappa(\rho(A_0))}{\varphi_{\min}(\rho(A_0))}\lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1} + \frac{C' r_b \kappa(\rho(B_0))}{\varphi_{\min}(\rho(B_0))}\lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1} \\
+ \; & 2CC' r_a r_b \lambda_{A_0}\lambda_{B_0}\frac{\kappa(\rho(A_0))\,\kappa(\rho(B_0))}{\varphi_{\min}(\rho(A_0))\varphi_{\min}(\rho(B_0))}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1} \\
\asymp \; & \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2} + \log^{1/2}(m \vee f)\left(\sqrt{\frac{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}{nf}} + \sqrt{\frac{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1}{nm}}\right) + o(1).
\end{aligned}$$

For the inverse, we plug in the bounds as in (49) and (50) in Theorem 11.1 to obtain $\left\|\widehat{A \otimes B}^{-1} - A_0^{-1} \otimes B_0^{-1}\right\|_2 \le \left\|B_0^{-1}\right\|_2\left\|A_0^{-1}\right\|_2\delta'$ where

$$\begin{aligned}
\delta' \; = \; & (\lambda_{A_0} \wedge \lambda_{B_0})/3 + \frac{C r_a \lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}}{\varphi_{\min}(\rho(A_0))} + \frac{C' r_b \lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1}}{\varphi_{\min}(\rho(B_0))} \\
+ \; & \frac{3CC' r_a r_b \lambda_{A_0}\lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee 1}}{2\varphi_{\min}(\rho(A_0))\varphi_{\min}(\rho(B_0))}.
\end{aligned}$$

Hence

$$\delta' \asymp \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{3} + \log^{1/2}(m \vee f)\left(\sqrt{\frac{|A_0^{-1}|_{0,\text{off}} \vee 1}{nf}} + \sqrt{\frac{|B_0^{-1}|_{0,\text{off}} \vee 1}{nm}}\right) + o(1).$$

□

We need the following bounds in the proof of Theorem 3.2:

$$(53) \qquad (a_{\min} \vee \varphi_{\min}(A_0))\sqrt{m} \le \|A_0\|_F = \left(\sum_{i=1}^{m} \varphi_i(A_0)^2\right)^{1/2} \le \sqrt{m}\,\|A_0\|_2\,,$$

$$(54) \qquad (b_{\min} \vee \varphi_{\min}(B_0))\sqrt{f} \le \|B_0\|_F = \left(\sum_{i=1}^{f} \varphi_i(B_0)^2\right)^{1/2} \le \sqrt{m}\,\|B_0\|_2\,,$$

$$(55) \qquad \sqrt{m}/a_{\max} = \left(\frac{1}{a_{\max}} \vee \frac{1}{\varphi_{\max}(A_0)}\right)\sqrt{m} \le \left\|A_0^{-1}\right\|_F \le \sqrt{m}\left\|A_0^{-1}\right\|_2,$$

$$(56) \qquad \sqrt{f}/b_{\max} = \left(\frac{1}{b_{\max}} \vee \frac{1}{\varphi_{\max}(B_0)}\right)\sqrt{f} \le \left\|B_0^{-1}\right\|_F \le \sqrt{m}\left\|B_0^{-1}\right\|_2.$$

Proof of Theorem 3.2. We first plug in the lower bounds as in (53) and (54) in Theorem 11.2 to obtain under (A1) and (A2), $\left\|\widehat{A \otimes B} - A_0 \otimes B_0\right\|_F \le \|A_0\|_F \|B_0\|_F \,\delta$ where $\delta = (\lambda_{A_0} \wedge \lambda_{B_0})/2+$

$$\begin{aligned} & C r_a \kappa\,(\rho(A_0))^2\,\lambda_{B_0}\sqrt{\frac{|A_0^{-1}|_{0,\text{off}} \vee m}{m}} + C' r_b \kappa\,(\rho(B_0))^2\,\lambda_{A_0}\sqrt{\frac{|B_0^{-1}|_{0,\text{off}} \vee f}{f}} \\ + \quad & 2CC' r_a r_b \lambda_{A_0}\lambda_{B_0}\kappa\,(\rho(A_0))^2\,\kappa\,(\rho(B_0))^2\sqrt{\frac{|A_0^{-1}|_{0,\text{off}} \vee m}{m}}\sqrt{\frac{|B_0^{-1}|_{0,\text{off}} \vee f}{f}} \\ = \quad & O\left(\lambda_{B_0}\sqrt{\frac{|A_0^{-1}|_{0,\text{off}} \vee m}{m}} + \lambda_{A_0}\sqrt{\frac{|B_0^{-1}|_{0,\text{off}} \vee f}{f}}\right) \to 0. \end{aligned}$$

To see the last equality: under (A1)

1. If $\left|A_0^{-1}\right|_{0,\text{off}} = O(m)$, then $\lambda_{B_0}\sqrt{(\left|A_0^{-1}\right|_{0,\text{off}} \vee m)/m} = O(\lambda_{B_0})$; otherwise, $\lambda_{B_0}\sqrt{(\left|A_0^{-1}\right|_{0,\text{off}} \vee m)/m} = \lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}}/m} = o(1/\sqrt{m})$;
2. If $\left|B_0^{-1}\right|_{0,\text{off}} = O(f)$, then $\lambda_{A_0}\sqrt{(\left|B_0^{-1}\right|_{0,\text{off}} \vee f)/f} = O(\lambda_{A_0})$; otherwise, $\lambda_{A_0}\sqrt{(\left|B_0^{-1}\right|_{0,\text{off}} \vee f)/f} = O(\lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}/f}) = o(1/\sqrt{f})$.

Moreover, the expression for δ can be simplified as follows:

1. If $\left|A_0^{-1}\right|_{0,\text{off}} = O(m)$ and $\left|B_0^{-1}\right|_{0,\text{off}} = O(f)$, then $\delta = O(\lambda_{B_0} + \lambda_{A_0})$.
2. If $1 \le n \le \log(m \vee f)$, then $(\lambda_{A_0} + \lambda_{B_0})$ dominates $(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{f}})$. Thus

$$\delta = O\left(\frac{1}{\sqrt{f}} + \frac{1}{\sqrt{m}} + \lambda_{A_0} + \lambda_{B_0}\right) = O(\lambda_{A_0} + \lambda_{B_0}).$$

We now insert the lower bounds as in (55) and (56) in Theorem 11.2 to obtain $\left\|\widehat{A \otimes B}^{-1} - A_0^{-1} \otimes B_0^{-1}\right\|_F \le \left\|A_0^{-1}\right\|_F \left\|B_0^{-1}\right\|_F \delta'$ where

$$\begin{aligned}\delta' \quad = \quad & \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{3} + \frac{C\lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}}{a_{\min}\left\|A_0^{-1}\right\|_F \varphi^2_{\min}(\rho(A_0))} + \frac{C'\lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f}}{b_{\min}\left\|B_0^{-1}\right\|_F \varphi^2_{\min}(\rho(B_0))} + \\ & + \frac{7CC'\lambda_{A_0}\lambda_{B_0}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} \vee f}}{5a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(A_0))\left\|A_0^{-1}\right\|_F\left\|B_0^{-1}\right\|_F}.\end{aligned}$$

Thus

$$\begin{aligned}\delta' \quad \le \quad & \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{3} + \frac{Cr_a\lambda_{B_0}}{\varphi^2_{\min}(\rho(A_0))}\sqrt{\frac{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}{m}} + \frac{C'r_b\lambda_{A_0}}{\varphi^2_{\min}(\rho(B_0))}\sqrt{\frac{\left|B_0^{-1}\right|_{0,\text{off}} \vee f}{f}} \\ & + \frac{7CC'r_ar_b\lambda_{A_0}\lambda_{B_0}}{5\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))}\sqrt{\frac{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}{m}}\sqrt{\frac{\left|B_0^{-1}\right|_{0,\text{off}} \vee f}{f}}.\end{aligned}$$

The upper bounds on δ' are dominated by that of δ under each case; hence the same conclusions hold. □

12. Proofs of Theorem 4.1 and Lemma 4.3. In order to prove Lemma 4.3 and Theorem 13.1, we need the following form of the Hanson-Wright inequality as recently derived in [4], which improves Theorem B.1 in a technical report version of the present work [5].

THEOREM 12.1. *Let $X = (X_1, \ldots, X_m) \in \mathbf{R}^m$ be a random vector with independent components X_i which satisfy $\mathbb{E}[X_i] = 0$ and $\|X_i\|_{\psi_2} \le \widetilde{K}$. Let A be an $m \times m$ matrix. Then, for every $t > 0$,*

$$\mathbb{P}\left(\left|X^T AX - \mathbb{E}\left[X^T AX\right]\right| > t\right) \le 2\exp\left[-c\min\left(\frac{t^2}{\widetilde{K}^4\|A\|_F^2}, \frac{t}{\widetilde{K}^2\|A\|_2}\right)\right].$$

By constructing a new matrix A_n which is block diagonal with n identical submatrices A/n along its diagonal, we prove the following Corollary 12.2.

COROLLARY 12.2. *Let A, X, and $\widetilde{K}$ be defined as in Theorem 12.1, and let $Y_1, \ldots, Y_n$ be independent copies of $X^T AX$. Then for every $t > 0$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i]) > t\right) \leq 2\exp\left[-c\min\left(\frac{nt^2}{\widetilde{K}^4\|A\|_F^2}, \frac{nt}{\widetilde{K}^2\|A\|_2}\right)\right].$$

Proof of Theorem 4.1. Throughout this proof, we assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds. First we obtain the large deviation bounds on the estimated correlation coefficients. We prove it only for $\widehat{\Gamma}(B_0)$, as a similar argument also works for $\widehat{\Gamma}(A_0)$. For all i, j, and $\alpha_n := \frac{\|A_0\|_F \tau_0}{\mathrm{tr}(A_0)}$,

$$\begin{aligned}
\left|\widehat{\Gamma}_{ij}(B_0) - \rho_{ij}(B_0)\right| &:= \left|\frac{\sum_{t=1}^n \langle y(t)^i, y(t)^j \rangle}{\sqrt{\sum_{t=1}^n \|y(t)^i\|_2^2}\sqrt{\sum_{t=1}^n \|y(t)^j\|_2^2}} - \rho_{ij}(B_0)\right| \\
&= \left|\frac{\frac{1}{n}\sum_{t=1}^n \langle y(t)^i, y(t)^j \rangle / (\mathrm{tr}(A_0)\sqrt{b_{ii}b_{jj}})}{(\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|y(t)^i\right\|_2^2/(b_{ii}\mathrm{tr}(A_0))})(\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|y(t)^j\right\|_2^2/(b_{jj}\mathrm{tr}(A_0))})} - \rho_{ij}(B_0)\right| \\
&= \left|\frac{\frac{1}{n}\sum_{t=1}^n \langle y(t)^i, y(t)^j \rangle / (\mathrm{tr}(A_0)\sqrt{b_{ii}b_{jj}}) - \rho_{ij}(B_0)}{(\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|y(t)^i\right\|_2^2/(b_{ii}\mathrm{tr}(A_0))})(\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|y(t)^j\right\|_2^2/(b_{jj}\mathrm{tr}(A_0))})}\right| \\
&+ \left|\frac{\rho_{ij}(B_0)}{(\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|y(t)^i\right\|_2^2/(b_{ii}\mathrm{tr}(A_0))})(\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|y(t)^j\right\|_2^2/(b_{jj}\mathrm{tr}(A_0))})} - \rho_{ij}(B_0)\right| \\
&\leq \frac{\alpha_n}{1-\alpha_n} + |\rho_{ij}(B_0)|\left|\frac{1}{1-\alpha_n} - 1\right|
\end{aligned}$$

where by Lemma 4.3, we have for all i, $\frac{\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|y(t)^i\right\|_2^2}}{\sqrt{b_{ii}\mathrm{tr}(A_0)}} \geq \sqrt{1-\alpha_n}$ and

$$\forall i, j, \qquad \left|\frac{\frac{1}{n}\sum_{t=1}^n \langle y(t)^i, y(t)^j \rangle}{\mathrm{tr}(A_0)\sqrt{b_{ii}b_{jj}}} - \rho_{ij}(B_0)\right| \leq \alpha_n.$$

The last inequality follows immediately by summing over the large deviation inequalities on the ℓ_2^2 norms for the row vectors and column vectors,

$$\begin{aligned}
\left|\tfrac{1}{n}\sum_{t=1}^n \|X(t)\|_F^2 - \mathrm{tr}(A_0)\mathrm{tr}(B_0)\right| &= \left|\tfrac{1}{n}\sum_{t=1}^n\sum_{i=1}^m \|x(t)\|_2^2 - \mathrm{tr}(A_0)\mathrm{tr}(B_0)\right| \\
&= \left|\tfrac{1}{n}\sum_{t=1}^n\sum_{j=1}^f \|y(t)\|_2^2 - \mathrm{tr}(A_0)\mathrm{tr}(B_0)\right| \\
&\leq \mathrm{tr}(A_0)\mathrm{tr}(B_0)\min\left\{\frac{\tau_0\|B_0\|_F}{\mathrm{tr}(B_0)}, \frac{\tau_0\|A_0\|_F}{\mathrm{tr}(A_0)}\right\} = (\alpha_n \wedge \beta_n)\mathrm{tr}(A_0)\mathrm{tr}(B_0).
\end{aligned}$$

□

Proof of Lemma 4.3. The conclusions of Lemma 4.3 follow from Theorem 13.1 by setting $M = I$, and $N = I$, and thus $D = \|A_0\|_F$ and $D' = \|B_0\|_F$, given that the stable rank for $A = A_0$ and $B = B_0$ satisfies $r(A_0), r(B_0) \geq 4 \log(m \vee f)/n$ under (A2). □

Proof of Corollary 12.2. Let $X_1, \ldots, X_n$ be independent copies of X. Let $X_{(n)}$ represent a long vector in $\mathbf{R}^{mn}$ such that $X_{(n)}^T := X_1^T|X_2^T|...|X_n^T$ is concatenated from individuals vectors $X_1, X_2, ..., X_n \in \mathbf{R}^m$. Let us construct a new block-diagonal matrix A_n of size $mn \times mn$ with n copies of the matrix A/n on the main diagonal and 0 outside. Clearly

$$\|A_n\|_F^2 = \|A\|_F^2/n \quad \text{and} \quad \|A_n\|_2 = \|A\|_2/n.$$

We can now write $\frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n X_i A X_i = X_{(n)}^T A_n X_{(n)}$, and apply Theorem 12.1 to obtain:

$$\begin{aligned}
\mathbb{P}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]) > t\right) &= \mathbb{P}\left(\left|X_{(n)}^T A_n X_{(n)} - \mathbb{E}\left[X_{(n)}^T A_n X_{(n)}\right]\right| > t\right) \\
&\leq 2\exp\left[-c\min\left(\frac{t^2}{\widetilde{K}^4 \|A_n\|_F^2}, \frac{t}{\widetilde{K}^2 \|A_n\|_2}\right)\right] \\
&= 2\exp\left[-c\min\left(\frac{nt^2}{\widetilde{K}^4 \|A\|_F^2}, \frac{nt}{\widetilde{K}^2 \|A\|_2}\right)\right].
\end{aligned}$$

The corollary is thus proved. □

13. A general theory on concentration inequalities. We now provide a large deviation inequality for the weighted sum of submatrices $\widehat{S}_n^{\ell k}, \ell, k = 1, \ldots, m$ (or sum of $\widetilde{S}_n^{\ell k}, \ell, k = 1, \ldots, f$).

THEOREM 13.1. *Let $m \vee f \geq 2$. Let matrices $M_{m\times m}$ and $N_{f\times f}$ satisfy the following condition: $\frac{1}{m}\|M\|_F^2 < \infty$ and $\frac{1}{f}\|N\|_F^2 < \infty$. Let K be as in (17) and $\tau_0 = 2CK^2 \log^{1/2}(m \vee f)/\sqrt{n}$ as defined in (19), where*

$$C = \frac{1}{\sqrt{c}} \vee \frac{1}{c} \vee 1 \text{ for } c \text{ as in Corollary 12.2}, \quad \text{so that} \quad C^2 c, Cc \geq 1, \tag{57}$$

Suppose that the stable ranks for $A = A_0^{1/2} M A_0^{1/2}$ and $B = B_0^{1/2} N B_0^{1/2}$, *satisfy* $r(A), r(B) \geq 4\log(m \vee f)/n$. *Then, with probability* $1 - \frac{3}{(m \vee f)^2}$,

$$\left\| \mathrm{diag}(B_0)^{-1/2} \Big(\sum_{k=1}^{m} \sum_{\ell=1}^{m} M_{k\ell} \widehat{S}_n^{\ell k} \Big) \mathrm{diag}(B_0)^{-1/2} - \langle A_0, M \rangle \, \rho(B_0) \right\|_{\max} \leq D\tau_0$$

$$\left\| \mathrm{diag}(A_0)^{-1/2} \Big(\sum_{k=1}^{f} \sum_{\ell=1}^{f} N_{k\ell} \widetilde{S}_n^{\ell k} \Big) \mathrm{diag}(A_0)^{-1/2} - \langle B_0, N \rangle \, \rho(A_0) \right\|_{\max} \leq D'\tau_0$$

where $D = \left\| A_0^{1/2} M A_0^{1/2} \right\|_F$ *and* $D' = \left\| B_0^{1/2} N B_0^{1/2} \right\|_F$. *Otherwise, suppose that* $m \vee f = o(\exp(m \wedge f))$. *Then with probability* $1 - \frac{3}{(m \vee f)^2}$, *the above inequalities hold with* $D = 2\sqrt{m} \left\| A_0 \right\|_2 \left\| M \right\|_2$ *and* $D' = 2\sqrt{f} \left\| B_0 \right\|_2 \left\| N \right\|_2$.

Remark 13.2. *Theorem 13.1 is stated as a matrix max norm bound for convenience; when we look at the individual entry-wise bound, we will come to the same conclusions as in Lemma 4.3. Clearly, Theorem 4.1 is a special case of Theorem 13.1 when we set* $M = I$.

We need to state the following Proposition 13.3 before we proceed.

Proposition 13.3. *Let* $\begin{pmatrix} c_{ii} & c_{ij} \\ c_{ij} & c_{jj} \end{pmatrix}$ *be the unique symmetric square root of the positive definite matrix* $B_{0,\{i,j\}}$, *where* $B_{0,\{i,j\}}$ *is the submatrix of* B_0 *with rows and columns indexed by* $\{i, j\}$. *Let*

$$Z(i,j) := \sum_{k=1}^{2m} \sum_{\ell=1}^{2m} \frac{M'_{k\ell}(i,j)}{\sqrt{b_{ii} b_{jj}}} g_k g_\ell - \frac{b_{ij}}{\sqrt{b_{ii} b_{jj}}} \langle A_0, M \rangle \tag{58}$$

where $g_1, \ldots, g_{2m}$ *are independent standard Gaussian random variables, and*

$$M'(i,j) = \begin{pmatrix} c_{ii} c_{ij} & c_{ii} c_{jj} \\ c_{ij} c_{ij} & c_{ij} c_{jj} \end{pmatrix} \otimes A_0^{1/2} M A_0^{1/2}$$

Then $\mathbb{E}\left[Z(i,j) \right] = 0$, $\left\| M'(i,j) \right\|_F = \sqrt{b_{ii} b_{jj}} \left\| A_0^{1/2} M A_0^{1/2} \right\|_F$,

$$\left\| M'(ij) \right\|_2 \leq \sqrt{b_{ii} b_{jj}} \left\| A_0^{1/2} M A_0^{1/2} \right\|_2.$$

Proof of Theorem 13.1. Our analysis will focus on estimating B_0 and $\rho(B_0)$ which applies to A_0 and $\rho(A_0)$ with minor changes on notation.

Decorrelation. The quantity which we are interested in is the following weighted sum of random matrices:

$$\sum_{k=1}^{m}\sum_{\ell=1}^{m}\widehat{S}_n^{\ell k}M_{k\ell} - a_{\ell k}M_{k\ell}B_0 = \\ \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{m}\sum_{\ell=1}^{m}M_{k\ell}x(t)^{\ell}\otimes x(t)^{k} - \langle A_0, M\rangle B_0.$$

Let us focus on a particular entry b_{ij} in matrix B_0 and rewrite the sum

$$\sum_{k=1}^{m}\sum_{\ell=1}^{m}\widehat{S}_{ij}^{\ell k}M_{k\ell} - a_{\ell k}M_{k\ell}b_{ij} = \\ \frac{1}{n}\sum_{t=1}^{n}\Big(\sum_{k=1}^{m}\sum_{\ell=1}^{m}M_{k\ell}x(t)_i^{\ell}\times x(t)_j^{k} - \langle A_0, M\rangle b_{ij}\Big)$$

where each summand inside the round brackets is a Gaussian chaos of order 2. We explore the concentration of the sum on the RHS in the next subsection. We first need to decorrelate the vectors in the sum. First observe that when $i = j$, the Gaussian random vector $(x(1)_j^k)_{k=1}^m$ involved in the sum is of size m, with covariance being $b_{jj}A_0$. Without loss of generality, we write $(x(1)_j^k)_{k=1}^m = (b_{jj}A_0)^{1/2}(g_1,\ldots,g_m)^T$, where $g_1,\ldots,g_m$ i.i.d. $\sim N(0,1)$ and replace the sum with the following expression:

$$\begin{aligned}\sum_{k=1}^{m}\sum_{\ell=1}^{m}M_{k\ell}x(1)_j^{\ell}\times x(1)_j^{k} &= (g_1,\ldots,g_m)b_{jj}A_0^{1/2}MA_0^{1/2}(g_1,\ldots,g_m)^T \\ &= \sum_{k=1}^{m}\sum_{\ell=1}^{m}b_{jj}M'_{k\ell}g_k g_\ell \quad \text{where} \quad M' = A_0^{1/2}MA_0^{1/2}.\end{aligned}$$

We next fix $i = 1, j = 2$. Then for vectors $\big(x(1)_1^{\ell}\big)_{\ell=1}^m$ and $\big(x(1)_2^{k}\big)_{k=1}^m$, we concatenate them to form a Gaussian random vector of size $2m$ with covariance $B_{0,\{1,2\}}\otimes A_0 = \begin{pmatrix} b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\otimes A_0$ where $B_{0,\{1,2\}}$ is the submatrix of B_0 with rows and columns indexed by $\{1,2\}$. We write

$$\begin{aligned}\Big((x(1)_1^{\ell})_{\ell=1}^m, (x(1)_2^{k})_{k=1}^m\Big) &:= B_{0,\{1,2\}}^{1/2}\otimes A_0^{1/2}(g_1,\ldots,g_{2m})^T \\ &= \begin{pmatrix} c_{11}A_0^{1/2} & c_{12}A_0^{1/2}\\ c_{12}A_0^{1/2} & c_{22}A_0^{1/2}\end{pmatrix}(g_1,\ldots,g_{2m})^T\end{aligned}$$

where $g_1, \ldots, g_{2m}$ i.i.d. $\sim N(0,1)$. Thus

$$\begin{aligned}
\sum_{k=1}^{m}\sum_{\ell=1}^{m} M_{k\ell} x(1)_1^{\ell} \times x(1)_2^{k} &= (x(1)_1^1 \ldots, x(1)_1^m) M (x(1)_2^1, \ldots, x(1)_2^m)^T \\
&= \sum_{i=1}^{2m}\sum_{j=1}^{2m} M'_{ij}(1,2) g_i g_j \\
\text{where } \; M'(1,2) &= \begin{pmatrix} c_{11}c_{12} & c_{11}c_{22} \\ c_{12}c_{12} & c_{12}c_{22} \end{pmatrix} \otimes A_0^{1/2} M A_0^{1/2}
\end{aligned}$$

$$\text{and } \mathbb{E}\left[\sum_{i=1}^{2m}\sum_{j=1}^{2m} M'_{ij}(1,2) g_i g_j\right] = \mathrm{tr}(M'(1,2)) = b_{12}\mathrm{tr}(A_0 M).$$

Applying the union bound. Let $g_k, k = 1, 2, \ldots,$ be independent standard Gaussian random variables. For $j = 1, \ldots, f$, and

$$\sum_{k=1}^{m}\sum_{\ell=1}^{m} \widehat{S}_{jj}^{k\ell} M_{k\ell}/b_{jj} - \langle A_0, M \rangle = \frac{1}{n}\sum_{t=1}^{n} Z_t(j,j)$$

where $Z_1(j,j), \ldots, Z_n(j,j)$ are independent copies of random variable

$$Z_A := \sum_{k=1}^{m}\sum_{\ell=1}^{m} M'_{k\ell} g_k g_\ell - \langle A_0, M \rangle \quad \text{where} \quad M' = A_0^{1/2} M A_0^{1/2}.$$

By Corollary 12.2, we have for $A = M' = A_0^{1/2} M A_0^{1/2}$ and $t > 0$,

$$\mathbb{P}\left(\exists j : \left|\sum_{k=1}^{m}\sum_{\ell=1}^{m} \widehat{S}_{jj}^{k\ell} M_{k\ell}/b_{jj} - \langle A_0, M \rangle\right| \geq t\right) \leq \mathbb{P}\left(\exists j : \left|\frac{1}{n}\sum_{t=1}^{n} Z_t(j,j)\right| \geq t\right)$$

$$\leq 2f \exp\left[-c \min\left(\frac{nt^2}{K^4 \left\|A_0^{1/2} M A_0^{1/2}\right\|_F^2}, \frac{nt}{K^2 \left\|A_0^{1/2} M A_0^{1/2}\right\|_2}\right)\right] \tag{59}$$

$$\leq 2f \exp\left\{-c \min\left(\frac{nt^2}{K^4 \|A_0\|_2^2 \|M\|_F^2}, \frac{nt}{K^2 \|A_0\|_2 \|M\|_2}\right)\right\}. \tag{60}$$

Let $g(t)_k, k = 1, 2, \ldots, m, t = 1, \ldots, n$ be independent standard Gaussian random variables. Let $Z_1(i,j), \ldots, Z_n(i,j)$ be independent copies of random variable $Z(i,j)$ defined as in (58). Following the Decorrelation step, we write

$$\begin{aligned}
&\sum_{k=1}^{m}\sum_{\ell=1}^{m} \widehat{S}_{ij}^{k\ell} M_{k\ell}/\sqrt{b_{ii}b_{jj}} - \langle A_0, M \rangle \rho_{ij}(B_0) \\
&= \sum_{k=1}^{m}\sum_{\ell=1}^{m} \frac{1}{n}\sum_{t=1}^{n} \frac{M_{k\ell} x(t)_i^{\ell} \times x(t)_j^{k}}{\sqrt{b_{ii}b_{jj}}} - \langle A_0, M \rangle \rho_{ij}(B_0) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{k=1}^{2m}\sum_{\ell=1}^{2m} \frac{M'_{k\ell}(i,j)}{\sqrt{b_{ii}b_{jj}}} g(t)_k g(t)_\ell - \langle A_0, M \rangle \rho_{ij}(B_0)\right) =: \frac{1}{n}\sum_{t=1}^{n} Z_t(i,j).
\end{aligned}$$

We now apply Corollary 12.2 with $A(i,j) = M'(i,j)/\sqrt{b_{ii}b_{jj}}$, and some $t > 0$ to be specified while summing over all events for $i \neq j$,

$$\mathbb{P}\left(\exists i \neq j : \frac{1}{n}\left|\sum_{t=1}^{n} Z_t(i,j)\right| \geq t\right) \leq \sum_{i\neq j} \exp\left[-c\min\left(\frac{nt^2}{K^4\|A(i,j)\|_F^2}, \frac{nt}{K^2\|A(i,j)\|_2}\right)\right]$$

$$\leq f(1-f)\exp\left[-c\min\left(\frac{nt^2}{K^4\left\|A_0^{1/2}MA_0^{1/2}\right\|_F^2}, \frac{nt}{K^2\left\|A_0^{1/2}MA_0^{1/2}\right\|_2}\right)\right] \tag{61}$$

where $\|A(i,j)\|_F = \|M'(i,j)\|_F/\sqrt{b_{ii}b_{jj}} = \left\|A_0^{1/2}MA_0^{1/2}\right\|_F$ and $\|A(i,j)\|_2 = \|M'(i,j)\|_2/\sqrt{b_{ii}b_{jj}} \leq \left\|A_0^{1/2}MA_0^{1/2}\right\|_2$ by Proposition 13.3.

Case 1. When the stable rank satisfies $r(A_0^{1/2}MA_0^{1/2}) \geq 4\log(m\vee f)/n$, set $t = \left\|A_0^{1/2}MA_0^{1/2}\right\|_F \tau_0$. Then by (59) and (61) we have for $A = A_0^{1/2}MA_0^{1/2}$,

$$\begin{aligned}
&\mathbb{P}\left(\exists i,j : \left|\frac{1}{n}\sum_{t=1}^{n} Z_t(j,j)\right| \geq 2CK^2\|A\|_F \log^{1/2}(m\vee f)/\sqrt{n}\right) \leq \\
&\quad f(f+1)\exp\left[-c\min\left(\frac{4C^2K^4\|A\|_F^2\log(m\vee f)}{K^4\|A\|_F^2}, \frac{2\sqrt{n}CK^2\|A\|_F\log^{1/2}(m\vee f)}{K^2\|A\|_2}\right)\right] \\
&= f(f+1)\exp\left[-\min\left(4\log(m\vee f), \frac{\sqrt{n}2\|A\|_F\log^{1/2}(m\vee f)}{\|A\|_2}\right)\right] \leq \frac{f(f+1)}{(m\vee f)^4}.
\end{aligned}$$

Case 2. Otherwise, set

$$t = 2\|A_0\|_2\|M\|_2\sqrt{m}\tau_0 = 4CK^2\|A_0\|_2\|M\|_2\sqrt{m}\log^{1/2}(m\vee f)/\sqrt{n}.$$

By (60) and (61), we have for $m \geq \log(m\vee f)$, for which a sufficient condition is to assume that $m \vee f = o(e^m)$,

$$\begin{aligned}
&\mathbb{P}\left(\exists i,j: \quad \left|\frac{1}{n}\sum_{t=1}^{n} Z_t(j,j)\right| \geq 4CK^2 \left\|A_0\right\|_2 \left\|M\right\|_2 \sqrt{m}\log^{1/2}(m \vee f)/\sqrt{n}\right) \leq \\
&= \quad f(f+1)\exp\left\{-c\min\left(\frac{16C^2 \left\|M\right\|_2^2 m\log(m \vee f)}{\left\|M\right\|_F^2}, 4C\sqrt{nm}\log^{1/2}(m \vee f)\right)\right\} \\
&\leq \quad f(f+1)\exp\left\{-\min\left(16\log(m \vee f), 4\sqrt{nm}\log^{1/2}(m \vee f)\right)\right\} \\
&\leq \quad f(f+1)\exp\{-4\log(m \vee f)\} \leq \frac{f(f+1)}{(m \vee f)^4}.
\end{aligned}$$

Assuming $m \vee f = o(\exp(f \wedge m))$, we can show the same type of large deviation bound holds for the union of events concerning the weighted sum of submatrices $\widetilde{S}_n^{\ell k}, \ell, k = 1, \ldots, f$ where the weights correspond to entries in an $f \times f$ matrix N:

$$\begin{aligned}
&\mathbb{P}\left(\left\|\mathrm{diag}(A_0)^{-1/2}\Big(\sum_{k=1}^{f}\sum_{\ell=1}^{f} N_{k\ell}\widetilde{S}_n^{\ell k}\Big)\mathrm{diag}(A_0)^{-1/2} - \mathrm{tr}(B_0 N)\rho(A_0)\right\|_{\max} \geq D'\tau_0\right) \\
&\leq m(m+1)/(m \vee f)^4.
\end{aligned}$$

The theorem is thus proved by applying the union bound to sum up the probability for all bad events and using the fact that for $(m \vee f) \geq 2$,

$$\frac{m(m+1)+f(f+1)}{(m \vee f)^4} \leq \frac{m^2+f^2+m+f}{(m \vee f)^4} \leq \frac{3}{(m \vee f)^2}.$$

□

Proof of Proposition 13.3. The proof for $\mathbb{E}\left[Z(i,j)\right] = 0$ follows exactly the arguments in the Decorrelation step in the proof of Theorem 13.1. Note that for the symmetric positive definite matrix $B_{0,\{i,j\}}^{1/2}$, we have

$$\left\|B_{0,\{i,j\}}^{1/2}\right\|_F = \sqrt{\mathrm{tr}(B_{0,\{i,j\}}^{1/2}B_{0,\{i,j\}}^{1/2})} = \sqrt{\mathrm{tr}(B_{0,\{i,j\}})} = \sqrt{b_{ii}b_{jj}},$$

and for $A = A_0^{1/2} M A_0^{1/2}$,

$$\left\|M'(i,j)\right\|_F = \left\|B_{0,\{i,j\}}^{1/2}\right\|_F \left\|A\right\|_F = \sqrt{b_{ii}b_{jj}}\left\|A\right\|_F$$

Moreover,

$$\left\|\frac{M'(ij)}{\sqrt{b_{ii}b_{jj}}}\right\|_2 = \frac{\left\|B_{0,\{i,j\}}^{1/2}\right\|_2 \left\|A\right\|_2}{\sqrt{b_{ii}b_{jj}}} \leq \frac{\left\|B_{0,\{i,j\}}^{1/2}\right\|_F \left\|A\right\|_2}{\sqrt{b_{ii}b_{jj}}} = \left\|A\right\|_2.$$

□

14. Proof of Corollary 14.1. Let $\Theta_0 := \rho(A_0)^{-1} \succ 0$ and $\Phi_0 := \rho(B_0)^{-1} \succ 0$. Let $\Theta_0 = (\theta_{ij})$ and $\Phi_0 = (\phi_{ij})$. Then,

$$0 < \varphi_{\min}(\rho(A_0)) \le 1 \le \varphi_{\max}(\rho(A_0)) < +\infty \text{ and hence } \kappa(\rho(A_0)) \ge 1, \tag{62}$$
$$0 < \varphi_{\min}(\rho(B_0)) \le 1 \le \varphi_{\max}(\rho(B_0)) < +\infty \text{ and hence } \kappa(\rho(B_0)) \ge 1, \tag{63}$$

given that $\sum_{i=1}^{m} \varphi_i(\rho(A_0)) = \mathrm{tr}(\rho(A_0)) = m$ and $\sum_{i=1}^{f} \varphi_i(\rho(B_0)) = f$.

We first write our estimator $\widehat{\Theta}$ for $\rho(A_0)^{-1}$ and $\widehat{\Phi}$ for $\rho(B_0)^{-1}$ as follows:

$$\widehat{\Theta} = \arg\min_{\Theta \succ 0} \left\{ \mathrm{tr}(\Theta \widehat{\Gamma}(A_0)) - \log|\Theta| + \lambda_B |\Theta|_{1,\mathrm{off}} \right\}, \tag{64a}$$
$$\widehat{\Phi} = \arg\min_{\Phi \succ 0} \left\{ \mathrm{tr}(\Phi \widehat{\Gamma}(B_0)) - \log|\Phi| + \lambda_A |\Phi|_{1,\mathrm{off}} \right\}, \tag{64b}$$

where λ_A, λ_B are non-negative regularization parameters, and $\widehat{\Gamma}(A_0)$ and $\widehat{\Gamma}(B_0)$ are sample correlation matrices. A unique minimizer exists for each objective function above; see [2]. Then $\widehat{A}_\rho := \widehat{\Theta}^{-1}$ and $\widehat{B}_\rho := \widehat{\Phi}^{-1}$, where $\widehat{A}_\rho$ and $\widehat{B}_\rho$ are the unique minimizers as defined in Theorem 4.5. Let us define the following shorthand notation:

$$S_{A_0} = \{(i,j) : \theta_{ij} \neq 0,\ i \neq j\} \quad \text{and} \quad S_{B_0} = \{(i,j) : \phi_{ij} \neq 0,\ i \neq j\}.$$

Let $\Delta_{A_0} := \widehat{\Theta} - \Theta_0$ and $\Delta_{B_0} := \widehat{\Phi} - \Phi_0$.

We use Corollary 14.1 in our analysis of the flip-flip algorithm.

Corollary 14.1. *Suppose conditions in Theorem 4.5 hold, and λ_B and λ_A are chosen as in (26). Let $S = \{(i,j) :\ \Theta_{0ij} \neq 0,\ i \neq j\}$ and $S^c = \{(i,j) :\ \Theta_{0ij} = 0,\ i \neq j\}$. Then on $\mathcal{T}(A_0)$ as defined in Theorem 4.5,*

$$\begin{aligned} |\Delta_{S^c}|_1 &\le \frac{1+\varepsilon}{1-\varepsilon} |\Delta_S|_1 \text{ where } \Delta = \widehat{A}_\rho^{-1} - \Theta_0 =: \Delta_{A_0} \\ \text{and } |\Delta_{A_0}|_{1,\mathrm{off}} &\le \frac{1+\varepsilon}{1-\varepsilon} 9\lambda_B \sqrt{|A_0^{-1}|_{0,\mathrm{off}} \vee 1}\sqrt{|A_0^{-1}|_{0,\mathrm{off}}}/\varphi^2_{\min}(\rho(A_0)). \end{aligned} \tag{65}$$

Similarly, for $\Delta := \Delta_{B_0} = \widehat{B}_\rho^{-1} - \Phi_0$, $|\Delta_{S^c}|_1 \le \frac{1+\epsilon}{1-\epsilon} |\Delta_S|_1$, where $S := \{(i,j) : \Phi_{0ij} \neq 0,\ i \neq j\}$ and $S^c := \{(i,j) :\ \Phi_{0ij} = 0,\ i \neq j\}$, holds on $\mathcal{T}(B_0)$, and

$$|\Delta_{B_0}|_{1,\mathrm{off}} \le \frac{1+\epsilon}{1-\epsilon} 9\lambda_A \sqrt{|B_0^{-1}|_{0,\mathrm{off}} \vee 1}\sqrt{|B_0^{-1}|_{0,\mathrm{off}}}/\varphi^2_{\min}(\rho(B_0)).$$

We need the following lemma. For an index set S and a matrix $W = [w_{ij}]$, write $W_S \equiv (w_{ij} I((i,j) \in S))$, where $I(\cdot)$ is an indicator function. Inequality (66) appears in [3]. We include its proof for convenience.

Lemma 14.2. *Let $\Theta_0 \succ 0$. Let $S = \{(i,j) : \Theta_{0ij} \neq 0,\ i \neq j\}$ and $S^c = \{(i,j) : \Theta_{0ij} = 0,\ i \neq j\}$. Then for all $\Delta \in \mathbf{R}^{m \times m}$, we have*

$$|\Theta_0 + \Delta|_{1,\text{off}} - |\Theta_0|_{1,\text{off}} \geq |\Delta_{S^c}|_1 - |\Delta_S|_1 \tag{66}$$

Moreover, we have on event $\mathcal{T}(A_0)$, for all $\Delta \in \mathbf{R}^{m \times m}$,

$$\left|\text{tr}\big(\Delta(\widehat{\Gamma}(A_0) - \rho(A_0))\big)\right| \leq \delta_{n,f}\,|\Delta|_{1,\text{off}} = \delta_{n,f}\,(|\Delta_{S^c}|_1 + |\Delta_S|_1)\,. \tag{67}$$

Proposition 14.3 is a standard result.

Proposition 14.3. *Let B be a $p \times p$ matrix. If $B \succ 0$ and $B + D \succ 0$, then $B + vD \succ 0$ for all $v \in [0,1]$.*

Proof of Corollary 14.1. Let $\widehat{\Gamma}(A_0)$ be the sample correlation matrix. Let $\lambda := \lambda_B$. Let $\widehat{\Theta}$ be the optimal solution to (64a). Let $\Theta_0 := \rho(A_0)^{-1}$. Then $\Delta = \widehat{\Theta} - \Theta_0$ minimizes $G(\Delta)$ defined as follows:

$$\begin{aligned} G(\Delta) &:= \log|\Theta_0| - \log|\Theta_0 + \Delta| + \text{tr}\,(\Delta\rho(A_0)) + \text{tr}\left(\Delta(\widehat{\Gamma}(A_0) - \rho(A_0))\right) \\ &\quad + \lambda(|\Theta_0 + \Delta|_{1,\text{off}} - |\Theta_0|_{1,\text{off}}) \\ &= W + \text{tr}\left(\Delta(\widehat{\Gamma}(A) - \rho(A_0)\right) + \lambda_{n,f}\left(|\Theta_0 + \Delta|_{1,\text{off}} - |\Theta_0|_{1,\text{off}}\right) && (68) \\ &\geq W - \delta_{n,f}\,|\Delta_{S^c}|_1 - \delta_{n,f}\,|\Delta_S|_1 + \lambda_{n,f}\,|\Delta_{S^c}|_1 - \lambda_{n,f}\,|\Delta_S|_1 && (69) \\ &= W - (\delta_{n,f} + \lambda_{n,f})\,|\Delta_S|_1 + (\lambda_{n,f} - \delta_{n,f})\,|\Delta_{S^c}|_1 && (70) \end{aligned}$$

where $W = \text{vec}\{\Delta\}^T \left(\int_0^1 (1-\text{v})(\Theta_0 + \text{v}\Delta)^{-1} \otimes (\Theta_0 + \text{v}\Delta)^{-1} d\text{v}\right) \text{vec}\{\Delta\}$, and in (69), we have applied (66) and (67).

Clearly $G(\underline{0}) = 0$ and hence $G(\Delta) \leq G(\underline{0}) = 0$. Hence for $\lambda_{n,f} := \lambda_B \geq \delta_{n,f}/\varepsilon$, we have by (70)

$$\begin{aligned} -W + (\delta_{n,f} + \lambda_{n,f})\,|\Delta_S|_1 &\geq (\lambda_{n,f} - \delta_{n,f})\,|\Delta_{S^c}|_1 \geq (1-\varepsilon)\lambda_{n,f}\,|\Delta_{S^c}|_1 \\ \text{and hence } (1-\varepsilon)\lambda_{n,f}\,|\Delta_{S^c}|_1 &\leq -W + (\delta_{n,f} + \lambda_{n,f})\,|\Delta_S|_1 \\ &\leq -W + (1+\varepsilon)\lambda_{n,f}\,|\Delta_S|_1 \end{aligned}$$

Thus we have $|\Delta_{S^c}|_1 \leq \frac{1+\varepsilon}{1-\varepsilon}\,|\Delta_S|_1$ given that $W \geq 0$; To see this, we note that $\Theta_0 \succ 0$, and hence $\Theta_0 + v\Delta \succ 0$ for all $v \in [0,1]$ in view of Proposition 14.3, given that $\widehat{\Theta} = \Theta_0 + \Delta \succ 0$ as an optimal solution to (64a). To see the last inequality, we have

$$\begin{aligned} \left|\widehat{\Theta} - \Theta_0\right|_{1,\text{off}} = |\Delta_{A_0}|_{1,\text{off}} &\leq \frac{2}{1-\varepsilon}\,|\Delta_S|_1 \leq \frac{2}{1-\varepsilon}\sqrt{|A_0^{-1}|_{0,\text{off}}}\,\|\Delta_S\|_F \\ &\leq \frac{2}{1-\varepsilon}\sqrt{|A_0^{-1}|_{0,\text{off}}}\,\|\Delta_{A_0}\|_F\,, \end{aligned}$$

and it holds by plugging in the bound on $\|\Delta_{A_0}\|_F$. $\square$

Proof of Lemma 14.2. We write $\Theta_0 = \mathrm{diag}(\Theta_0) + \Theta_{0,S} + \Theta_{0,S^c}$ and observe that $\Theta_{0,S^c} = \underline{0}$ and hence $|\Theta_0|_{1,\mathrm{off}} = |\Theta_{0,S}|_1 = |\Theta_{0,S}|_{1,\mathrm{off}}$. Thus

$$\begin{aligned} |\Theta_0 + \Delta|_{1,\mathrm{off}} &= |\Theta_{0S} + \Delta_S|_{1,\mathrm{off}} + |\Delta_{S^c}|_{1,\mathrm{off}}, \quad \text{hence} \\ |\Theta_0 + \Delta|_{1,\mathrm{off}} - |\Theta_0|_{1,\mathrm{off}} &\geq |\Theta_{0S} + \Delta_S|_{1,\mathrm{off}} - |\Theta_{0,S}|_{1,\mathrm{off}} + |\Delta_{S^c}|_{1,\mathrm{off}} \qquad (71) \\ &\geq |\Delta_{S^c}|_{1,\mathrm{off}} - |\Delta_S|_{1,\mathrm{off}} = |\Delta_{S^c}|_1 - |\Delta_S|_1 \qquad (72) \end{aligned}$$

where (71) follows from the triangle inequality and (72) follows from definition of S and S^c. The "moreover" part follows from Lemma 4.6 and definition of event $\mathcal{T}(A_0)$. $\square$

14.1. *Discussion.* To get a sense of how tightly we can bound these entries in Δ_{A_0}, we do the following calculations. For $\varepsilon = 2/3$, we have $|\Delta_{S^c}|_1 \leq 5\,|\Delta_S|_1$, and hence

$$\begin{aligned} \left|\widehat{\Theta} - \Theta_0\right|_{1,\mathrm{off}} &= |\Delta_{A_0}|_{1,\mathrm{off}} \leq 6\,|\Delta_S|_1 \leq 6\sqrt{|A_0^{-1}|_{0,\mathrm{off}}}\,\|\Delta_S\|_F \\ &\leq 6\sqrt{|A_0^{-1}|_{0,\mathrm{off}}}\frac{9(1+\varepsilon)\lambda_B\sqrt{|A_0^{-1}|_{0,\mathrm{off}} \vee 1}}{2\varphi^2_{\min}(\rho(A_0))} \leq \frac{27(1+\varepsilon)}{\varphi^2_{\min}(\rho(A_0))}\lambda_B\left(|A_0^{-1}|_{0,\mathrm{off}} \vee 1\right). \end{aligned}$$

From Corollary 14.1, we have for $\Theta_0 = \rho(A_0)^{-1}$ and for $S_{A_0} \neq \emptyset$,

$$\frac{|\Delta_{A_0}|_{1,\mathrm{off}}}{|A_0^{-1}|_{0,\mathrm{off}}} \leq \frac{27(1+\varepsilon)}{\varphi^2_{\min}(\rho(A_0))}\lambda_B \asymp \frac{\delta_{n,f}}{\varphi^2_{\min}(\rho(A_0))} \to 0.$$

Such results are useful for bounding the number of falsely selected edges and the number of falsely deleted edges under certain separation assumption, which states that: the minimum absolute value of θ_{ij}: $\min_{i,j,i\neq j}|\theta_{ij}|$ among all non-zero off-diagonal entries in Θ_0 is bounded away from 0.

15. Proofs of Theorems 11.1 and 11.2.

We need some auxiliary results for proving Theorems 11.1 and 11.2. Throughout this section, we assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds, α_n and β_n are as defined in Theorem 4.1, and $1 > \lambda_{A_0} > 3\alpha_n$ and $1 > \lambda_{B_0} > 3\beta_n$.

15.1. *Some auxiliary results.* Let $\widehat{W}_1$ and $\widehat{W}_2$ be as defined in (8a) and (8b). Claim 15.1 provides the large deviation bounds on the spectral norm for estimating $W_1 = \sqrt{\mathrm{tr}(B_0)}\mathrm{diag}(a_{11},\ldots,a_{mm})^{1/2}$ and $W_2 = \sqrt{\mathrm{tr}(A_0)}\mathrm{diag}(b_{11},\ldots,b_{ff})^{1/2}$.

Claim 15.1. *On $\mathcal{X}_0$, for $\beta'_n := \frac{\beta_n}{\sqrt{1-\beta_n}} \le \frac{\lambda_{B_0}}{\sqrt{6}}$ and $\alpha'_n = \frac{\alpha_n}{\sqrt{1-\alpha_n}} \le \frac{\lambda_{A_0}}{\sqrt{6}}$,*

$$\begin{aligned}
\left\|\widehat{W}_1 - W_1\right\|_2 &\le \beta_n \sqrt{\mathrm{tr}(B_0)}\sqrt{a_{\max}}, \quad \left\|\widehat{W}_1^{-1} - W_1^{-1}\right\|_2 \le \beta'_n/(\sqrt{\mathrm{tr}(B_0)}\sqrt{a_{\min}}),\\
\left\|\widehat{W}_2 - W_2\right\|_2 &\le \alpha_n \sqrt{\mathrm{tr}(A_0)}\sqrt{b_{\max}}, \quad \textit{and} \quad \left\|\widehat{W}_2^{-1} - W_2^{-1}\right\|_2 \le \alpha'_n/(\sqrt{\mathrm{tr}(A_0)}\sqrt{b_{\min}}).
\end{aligned}$$

We also need the following proposition in proving Lemma 15.3.

Proposition 15.2. *Let $\widehat{W}$ and W be diagonal positive definite matrices. Let $\widehat{\Psi}$ and Ψ be symmetric positive definite matrices. Then*

$$\begin{aligned}
\left\|\widehat{W}\widehat{\Psi}\widehat{W} - W\Psi W\right\|_2 &\le \left(\left\|\widehat{W} - W\right\|_2 + \|W\|_2\right)^2 \left\|\widehat{\Psi} - \Psi\right\|_2\\
&\quad + \left\|\widehat{W} - W\right\|_2 \left(\left\|\widehat{W} - W\right\|_2 + 2\right) \|\Psi\|_2\\
\left\|\widehat{W}\widehat{\Psi}\widehat{W} - W\Psi W\right\|_F &\le \left(\left\|\widehat{W} - W\right\|_2 + \|W\|_2\right)^2 \left\|\widehat{\Psi} - \Psi\right\|_F\\
&\quad + \left\|\widehat{W} - W\right\|_2 \left(\left\|\widehat{W} - W\right\|_2 + 2\right) \|\Psi\|_F .
\end{aligned}$$

It is clear from the proof of Theorem 4.1 that $\mathcal{T}(A_0) \cap \mathcal{T}(B_0)$ holds on event $\mathcal{X}_0$, and thus all statements in Theorem 4.5 hold on $\mathcal{X}_0$. Lemma 15.3 follows from Theorem 4.5 and Claim 15.1 in view of Proposition 15.2.

It will become clear from Lemma 15.4 that the intermediate results in Lemma 15.3 are useful in bounding the operator norm and the Frobenius norm on Δ and Δ' to be defined in (73) and (74).

Lemma 15.3. *Suppose (A1) and (A2) hold. We have on $\mathcal{X}_0$ for some absolute constants $18 < C, C' < 36$,*

$$\begin{aligned}
\delta_{A,2} := \left\|\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1/\mathrm{tr}(B_0) - A_0\right\|_2 &\le C a_{\max} \kappa \left(\rho(A_0)\right)^2 \lambda_{B_0} \sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee 1},\\
\delta_{B,2} := \left\|\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2/\mathrm{tr}(A_0) - B_0\right\|_2 &\le C' b_{\max} \kappa \left(\rho(B_0)\right)^2 \lambda_{A_0} \sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee 1},\\
\delta_{A,F} := \left\|\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1/\mathrm{tr}(B_0) - A_0\right\|_F &\le C a_{\max} \kappa \left(\rho(A_0)\right)^2 \lambda_{B_0} \sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m},\\
\delta_{B,F} := \left\|\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2/\mathrm{tr}(A_0) - B_0\right\|_F &\le C' b_{\max} \kappa \left(\rho(B_0)\right)^2 \lambda_{A_0} \sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f};
\end{aligned}$$

and for some $10 < C, C' < 19$,

$$\delta^-_{A,2} := \left\| \mathrm{tr}(B_0) \left(\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1 \right)^{-1} - A_0^{-1} \right\|_2 \le \frac{C\lambda_{B_0}\sqrt{|A_0^{-1}|_{0,\mathrm{off}} \vee 1}}{a_{\min}\varphi^2_{\min}(\rho(A_0))},$$

$$\delta^-_{B,2} := \left\| \mathrm{tr}(A_0) \left(\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2 \right)^{-1} - B_0^{-1} \right\|_2 \le \frac{C'\lambda_{A_0}\sqrt{|B_0^{-1}|_{0,\mathrm{off}} \vee 1}}{b_{\min}\varphi^2_{\min}(\rho(B_0))},$$

$$\delta^-_{A,F} := \left\| \mathrm{tr}(B_0) \left(\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1 \right)^{-1} - A_0^{-1} \right\|_F \le \frac{C\lambda_{B_0}\sqrt{|A_0^{-1}|_{0,\mathrm{off}} \vee m}}{a_{\min}\varphi^2_{\min}(\rho(A_0))},$$

$$\delta^-_{B,F} := \left\| \mathrm{tr}(A_0) \left(\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2 \right)^{-1} - B_0^{-1} \right\|_F \le \frac{C'\lambda_{A_0}\sqrt{|B_0^{-1}|_{0,\mathrm{off}} \vee f}}{b_{\min}\varphi^2_{\min}(\rho(B_0))}.$$

We are now ready to state the following convenient results. Let $\|\cdot\|$ be a matrix norm which satisfies the triangle inequality and is multiplicative with respect to the Kronecker product: $\|A \otimes B\| = \|A\|\,\|B\|$. Let

$$\Delta := \widehat{W}_1 \widehat{A}_\rho \widehat{W}_1 \otimes \widehat{W}_2 \widehat{B}_\rho \widehat{W}_2 / (\mathrm{tr}(B_0)\mathrm{tr}(A_0)) - A_0 \otimes B_0, \tag{73}$$

$$\Delta' := \mathrm{tr}(B_0)\mathrm{tr}(A_0) \left(\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1 \right)^{-1} \otimes \left(\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2 \right)^{-1} - A_0^{-1} \otimes B_0^{-1}. \tag{74}$$

Lemma 15.4. *Let* $\Delta_A := A_1 - A$ *and* $\Delta_B := B_1 - B$. *Then*

$$\|A_1 \otimes B_1 - A \otimes B\| \le \|\Delta_A\|\,\|B\| + \|\Delta_B\|\,\|A\| + \|\Delta_A\|\,\|\Delta_B\|.$$

Lemma 15.5. *Let* $\widehat{A \otimes B}$ *be as in* (9). *Then for* $\Sigma_0 = A_0 \otimes B_0$,

$$\left\| \widehat{A \otimes B}^{-1} - \Sigma_0^{-1} \right\| \le (\alpha_n \wedge \beta_n) \left\| A_0^{-1} \right\| \left\| B_0^{-1} \right\| + (1 + \alpha_n \wedge \beta_n) \left\| \Delta' \right\|, \tag{75}$$

$$\left\| \widehat{A \otimes B} - \Sigma_0 \right\| \le \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2} \|A_0\|\,\|B_0\| + (1 + \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2}) \|\Delta\|. \tag{76}$$

15.2. *Proof of Corollary 10.1.* Throughout this proof, we assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds. We first note that the bounds on $\left\| \widehat{B}_* - B_* \right\|$ and $\left\| \widehat{B}_*^{-1} - B_*^{-1} \right\|$ follow from Lemma 15.3 immediately: for $W_2 = \sqrt{\mathrm{tr}(A_0)}\mathrm{diag}(B_0)^{1/2}$ and $\widehat{W}_2$ as defined in (8b)

$$\left\| \widehat{B}_* - B_* \right\| = \tfrac{1}{m} \left\| \widehat{W}_2 \widehat{B}_\rho \widehat{W}_2 - \mathrm{tr}(A_0) B_0 \right\| = \tfrac{\mathrm{tr}(A_0)}{m} \left\| \widehat{W}_2 \widehat{B}_\rho \widehat{W}_2 / \mathrm{tr}(A_0) - B_0 \right\|.$$

We can now plug in the bounds on the operator and the Frobenius norm from Lemma 15.3. The proof for $\widehat{A}_*$ is long and monotonous and hence omitted. □

15.3. *Proof of Theorem 11.1.* Assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds. Let and $\widehat{W}_1$ and $\widehat{W}_2$ be as defined in (8a) and (8b). First we bound $\|\Delta'\|_2$ using Lemmata 15.4 and 15.3: for $10 < C, C' < 19$,

$$\begin{aligned}
\left\|\Delta'\right\|_2 &\leq \delta^-_{A,2}\left\|B_0^{-1}\right\|_2 + \left\|A_0^{-1}\right\|_2 \delta^-_{B,2} + \delta^-_{A,2}\delta^-_{B,2} \\
&\leq \frac{C\lambda_{B_0}\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}}{a_{\min}\varphi^2_{\min}(\rho(A_0))\varphi_{\min}(B_0)} + \frac{C'\lambda_{A_0}\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1}}{b_{\min}\varphi^2_{\min}(\rho(B_0))\varphi_{\min}(A_0)} \\
&\quad + \frac{CC'\lambda_{A_0}\lambda_{B_0}}{a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))}\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1},
\end{aligned} \tag{77}$$

and thus by (62), (63), (47), (48), and for $\|\Delta'\|_2$ bounded as in (77),

$$\begin{aligned}
(\alpha_n \wedge \beta_n)\left\|\Delta'\right\|_2 &\leq \frac{CC'\lambda_{A_0}\lambda_{B_0}\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1}}{3a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))} \times \\
&\quad \left(\frac{\varphi_{\min}(\rho(B_0))}{C'\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1}} + \frac{\varphi_{\min}(\rho(A_0))}{C\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}} + \lambda_{A_0} \wedge \lambda_{B_0}\right) \\
&\leq \frac{2CC'\lambda_{A_0}\lambda_{B_0}}{5a_{\min}b_{\min}}\left(\frac{\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}}{\varphi^2_{\min}(\rho(A_0))}\frac{\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1}}{\varphi^2_{\min}(\rho(B_0))}\right).
\end{aligned} \tag{78}$$

Next we bound the error term $\|\Delta\|_2$. By Lemmata 15.3 and 15.4, we have

$$\begin{aligned}
\|\Delta\|_2 &= \left\|\left(\frac{\widehat{W}_1}{\sqrt{\text{tr}(B_0)}}\right)\widehat{A}_\rho\left(\frac{\widehat{W}_1}{\sqrt{\text{tr}(B_0)}}\right) \otimes \left(\frac{\widehat{W}_2}{\sqrt{\text{tr}(A_0)}}\right)\widehat{B}_\rho\left(\frac{\widehat{W}_2}{\sqrt{\text{tr}(A_0)}}\right) - A_0 \otimes B_0\right\|_2 \\
&\leq \delta_{A,2}\|B_0\|_2 + \|A_0\|_2\delta_{B,2} + \delta_{A,2}\delta_{B,2} \leq C\lambda_{B_0}a_{\max}\|B_0\|_2\kappa(\rho(A_0))^2 \times \\
&\quad \sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1} + C'\lambda_{A_0}b_{\max}\|A_0\|_2\kappa(\rho(B_0))^2\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1} \\
&\quad + CC'\lambda_{A_0}\lambda_{B_0}a_{\max}b_{\max}\kappa(\rho(A_0))^2\kappa(\rho(B_0))^2\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1},
\end{aligned}$$

where $18 < C, C' < 36$, and hence by (62), (63), and (51),

$$\begin{aligned}
\frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2}\|\Delta\|_2 &\leq (CC'/2)\lambda_{A_0}\lambda_{B_0}a_{\max}b_{\max}\kappa(\rho(A_0))^2\kappa(\rho(B_0))^2 \times \\
&\quad \left(\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1}\right) \times \left(\frac{1}{C} + \frac{1}{C'} + (\lambda_{A_0} \wedge \lambda_{B_0})\right) \\
&\leq \frac{5CC'}{9}\lambda_{A_0}\lambda_{B_0}a_{\max}b_{\max}\kappa(\rho(A_0))^2\kappa(\rho(B_0))^2\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}\sqrt{|B_0^{-1}|_{0,\text{off}} \vee 1}.
\end{aligned}$$

The theorem is thus proved by inserting the bounds immediately above in (76), and by inserting (77) and (78) in (75). □

15.4. *Proof of Theorem 11.2.* Assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds. First by Lemmata 15.4 and 15.3, we have

$$\begin{aligned}\|\Delta\|_F &\le \delta_{A,F}\|B_0\|_F + \|A_0\|_F\,\delta_{B,F} + \delta_{A,F}\delta_{B,F} \le C\lambda_{B_0}a_{\max}\|B_0\|_F\,\kappa(\rho(A_0))^2 \times \\ &\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m} + C'\lambda_{A_0}b_{\max}\|A_0\|_F\,\kappa(\rho(B_0))^2\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f} \\ &+ CC'\lambda_{A_0}\lambda_{B_0}a_{\max}b_{\max}\kappa(\rho(A_0))^2\kappa(\rho(B_0))^2\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f}\end{aligned}$$

where $18 < C, C' < 36$. Hence by (51), (62), and (63), and the fact that $\|B_0\|_F \le \sqrt{f}\,\|B_0\|_2$, and $\|A_0\|_F \le \sqrt{m}\,\|A_0\|_2$,

$$\begin{aligned}\frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2}\|\Delta\|_F &\le (CC'/2)\lambda_{A_0}\lambda_{B_0}a_{\max}b_{\max}\kappa(\rho(A_0))^2\kappa(\rho(B_0))^2 \cdot \\ &\left(\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f}\right) \times \left(\frac{1}{C} + \frac{1}{C'} + (\lambda_{A_0} \wedge \lambda_{B_0})\right) \\ &\le \frac{5CC'}{9}\lambda_{A_0}\lambda_{B_0}a_{\max}b_{\max}\kappa(\rho(A_0))^2\kappa(\rho(B_0))^2\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f}.\end{aligned}$$

We next bound $\|\Delta'\|_F$. Let $10 < C, C' < 19$. By Lemmata 15.4 and 15.3,

$$\begin{aligned}\|\Delta'\|_F &\le \delta^-_{A,F}\left\|B_0^{-1}\right\|_F + \left\|A_0^{-1}\right\|_F\delta^-_{B,F} + \delta^-_{A,F}\delta^-_{B,F} \\ &\le \frac{C\lambda_{B_0}\left\|B_0^{-1}\right\|_F\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}}{a_{\min}\varphi^2_{\min}(\rho(A_0))} + \frac{C'\lambda_{A_0}\left\|A_0^{-1}\right\|_F\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f}}{b_{\min}\varphi^2_{\min}(\rho(B_0))} + \\ &+ \frac{CC'\lambda_{A_0}\lambda_{B_0}}{a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f}.\end{aligned}$$

And thus by (47), (48), (62), and (63), $\frac{\lambda_{A_0} \wedge \lambda_{B_0}}{3}\|\Delta'\|_F \le$

$$\begin{aligned}&\frac{C\lambda_{B_0}\lambda_{A_0}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}\sqrt{f}}{3a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi_{\min}(\rho(B_0))} + \frac{C'\lambda_{A_0}\lambda_{B_0}\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f}\sqrt{m}}{3a_{\min}b_{\min}\varphi^2_{\min}(\rho(B_0))\varphi_{\min}(\rho(A_0))} \\ &+ \frac{(\lambda_{A_0} \wedge \lambda_{B_0})CC'\lambda_{A_0}\lambda_{B_0}}{3a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f} \\ &\le \frac{2CC'\lambda_{A_0}\lambda_{B_0}}{5a_{\min}b_{\min}\varphi^2_{\min}(\rho(A_0))\varphi^2_{\min}(\rho(B_0))}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}\sqrt{\left|B_0^{-1}\right|_{0,\mathrm{off}} \vee f}.\end{aligned}$$

The theorem is thus proved by inserting the bounds above in (75) and (76).
□

15.5. *Proof of Claim 15.1.* Assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds. Then for $\beta := \beta_n$, $\forall i = 1, \ldots, m$,

$$\max_{i=1,\ldots,m} \left| \frac{\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|x(t)^i\right\|_2^2}}{\sqrt{a_{ii}\mathrm{tr}(B_0)}} - 1 \right| \leq \left(1 - \sqrt{1-\beta}\right) \vee \left(\sqrt{1+\beta} - 1\right) \leq \beta.$$

Thus the first inequality holds. Similarly, on $\mathcal{X}_0$, we obtain

$$\forall i = 1, \ldots, m, \quad \frac{1}{\sqrt{1+\beta}} \leq \frac{\sqrt{a_{ii}\mathrm{tr}(B_0)}}{\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|x(t)^i\right\|_2^2}} \leq \frac{1}{\sqrt{1-\beta}} \quad \text{and hence}$$

$$\left| \frac{\sqrt{a_{ii}\mathrm{tr}(B_0)}}{\sqrt{\frac{1}{n}\sum_{t=1}^n \left\|x(t)^i\right\|_2^2}} - 1 \right| \leq \left(\frac{\sqrt{1+\beta}-1}{\sqrt{1+\beta}}\right) \vee \left(\frac{1-\sqrt{1-\beta}}{\sqrt{1-\beta}}\right) \leq \frac{\beta}{\sqrt{1-\beta}}.$$

The statements about $\widehat{W}_2$ and $\widehat{W}_2^{-1}$ hold following the same line of arguments. □

15.6. *Proof of Proposition 15.2.* By the triangle inequality,

$$\begin{aligned}
&\left\|\widehat{W}\widehat{\Psi}\widehat{W} - W\Psi W\right\|_2 = \\
&\quad \left\|(\widehat{W} - W)\widehat{\Psi}(\widehat{W} - W) + W\widehat{\Psi}(\widehat{W} - W) + (\widehat{W} - W)\widehat{\Psi}W + W(\widehat{\Psi} - \Psi)W\right\|_2 \\
&\leq \left(\left\|\widehat{W} - W\right\|_2^2 + 2\left\|\widehat{W} - W\right\|_2 \|W\|_2\right)\|\Psi\|_2 + \left(\left\|\widehat{W} - W\right\|_2 + \|W\|_2\right)^2 \left\|\widehat{\Psi} - \Psi\right\|_2.
\end{aligned}$$

Similarly, we have $\left\|\widehat{W}\widehat{\Psi}\widehat{W} - W\Psi W\right\|_F \leq$

$$\begin{aligned}
&\left\|\widehat{W} - W\right\|_2^2 \left\|\widehat{\Psi}\right\|_F + 2\left\|\widehat{W} - W\right\|_2 \|W\|_2 \left\|\widehat{\Psi}\right\|_F + \|W\|_2 \left\|\widehat{\Psi} - \Psi\right\|_F \|W\|_2 \\
&\leq \left\|\widehat{W} - W\right\|_2 \left(\left\|\widehat{W} - W\right\|_2 + 2\right)\|\Psi\|_F + \left(\left\|\widehat{W} - W\right\|_2 + \|W\|_2\right)^2 \left\|\widehat{\Psi} - \Psi\right\|_F.
\end{aligned}$$

□

15.7. *Proof of Lemma 15.3.* We assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds. Let $W := W_1/\sqrt{\mathrm{tr}(B_0)}$ and $\widehat{W} := \widehat{W}_1/\sqrt{\mathrm{tr}(B_0)}$. By Proposition 15.2, Claim 15.1, and Theorem 4.5, for $\beta := \beta_n \leq \lambda_{B_0}/3 < 1/3$,

$$\begin{aligned}
\delta_{A,2} &:= \left\|\widehat{W}\widehat{A}_\rho\widehat{W} - \mathrm{diag}(A_0)^{1/2}\rho(A_0)\mathrm{diag}(A_0)^{1/2}\right\|_2 \\
&\leq (1+\beta)^2 a_{\max}\left\|\widehat{A}_\rho - \rho(A_0)\right\|_2 + (\beta^2 + 2\beta)a_{\max}\|\rho(A_0)\|_2 \\
&\leq C\lambda_{B_0} a_{\max}\kappa(\rho(A_0))^2\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee 1},
\end{aligned}$$

where we used the fact that $\kappa(\rho(A_0)) \geq \|\rho(A_0)\|_2 \geq 1$, given that $\varphi_{\min}(\rho(A_0)) \leq 1$ as shown in (62); and

$$\begin{aligned}\delta_{A,F} &\leq (1+\beta)^2 a_{\max} \left\|\widehat{A}_\rho - \rho(A_0)\right\|_F + (\beta^2 + 2\beta) a_{\max} \|\rho(A_0)\|_F \\ &\leq C\lambda_{B_0} a_{\max} \kappa\left(\rho(A_0)\right)^2 \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}.\end{aligned}$$

Similarly, for $\beta' := \beta_n/\sqrt{1-\beta_n} \leq \lambda_{B_0}/\sqrt{6}$, where $\beta_n < 1/3$, we have by Proposition 15.2, Claim 15.1, and Theorem 4.5,

$$\begin{aligned}\delta^-_{A,2} &\leq \frac{(1+\beta')^2}{a_{\min}} \left\|\widehat{A}_\rho^{-1} - \rho(A_0)^{-1}\right\|_2 + \frac{(\beta'+2)\beta'}{a_{\min}} \left\|\rho(A_0)^{-1}\right\|_2 \\ &\leq (2C+1)\lambda_{B_0} \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}/(a_{\min}\varphi^2_{\min}(\rho(A_0))),\end{aligned}$$

and

$$\begin{aligned}\delta^-_{A,F} &\leq \frac{(1+\beta')^2}{a_{\min}} \left\|\widehat{A}_\rho^{-1} - \rho(A_0)^{-1}\right\|_F + \frac{(\beta'+2)\beta'}{a_{\min}} \left\|\rho(A_0)^{-1}\right\|_F \\ &\leq (2C+1)\lambda_{B_0} \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}/(a_{\min}\varphi^2_{\min}(\rho(A_0))),\end{aligned}$$

where $9/2 < C < 9$ and hence the statement in the Lemma holds. The bounds for B_0 and B_0^{-1} can be derived in exactly the same manner, and hence omitted. □

15.8. *Proof of Lemma 15.4* . By distributivity of the Kronecker product,

$$\begin{aligned}A_1 \otimes B_1 &= (A + \Delta_A) \otimes (B + \Delta_B) = A \otimes (B + \Delta_B) + \Delta_A \otimes (B + \Delta_B) \\ &= A \otimes B + A \otimes \Delta_B + \Delta_A \otimes B + \Delta_A \otimes \Delta_B,\end{aligned}$$

hence by the triangle inequality and the multiplicativity of the norm,

$$\begin{aligned}&\|A_1 \otimes B_1 - A \otimes B\| = \|A \otimes \Delta_B + \Delta_A \otimes B + \Delta_A \otimes \Delta_B\| \\ &\leq \|A \otimes \Delta_B\| + \|\Delta_A \otimes B\| + \|\Delta_A \otimes \Delta_B\| = \|\Delta_B\| \|A\| + \|\Delta_A\| \|B\| + \|\Delta_A\| \|\Delta_B\|.\end{aligned}$$

□

15.9. *Proof of Lemma 15.5.* By the triangle inequality and the multiplicativity of the norm with respect to the Kronecker product, (73), and (74)

$$\text{tr}(A_0)\text{tr}(B_0) \left\|\left(\widehat{W}_1^{-1}\widehat{A}_\rho^{-1}\widehat{W}_1^{-1}\right) \otimes \left(\widehat{W}_2^{-1}\widehat{B}_\rho^{-1}\widehat{W}_2^{-1}\right)\right\| \leq$$

$$\left\|A_0^{-1}\right\| \left\|B_0^{-1}\right\| + \left\|\Delta'\right\| \tag{79}$$

$$\left\|(\widehat{W}_1\widehat{A}_\rho\widehat{W}_1) \otimes (\widehat{W}_2\widehat{B}_\rho\widehat{W}_2)/\text{tr}(A_0)\text{tr}(B_0)\right\| \leq \|A_0\| \|B_0\| + \|\Delta\|. \tag{80}$$

By (20), we have for $\lambda_{A_0} \geq 3\alpha_n$ and $\lambda_{B_0} \geq 3\beta_n$, where $\alpha_n \wedge \beta_n \leq \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{3}$,

$$\left| \frac{1}{\frac{1}{n}\sum_{t=1}^n \|X(t)\|_F^2} - \frac{1}{\mathrm{tr}(A_0)\mathrm{tr}(B_0)} \right| = \left| \frac{\frac{1}{n}\sum_{t=1}^n \|X(t)\|_F^2 - \mathrm{tr}(A_0)\mathrm{tr}(B_0)}{\frac{1}{n}\sum_{t=1}^n \|X(t)\|_F^2 \cdot \mathrm{tr}(A_0)\mathrm{tr}(B_0)} \right|$$
$$\leq \left| \frac{(\alpha_n \wedge \beta_n)}{\frac{1}{n}\sum_{t=1}^n \|X(t)\|_F^2} \right| \leq \frac{\alpha_n \wedge \beta_n}{\mathrm{tr}(A_0)\mathrm{tr}(B_0)(1-\alpha_n \wedge \beta_n)}$$
$$\text{thus } \left| \frac{\mathrm{tr}(A_0)\mathrm{tr}(B_0)}{\frac{1}{n}\sum_{t=1}^n \|X(t)\|_F^2} - 1 \right| \leq \frac{\alpha_n \wedge \beta_n}{1-\alpha_n \wedge \beta_n} \leq \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2}. \tag{81}$$

By the triangle inequality, definition of Δ, (80), and (81), $\left\| \widehat{A \otimes B}_0 - A_0 \otimes B_0 \right\|$

$$\leq \left| \frac{1}{\frac{1}{n}\sum_{i=1}^n \|X(i)\|_F^2} - \frac{1}{\mathrm{tr}(A_0)\mathrm{tr}(B_0)} \right| \left\| \left(\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1\right) \otimes \left(\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2\right) \right\|$$
$$+ \left\| \left(\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1\right) \otimes \left(\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2\right) / \mathrm{tr}(A_0)\mathrm{tr}(B_0) - A_0 \otimes B_0 \right\|$$
$$\leq \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2} \|A_0\| \|B_0\| + (1 + \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2}) \|\Delta\|.$$

By definition of Δ', (20), and (79), $\left\| \widehat{A \otimes B}^{-1} - A_0^{-1} \otimes B_0^{-1} \right\| =$

$$\left\| \left(\frac{1}{n}\sum_{t=1}^n \|X(t)\|_F^2\right) \left(\widehat{W}_1^{-1} \widehat{A}_\rho^{-1} \widehat{W}_1^{-1}\right) \otimes \left(\widehat{W}_2^{-1} \widehat{B}_\rho^{-1} \widehat{W}_2^{-1}\right) - A_0^{-1} \otimes B_0^{-1} \right\|$$
$$\leq \left| \frac{1}{n}\sum_{i=1}^n \|X(i)\|_F^2 - \mathrm{tr}(A_0)\mathrm{tr}(B_0) \right| \left\| \left(\widehat{W}_1^{-1} \widehat{A}_\rho^{-1} \widehat{W}_1^{-1}\right) \otimes \left(\widehat{W}_2^{-1} \widehat{B}_\rho^{-1} \widehat{W}_2^{-1}\right) \right\|$$
$$+ \left\| \mathrm{tr}(A_0)\mathrm{tr}(B_0) \left(\widehat{W}_1^{-1} \widehat{A}_\rho^{-1} \widehat{W}_1^{-1}\right) \otimes \left(\widehat{W}_2^{-1} \widehat{B}_\rho^{-1} \widehat{W}_2^{-1}\right) - A_0^{-1} \otimes B_0^{-1} \right\|$$
$$\leq (\alpha_n \wedge \beta_n) \left(\left\|A_0^{-1}\right\| \left\|B_0^{-1}\right\| + \left\|\Delta'\right\| \right) + \left\|\Delta'\right\|. \quad \square$$

16. Proof of Theorem 3.3. First we have the following bound following Theorem 6 in [1]. We show here the rates in the operator norm.

THEOREM 16.1. *Suppose that (A2) hold. Suppose that* $\Theta_0 := \rho(A_0)^{-1} \in \mathcal{U}_0(d_0(m), M)$ *and* $\Phi_0 := \rho(B_0)^{-1} \in \mathcal{U}_0(d_0(f), K)$*. Let* $\widehat{\Theta}_{\mathrm{CL}}$ *and* $\widehat{\Phi}_{\mathrm{CL}}$ *be the CLIME estimators as defined in* (15) *and* (16) *with sample correlation matrices* $\widehat{\Gamma}(A_0)$ *and* $\widehat{\Gamma}(B_0)$ *as in* (7a) *and* (7b) *as their input. Suppose that event* $\mathcal{T}(A_0)$ *holds for* $\widehat{\Gamma}(A_0)$ *for some* $\delta_{n,f}$ *and event* $\mathcal{T}(B_0)$ *holds for* $\widehat{\Gamma}(B_0)$ *for some* $\delta_{n,m}$*, such that*

$$\delta_{n,f} d_0(m) \|\Theta_0\|_1^2 = o(1) \quad \text{and} \quad \delta_{n,m} d_0(f) \|\Phi_0\|_1^2 = o(1).$$

Suppose that for some $0 < \epsilon, \varepsilon < 1$, $\lambda_{B_0} = \delta_{n,f}/\varepsilon$ *and* $\lambda_{A_0} = \delta_{n,m}/\epsilon$. *Let*

$$\lambda_M = \|\Theta_0\|_1 \lambda_{B_0} \quad \text{and} \quad \lambda_K = \|\Phi_0\|_1 \lambda_{A_0}.$$

Then on event $\mathcal{T}(A_0) \cap \mathcal{T}(B_0)$, *for some absolute constant* C, $\widehat{\Theta}_{\mathrm{CL}}, \widehat{\Phi}_{\mathrm{CL}} \succ 0$ *under (A2), for* $\widehat{A}_\rho := \widehat{\Theta}_{\mathrm{CL}}^{-1}$ *and* $\widehat{B}_\rho := \widehat{\Phi}_{\mathrm{CL}}^{-1}$, *we have*

$$\begin{aligned}
\left\|\widehat{\Theta}_{\mathrm{CL}} - \Theta_0\right\|_2 &\leq Cd_0(m) \|\Theta_0\|_1 \lambda_M \asymp d_0(m) \|\Theta_0\|_1^2 \lambda_{B_0} \\
\left\|\widehat{\Phi}_{\mathrm{CL}} - \Phi_0^{-1}\right\|_2 &\leq Cd_0(f) \|\Phi_0\|_1 \lambda_K \asymp d_0(f) \|\Phi_0\|_1^2 \lambda_{A_0} \\
\left\|\widehat{A}_\rho - \rho(A_0)\right\|_2 &\leq C(1+o(1)) \|\rho(A_0)\|_2^2 d_0(m) \|\Theta_0\|_1^2 \lambda_{B_0}, \\
\left\|\widehat{B}_\rho - \rho(B_0)\right\|_2 &\leq C(1+o(1)) \|\rho(B_0)\|_2^2 d_0(f) \|\Phi_0\|_1^2 \lambda_{A_0}
\end{aligned}$$

PROOF. Note that on event $\mathcal{T}(A_0)$ and $\mathcal{T}(B_0)$, the choices of the penalties satisfy $\lambda_M \geq \|\Theta_0\|_1 \delta_{n,f}$ $\lambda_K \geq \|\Phi_0\|_1 \delta_{n,m}$ as desired. Then the bounds on $\left\|\widehat{\Theta}_{\mathrm{CL}} - \Theta_0\right\|_2$ and $\left\|\widehat{\Phi} - \Phi_0^{-1}\right\|_2$ in the theorem statement hold by Theorem 6 in [1]. To see the last two inequalities, we have for $\widehat{A}_\rho$,

$$\left\|\widehat{A}_\rho - \rho(A_0)\right\|_2 \leq \frac{\left\|\widehat{\Theta}_{\mathrm{CL}} - \Theta_0\right\|_2}{\varphi_{\min}(\widehat{\Theta}_{\mathrm{CL}})\varphi_{\min}(\Theta_0)} \leq (1+o(1)) d_0(m) \|\Theta_0\|_1^2 \lambda_{B_0} \varphi_{\max}^2(\rho(A_0))$$

as desired. Similarly, we obtain the bound for $\left\|\widehat{B}_\rho - \rho(B_0)\right\|$. □

It is clear from Lemma 15.4 that the intermediate results in Lemma 16.2 are useful in bounding the operator norm on Δ and Δ' as defined in (73) and (74). The proof follows exactly that of Lemma 15.3, hence omitted.

LEMMA 16.2. *Suppose (A1') and (A2) hold. Let* $\alpha_n < \lambda_{A_0}/3$ *and* $\beta_n < \lambda_{B_0}/3$ *where* $\lambda_{A_0}, \lambda_{B_0} < 1$. *We have on* $\mathcal{X}_0$ *for some absolute constants* $48 > C, C' > 5$ *as defined in Theorem 16.1,*

$$\begin{aligned}
\delta_{A,2} &:= \left\|\frac{\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1}{\mathrm{tr}(B_0)} - A_0\right\|_2 \leq 2Ca_{\max}\lambda_{B_0} d_0(m) \|\rho(A_0)\|_2^2 \|\Theta_0\|_1^2, \\
\delta_{B,2} &:= \left\|\frac{\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2}{\mathrm{tr}(A_0)} - B_0\right\|_2 \leq 2Cb_{\max}\lambda_{A_0} d_0(f) \|\rho(B_0)\|_2^2 \|\Phi_0\|_1^2, \\
\delta_{A,2}^- &:= \left\|\mathrm{tr}(B_0)\left(\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1\right)^{-1} - A_0^{-1}\right\|_2 \leq 2C\lambda_{B_0} \|\Theta_0\|_1^2 d_0(m)/a_{\min}, \\
\delta_{B,2}^- &:= \left\|\mathrm{tr}(A_0)\left(\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2\right)^{-1} - B_0^{-1}\right\|_2 \leq 2C\lambda_{A_0} \|\Phi_0\|_1^2 d_0(f)/b_{\min}.
\end{aligned}$$

Proof of Theorem 3.3. Throughout this proof, we assume that event $\mathcal{X}_0$ as defined in Lemma 4.3 holds. We only bound the relative error δ in the operator norm for estimating $A_0 \otimes B_0$. The bound for δ' for estimating $A_0^{-1} \otimes B_0^{-1}$ follows exactly the same line of arguments. Let $\widehat{W}_1$ and $\widehat{W}_2$ be as defined in (8a) and (8b). On $\mathcal{X}_0$, we have by Lemmata 15.4 and 16.2, and r_a, r_b as defined in (46),

$$
\begin{aligned}
r_A &:= \left\| \frac{\widehat{W}_1 \widehat{A}_\rho \widehat{W}_1}{\mathrm{tr}(B_0)} - A_0 \right\|_2 / \|A_0\|_2 \le 2C r_a \lambda_{B_0} d_0(m) \|\rho(A_0)\|_2 \|\Theta_0\|_1^2 \to 0 \\
r_B &:= \left\| \frac{\widehat{W}_2 \widehat{B}_\rho \widehat{W}_2}{\mathrm{tr}(A_0)} - B_0 \right\|_2 / \|B_0\|_2 \le 2C r_b \lambda_{A_0} d_0(f) \|\rho(B_0)\|_2 \|\Theta_0\|_1^2 \to 0
\end{aligned}
$$

under (A1'), and (76), for $r_A = \Omega(\lambda_{B_0})$, $r_B = \Omega(\lambda_{A_0})$, and $r_A, r_B \to 0$,

$$
\begin{aligned}
\frac{\left\| \widehat{A \otimes B} - A_0 \otimes B_0 \right\|_2}{\|A_0\|_2 \|B_0\|_2} &\le \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2} + (1 + \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2}) \frac{\|\Delta\|}{\|A_0\| \|B_0\|} \\
&\le \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2} + (1 + \frac{\lambda_{A_0} \wedge \lambda_{B_0}}{2})(r_A + r_B + r_A r_B) \\
&\asymp 2C r_a \lambda_{B_0} d_0(m) \|\rho(A_0)\|_2 \|\Theta_0\|_1^2 + 2C r_b \lambda_{A_0} d_0(f) \|\rho(B_0)\|_2 \|\Phi_0\|_1^2 \\
&\quad + 2C r_a \lambda_{B_0} d_0(m) \|\rho(A_0)\|_2 \|\Theta_0\|_1^2 \, 2C r_b \lambda_{A_0} d_0(f) \|\rho(B_0)\|_2 \|\Phi_0\|_1^2 \\
&= O\left(\lambda_{B_0} d_0(m) \|\Theta_0\|_1^2 + \lambda_{A_0} d_0(f) \|\Theta_0\|_1^2 \right)
\end{aligned}
$$

The theorem is thus proved by inserting the bounds immediately above in (76). □

17. Proofs for the Flip-Flop methods. We state in Lemma 17.1 a corollary of Theorem 13.1, where we set $M = A_0^{-1}$ and $N = B_0^{-1}$. Lemma 17.1 is needed in proving Theorem 6.2 and Theorem 6.4.

LEMMA 17.1. *Let $\lambda_{f,n}, \lambda_{m,n}$ be as in (32). On event $\mathcal{E}_0$,*

$$
\begin{aligned}
\left\| \mathrm{diag}(A_*)^{-1/2} \widetilde{A}(B_*) \mathrm{diag}(A_*)^{-1/2} - \rho(A_0) \right\|_{\max} &\le \lambda_{f,n}, \\
\left\| \mathrm{diag}(B_*)^{-1/2} \widetilde{B}(A_*) \mathrm{diag}(B_*)^{-1/2} - \rho(B_0) \right\|_{\max} &\le \lambda_{m,n}
\end{aligned}
$$

where $\mathbb{P}(\mathcal{E}_0) \ge 1 - \frac{3}{\max(m,f)^2}$.

17.1. *Convergence bounds for $\widehat{A}_*$ in Step 2.*

COROLLARY 17.2. *Suppose (A1), (A2), and (A3) hold. Assume that $\eta \le 1/4$. Then on event $\mathcal{A}_1$, for $\widehat{A}_*$ as constructed in Step 2 with λ_{B_1} as in (38), and some constant $9 < C < 18$,*

$$\begin{aligned}
\left\|\widehat{A}_* - A_*\right\|_2 &\le 2C\lambda_{B_1} a_{*,\max}\kappa\left(\rho(A_0)\right)^2 \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}, \\
\left\|\widehat{A}_* - A_*\right\|_F &\le 2C\lambda_{B_1} a_{*,\max}\kappa(\rho(A_0))^2 \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}, \\
\left\|\widehat{A}_*^{-1} - A_*^{-1}\right\|_2 &\le C\lambda_{B_1} \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}/(a_{*,\min}\varphi^2_{\min}(\rho(A_0))), \\
\textit{and} \quad \left\|\widehat{A}_*^{-1} - A_*^{-1}\right\|_F &\le C\lambda_{B_1} \sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}/(a_{*,\min}\varphi^2_{\min}(\rho(A_0))).
\end{aligned}$$

17.2. *Proof of Lemma 6.1.* In order to prove Lemma 6.1, we need the following auxiliary results, where we assume that λ_{A_0}, B_1, and $\widehat{B}_\rho$ are as defined in Lemma 6.1. Suppose all conditions in Lemma 6.1 hold.

CLAIM 17.3. *Suppose all conditions in Lemma 6.1 hold. Let $\check{B}_0 = B_1 - B_*$, where $B_* = (\text{tr}(A_0)/m)B_0$. Then on $\mathcal{X}_0$,*

$$\lambda_{A_0}\left|\widehat{B}_\rho^{-1}\right|_{1,\text{off}} - \frac{\alpha}{1-\alpha}\left|\widehat{B}_\rho^{-1}\right|_1 \le \text{tr}(\check{B}_0 B_1^{-1}) \le \lambda_{A_0}\left|\widehat{B}_\rho^{-1}\right|_{1,\text{off}} + \frac{\alpha}{1-\alpha}\left|\widehat{B}_\rho^{-1}\right|_1.$$

COROLLARY 17.4. *Suppose all conditions in Lemma 6.1 hold. On $\mathcal{X}_0$,*

$$\left|\text{tr}(\check{B}_0 B_1^{-1})\right|/f \le \widetilde{\mu} \le \mu = \frac{\alpha}{1-\alpha}\left|\rho(B_0)^{-1}\right|_1/f + \lambda_{A_0}\left|\rho(B_0)^{-1}\right|_{1,\text{off}}/f + o\left(\lambda_{A_0}\right).$$

Proof of Lemma 6.1. Throughout this proof, we assume (A1) (A2) and (A3) hold. Let

$$\begin{aligned}
\widehat{R}_A &= \left[\text{vec}\left\{\widetilde{\mathrm{S}}_{\mathrm{n}}^{11}\right\}\ldots\text{vec}\left\{\widetilde{\mathrm{S}}_{\mathrm{n}}^{1\mathrm{f}}\right\}\ldots\text{vec}\left\{\widetilde{\mathrm{S}}_{\mathrm{n}}^{\mathrm{ff}}\right\}\right] \\
R_A &= \left[b_{11}\text{vec}\left\{\mathrm{A}_0\right\}\ldots\mathrm{b}_{1\mathrm{f}}\text{vec}\left\{\mathrm{A}_0\right\}\ldots\mathrm{b}_{\mathrm{ff}}\text{vec}\left\{\mathrm{A}_0\right\}\right] = \text{vec}\left\{\mathrm{A}_0\right\}\otimes(\text{vec}\left\{\mathrm{B}_0^{\mathrm{T}}\right\})^{\mathrm{T}}
\end{aligned}$$

In order to obtain a bound on $\left\|\widetilde{A}(B_1) - A_*\right\|_{\max}$, we first

$$\text{vec}\left\{\widetilde{\mathrm{A}}(\mathrm{B}_1) - \mathrm{A}_*\right\} = \tfrac{1}{\mathrm{f}}(\widehat{\mathrm{R}}_{\mathrm{A}} - \mathrm{R}_{\mathrm{A}})\text{vec}\left\{\mathrm{B}_*^{-1}\right\} + \tfrac{1}{\mathrm{f}}\widehat{\mathrm{R}}_{\mathrm{A}}\text{vec}\left\{\mathrm{B}_1^{-1} - \mathrm{B}_*^{-1}\right\}, \tag{82}$$

where in (82) we use the following equation

$$\begin{aligned}
\tfrac{1}{f}R_A\text{vec}\left\{\mathrm{B}_*^{-1}\right\} &= = \tfrac{1}{f}\text{vec}\left\{\mathrm{A}_0\right\}\otimes(\text{vec}\left\{\mathrm{B}_0^{\mathrm{T}}\right\})^{\mathrm{T}}\text{vec}\left\{\left(\tfrac{\mathrm{tr}(\mathrm{A}_0)}{\mathrm{m}}\mathrm{B}_0\right)^{-1}\right\} \\
&= \frac{m}{\text{tr}(A_0)}\text{vec}\left\{\mathrm{A}_0\right\}\tfrac{1}{\mathrm{f}}(\text{vec}\left\{\mathrm{B}_0^{\mathrm{T}}\right\})^{\mathrm{T}}\text{vec}\left\{\mathrm{B}_0^{-1}\right\} = \text{vec}\left\{\mathrm{A}_*\right\}.
\end{aligned}$$

In order to proceed, we now compute $\widetilde{A}(B_1)$ as defined in (29), where an explicit formula for B_1^{-1} is needed. For $\check{B}_0 := B_1 - B_*$, we have $\Delta^0 := B_1^{-1} - B_*^{-1} = -B_1^{-1}(B_1 - B_*)B_*^{-1} = -B_1^{-1}\check{B}_0 B_*^{-1}$. Thus

$$\begin{aligned}\widehat{R}_A \mathrm{vec}\left\{ \mathrm{B}_1^{-1} - \mathrm{B}_*^{-1} \right\} &= \widehat{R}_A(-B_1^{-1}(B_1 - B_*)B_*^{-1}) = \widehat{R}_A \mathrm{vec}\left\{ -\mathrm{B}_1^{-1}\check{\mathrm{B}}_0\mathrm{B}_*^{-1} \right\} \\ &= R_A \mathrm{vec}\left\{ -\mathrm{B}_1^{-1}\check{\mathrm{B}}_0\mathrm{B}_*^{-1} \right\} + (\widehat{\mathrm{R}}_\mathrm{A} - \mathrm{R}_\mathrm{A})\mathrm{vec}\left\{ \Delta^0 \right\}.\end{aligned}$$

We now insert the equation immediately above in (82) to obtain

$$\begin{aligned}\mathrm{vec}\left\{ \widetilde{\mathrm{A}}(\mathrm{B}_1) - \mathrm{A}_* \right\} &= \tfrac{1}{\mathrm{f}}(\widehat{\mathrm{R}}_\mathrm{A} - \mathrm{R}_\mathrm{A})\mathrm{vec}\left\{ \mathrm{B}_*^{-1} \right\} \\ &\quad + \tfrac{1}{f} R_A \mathrm{vec}\left\{ -\mathrm{B}_1^{-1}\check{\mathrm{B}}_0\mathrm{B}_*^{-1} \right\} + \tfrac{1}{\mathrm{f}}(\widehat{\mathrm{R}}_\mathrm{A} - \mathrm{R}_\mathrm{A})\mathrm{vec}\left\{ \Delta^0 \right\} \\ &=: U_1 + U_2 + U_3, \end{aligned} \tag{83}$$

where the matrix correspondent of each summand will be denoted by M_1, M_2, and M_3 respectively. Now for the first summand on the RHS, we have

$$\begin{aligned} U_1 &= \tfrac{1}{f}(\widehat{R}_A - R_A)\mathrm{vec}\left\{ \mathrm{B}_*^{-1} \right\} = \mathrm{vec}\left\{ \widetilde{\mathrm{A}}(\mathrm{B}_*) - \mathrm{A}_* \right\} \\ &= \tfrac{1}{f}\textstyle\sum_{k=1}^{f}\sum_{j=1}^{f} \mathrm{vec}\left\{ \widetilde{\mathrm{S}}_\mathrm{n}^\mathrm{kj} - \mathrm{b}_\mathrm{kj}\mathrm{A}_0 \right\} (\mathrm{B}_*^{-1})_\mathrm{jk}, \end{aligned}$$

and $M_1 = \widetilde{A}(B_*) - A_* = \frac{1}{f}\sum_{k=1}^{f}\sum_{\ell=1}^{f} \widetilde{S}_n^{\ell k}(B_*^{-1})_{k\ell} - A_*$. By Lemma 17.1, we have on event $\mathcal{E}_0$,

$$\forall i, j \quad |M_{1,ij}| = \left| \widetilde{A}_{ij}(B_*) - a_{*,ij} \right| \leq \sqrt{a_{*,ii}a_{*,jj}}\lambda_{f,n}.$$

We now examine the second summand on the RHS of (83), where recall that $\check{B}_0 = B_1 - B_*$. We have

$$\begin{aligned} R_A \mathrm{vec}\left\{ \mathrm{B}_1^{-1}\check{\mathrm{B}}_0\mathrm{B}_*^{-1} \right\} &= \mathrm{vec}\left\{ \mathrm{A}_0 \right\} \mathrm{vec}\left\{ \mathrm{B}_0^\mathrm{T} \right\}^\mathrm{T} (\mathrm{B}_*^{-\mathrm{T}} \otimes \mathrm{B}_1^{-1})\mathrm{vec}\left\{ \check{\mathrm{B}}_0 \right\} \\ &= \mathrm{vec}\left\{ \mathrm{A}_* \right\} \mathrm{tr}(\mathrm{B}_1^{-1}\check{\mathrm{B}}_0). \end{aligned}$$

Then clearly on $\mathcal{X}_0$,

$$\begin{aligned} U_2 &= \tfrac{1}{f} R_A \mathrm{vec}\left\{ -\mathrm{B}_1^{-1}\check{\mathrm{B}}_0\mathrm{B}_*^{-1} \right\} = \tfrac{1}{\mathrm{f}}\mathrm{vec}\left\{ \mathrm{A}_* \right\} \mathrm{tr}(-\mathrm{B}_1^{-1}\check{\mathrm{B}}_0), \\ \text{and} \quad M_2 &= A_* \tfrac{1}{f}\mathrm{tr}(B_1^{-1}\check{B}_0). \end{aligned}$$

By Claim 17.3, we have on event $\mathcal{X}_0$ and for $\widetilde{\mu}$ as defined in (34),

$$|M_{2,ij}| = |a_{*,ij}| \left| \mathrm{tr}(\check{B}_0 B_1^{-1})/f \right| \leq |a_{*,ij}| \widetilde{\mu}$$

Finally, we bound the third summand. Let $\Delta^0 := B_1^{-1} - B_*^{-1}$,

$$U_3 := \tfrac{1}{f}(\widehat{R}_A - R_A)\mathrm{vec}\left\{ \Delta^0 \right\} = \tfrac{1}{\mathrm{f}}\textstyle\sum_{\mathrm{k}=1}^{\mathrm{f}}\sum_{\mathrm{j}=1}^{\mathrm{f}} \mathrm{vec}\left\{ \widetilde{\mathrm{S}}_\mathrm{n}^\mathrm{kj} - \mathrm{b}_\mathrm{kj}\mathrm{A}_0 \right\} \Delta^0_\mathrm{jk},$$

and $M_3 := \frac{1}{f}\left(\sum_{k=1}^{f}\sum_{j=1}^{f}\widetilde{S}_n^{kj}\Delta_{jk}^0 - \mathrm{tr}(B_0\Delta^0)A_0\right)$. Define event $\mathcal{E}_1$ as

$$\left\{\left\|\mathrm{diag}(A_0)^{-1/2}\left(\sum_{k=1}^{f}\sum_{j=1}^{f}\Delta_{kj}^0\widetilde{S}_n^{jk}\right)\mathrm{diag}(A_0)^{-1/2} - \langle B_0, \Delta^0\rangle\,\rho(A_0)\right\|_{\max} \le t\right\}$$

where $t = 2\sqrt{f}\left\|B_0\right\|_2\left\|\Delta^0\right\|_2\tau_0$ for τ_0 as defined in Theorem 13.1. By proof of Theorem 13.1, we have $\mathbb{P}\left(\mathcal{E}_1|\mathcal{X}_0\right) \ge 1 - \frac{2}{(m\vee f)^2}$. Thus under event $\mathcal{E}_1 \cap \mathcal{X}_0$,

$$|M_{3,ij}| \;\le\; \sqrt{a_{ii}a_{jj}}t/f = \sqrt{a_{ii}a_{jj}}2\lambda_{f,n}\left\|B_0\right\|_2\left\|\Delta^0\right\|_2$$

where by Corollary 10.1, $\left\|\Delta^0\right\|_2 = \left\|\widehat{B}_*^{-1} - B_*^{-1}\right\|_2 = o\left(1/b_{*,\min}\varphi_{\min}^2(\rho(B_0))\right)$. We write the matrix correspondent of equation (83), under event $\mathcal{X}_0\cap\mathcal{E}_0\cap\mathcal{E}_1$,

$$\begin{aligned}\left|\left(\widetilde{A}(B_1) - A_*\right)_{ij}\right| &\le \left|\left(\widetilde{A}(B_*) - A_*\right)_{ij}\right| + \left|a_{*,ij}\mathrm{tr}(B_1^{-1}\check{B}_0)/f\right| + |M_{3,ij}| \\ &\le \sqrt{a_{*,ii}a_{*,jj}}\lambda_{f,n}(1 + o\left(\left\|B_0\right\|_2/(b_{\min}\varphi_{\min}^2(\rho(B_0)))\right) + |a_{*,ij}|\,\widetilde{\mu}.\end{aligned}$$

The lemma is thus proved. □

It remains to prove Claim 17.3 and Corollary 17.4.

Proof of Claim 17.3. Throughout this proof, we assume that event $\mathcal{X}_0$ holds, and $\lambda_{A_0} > \frac{2\alpha}{1-\alpha}$. Let $\widetilde{B} := \widetilde{B}(I)$ and $\widetilde{W} := \mathrm{diag}(\widetilde{B})^{1/2}$. Thus $\widetilde{W} \succ 0$. By definition, we have

$$\widetilde{B} = \widetilde{W}\widehat{\Gamma}(B_0)\widetilde{W} \text{ and } \quad B_1 = \widetilde{W}\widehat{B}_\rho\widetilde{W}.$$

We have by the KKT conditions,

$$\begin{aligned}\left|\widehat{B}_{\rho,ij} - \widehat{\Gamma}_{ij}(B_0)\right| &\le \lambda_{A_0},\ \ \forall\widehat{B}_{\rho,ij}^{-1} = 0\ \ (\text{ hence } B_{1,ij}^{-1} = 0) \\ \widehat{B}_{\rho,ij} - \widehat{\Gamma}_{ij}(B_0) &= \lambda_{A_0},\ \ \forall\widehat{B}_{\rho,ij}^{-1} > 0\ \ (\text{ hence } B_{1,ij}^{-1} > 0) \\ \text{and }\ \widehat{B}_{\rho,ij} - \widehat{\Gamma}_{ij}(B_0) &= -\lambda_{A_0},\ \ \forall\widehat{B}_{\rho,ij}^{-1} < 0\ \ (\text{ hence } B_{1,ij}^{-1} < 0).\end{aligned}$$

Hence

$$\begin{aligned}\forall i,j\ \ B_{1,ij} - \widetilde{B}_{ij} &= \widetilde{W}_{ii}\left(\widehat{B}_{\rho,ij} - \widehat{\Gamma}_{ij}(B_0)\right)\widetilde{W}_{jj} \\ &= \begin{cases} 0 & \text{if } i = j \\ \widetilde{W}_{ii}\lambda_{A_0}\widetilde{W}_{jj} & \text{if } B_{1,ij}^{-1} > 0 \\ -\widetilde{W}_{ii}\lambda_{A_0}\widetilde{W}_{jj} & \text{if } B_{1,ij}^{-1} < 0 \\ \in [-\widetilde{W}_{ii}\lambda_{A_0}\widetilde{W}_{jj}, \widetilde{W}_{ii}\lambda_{A_0}\widetilde{W}_{jj}] & \text{if } B_{1,ij}^{-1} = 0 \end{cases}.\end{aligned}$$

Thus we have

$$\operatorname{tr}\left((B_1-\widetilde{B})B_1^{-1}\right) = \sum_{i\neq j}\widetilde{W}_{ii}\lambda_{A_0}\widetilde{W}_{jj}\left|B^{-1}_{1,ij}\right|$$
$$= \sum_{i\neq j}\widetilde{W}_{ii}\lambda_{A_0}\widetilde{W}_{jj}\left|\widetilde{W}_{ii}^{-1}\widehat{B}^{-1}_{\rho,ij}\widetilde{W}_{jj}^{-1}\right| = \lambda_{A_0}\left|\widehat{B}_\rho^{-1}\right|_{1,\mathrm{off}}.$$

The claim is proved once we show that

$$\left|\operatorname{tr}\left((\widetilde{B}-B_*)B_1^{-1}\right)\right| \leq \left|\widehat{B}_\rho^{-1}\right|_1 \frac{\alpha}{1-\alpha}. \tag{84}$$

Indeed, for $\check{B}_0 = B_1 - B_* = (B_1-\widetilde{B}) + (\widetilde{B}-B_*)$, we have

$$\lambda_{A_0}\left|\widehat{B}_\rho^{-1}\right|_{1,\mathrm{off}} - \frac{\alpha}{1-\alpha}\left|\widehat{B}_\rho^{-1}\right|_1 \leq \operatorname{tr}(\check{B}_0 B_1^{-1})$$
$$= \operatorname{tr}\left((B_1-\widetilde{B})B_1^{-1}\right) + \operatorname{tr}\left((\widetilde{B}-B_*)B_1^{-1}\right) \leq \lambda_{A_0}\left|\widehat{B}_\rho^{-1}\right|_{1,\mathrm{off}} + \frac{\alpha}{1-\alpha}\left|\widehat{B}_\rho^{-1}\right|_1.$$

It remains to show (84). Before we continue, we need the following inequalities. We have on $\mathcal{X}_0$ by Lemma 4.3 for $\alpha := \alpha_n = \frac{\|A_0\|_F\tau_0}{\operatorname{tr}(A_0)}$,

$$\forall i\neq j,\ \left|\frac{\frac{1}{n}\sum_{t=1}^n\langle y(t)^i, y(t)^j\rangle}{\sqrt{b_{ii}b_{jj}}\operatorname{tr}(A_0)} - \rho_{ij}(B_0)\right| \leq \alpha,$$
$$\text{and}\quad \forall i = 1,\ldots,f\ \left|\frac{\frac{1}{n}\sum_{t=1}^n\left\|y(t)^i\right\|_2^2}{b_{ii}\operatorname{tr}(A_0)} - 1\right| \leq \alpha,$$

and hence, for all i,

$$\frac{1}{1+\alpha} \leq \frac{b_{*,ii}}{\widetilde{W}_{ii}^2} = \frac{b_{ii}\operatorname{tr}(A_0)}{\frac{1}{n}\sum_{t=1}^n\left\|y(t)^i\right\|_2^2} \leq \frac{1}{1-\alpha},$$
$$\text{and}\ \forall i,\ \frac{\alpha}{1+\alpha} \geq 1 - \frac{b_{*,ii}}{\widetilde{W}_{ii}^2} \geq \frac{-\alpha}{1-\alpha}.$$

Thus we have for $\frac{1}{1+\alpha} \leq \sqrt{b_{*,ii}b_{*,jj}}/(\widetilde{W}_{ii}\widetilde{W}_{jj}) \leq \frac{1}{1-\alpha}$,

$$\left|\sum_{i\neq j}\widehat{B}^{-1}_{\rho,ij}\left(\frac{\frac{1}{n}\sum_{t=1}^n\langle y(t)^i, y(t)^j\rangle}{\sqrt{b_{ii}b_{jj}}\operatorname{tr}(A_0)} - \rho_{ij}(B_0)\right)\frac{\sqrt{b_{*,ii}}\sqrt{b_{*,jj}}}{\widetilde{W}_{ii}\widetilde{W}_{jj}}\right| \leq \left|\widehat{B}_\rho^{-1}\right|_{1,\mathrm{off}}\frac{\alpha}{(1-\alpha)}, \tag{85a}$$

$$\text{and}\quad \left|\sum_{i=1}^f\widehat{B}^{-1}_{\rho,ii}\left(1-\frac{b_{*,ii}}{\widetilde{W}_{ii}^2}\right)\right| \leq \frac{\alpha}{1-\alpha}\left|\operatorname{diag}(\widehat{B}_\rho^{-1})\right|_1. \tag{85b}$$

Thus $\operatorname{tr}\left((\widetilde{B}-B_*)B_1^{-1}\right) =$

$$\operatorname{tr}\left(\left(\widetilde{W}\widehat{\Gamma}(B_0)\widetilde{W} - \operatorname{diag}(B_*)^{1/2}\rho(B_0)\operatorname{diag}(B_*)^{1/2}\right)\widetilde{W}^{-1}\widehat{B}_\rho^{-1}\widetilde{W}^{-1}\right)$$

$$= \operatorname{tr}\left(\widehat{\Gamma}(B_0)\widehat{B}_\rho^{-1}\right) - \sum_{i,j=1}^{f}\rho_{ij}(B_0)\widehat{B}_{\rho,ij}^{-1}\left(\frac{\sqrt{b_{*,ii}}}{\widetilde{W}_{ii}}\frac{\sqrt{b_{*,jj}}}{\widetilde{W}_{jj}}\right)$$

$$= \sum_{i\neq j}\widehat{B}_{\rho,ij}^{-1}\left(\widehat{\Gamma}_{ij}(B_0) - \rho_{ij}(B_0)\frac{\sqrt{b_{*,ii}}}{\widetilde{W}_{ii}}\frac{\sqrt{b_{*,jj}}}{\widetilde{W}_{jj}}\right) + \sum_{i=1}^{f}\widehat{B}_{\rho,ii}^{-1}\left(1-\frac{b_{*,ii}}{\widetilde{W}_{ii}^2}\right)$$

$$= \sum_{i\neq j}\widehat{B}_{\rho,ij}^{-1}\left(\frac{\frac{1}{n}\sum_{t=1}^{n}\langle\, y(t)^i, y(t)^j\,\rangle}{\sqrt{b_{ii}b_{jj}}\operatorname{tr}(A_0)} - \rho_{ij}(B_0)\right)\frac{\sqrt{b_{*,ii}}}{\widetilde{W}_{ii}}\frac{\sqrt{b_{*,jj}}}{\widetilde{W}_{jj}} + \sum_{i=1}^{f}\widehat{B}_{\rho,ii}^{-1}\left(1-\frac{b_{*,ii}}{\widetilde{W}_{ii}^2}\right).$$

Clearly (84) holds by taking the absolute values on both sides of the last equation, applying the triangle inequality, and then plugging in the inequalities (85a) and (85b). □

Proof of Corollary 17.4. Throughout this proof, we assume that event $\mathcal{X}_0$ holds. Let $\Delta_{B_0} := \widehat{B}_\rho^{-1} - \rho(B_0)^{-1}$. For $\lambda_{A_0} = 2\alpha_n/\varepsilon(1-\alpha_n)$, where $0<\varepsilon\le 2/3$, we have by Corollary 14.1,

$$\begin{aligned}
|\operatorname{diag}(\Delta_{B_0})|_1 &\le \sqrt{f}\,\|\Delta_{B_0}\|_F \le \sqrt{f}\frac{9}{2}\frac{1+\varepsilon}{\varphi^2_{\min}(\rho(B_0))}\lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}\vee 1} = o\left(\sqrt{f}\right)\\
|\Delta_{B_0}|_{1,\text{off}} &\le \sqrt{\left|B_0^{-1}\right|_{0,\text{off}}}\frac{1+\varepsilon}{1-\varepsilon}(9\lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}\vee 1})/(\varphi^2_{\min}(\rho(B_0)))\\
&\le \sqrt{\left|B_0^{-1}\right|_{0,\text{off}}}\frac{1+\varepsilon}{(1-\varepsilon)\varepsilon}\frac{\alpha_n}{1-\alpha_n}(18\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}\vee 1})/(\varphi^2_{\min}(\rho(B_0)))\\
&= o\left(\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}}\right)
\end{aligned}$$

where $(9\lambda_{A_0}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}\vee 1})/(\varphi^2_{\min}(\rho(B_0))) = o(1)$ by (A1) and (A2), and $\frac{1+\varepsilon}{(1-\varepsilon)\varepsilon}$ is a constant so long as ε is bounded away from 0 and 1. Thus for $\sqrt{\left|B_0^{-1}\right|_{0,\text{off}} + f}/f \le 1$, we have

$$\frac{1}{f}\left|\Delta_{B_0}\right|_1 = o\left(1/\sqrt{f}\right) + o\left(\frac{1}{f}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}}\right) = o\left(\frac{1}{f}\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}+f}\right) = o(1).$$

Thus by the triangle inequality, we have for $\lambda_{A_0} \asymp \alpha_n/(1-\alpha_n)$,

$$\left|\widehat{B}_\rho^{-1}\right|_{1,\text{off}} \le \left|\rho(B_0)^{-1}\right|_{1,\text{off}} + o\left(\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}}\right)$$

$$\left|\widehat{B}_\rho^{-1}\right|_1 \le \left|\rho(B_0)^{-1}\right|_1 + \left|\Delta_{B_0}\right|_1 \le \left|\rho(B_0)^{-1}\right|_1 + o\left(\sqrt{\left|B_0^{-1}\right|_{0,\text{off}}+f}\right)$$

and hence $\widetilde{\mu} \le \dfrac{\alpha_n}{1-\alpha_n}\left|\rho(B_0)^{-1}\right|_1/f + \lambda_{A_0}\left|\rho(B_0)^{-1}\right|_{1,\text{off}}/f + o\left(\lambda_{A_0}\right)$

The corollary thus holds. □

17.3. *Proof of Theorem 6.2.* First we observe that (37) follows from (36) immediately given that $\widetilde{\eta} = \lambda_{f,n}(1+o(1)) + \widetilde{\mu} < \eta$. On $\mathcal{A}_1$, we have by (33),

$$\forall i, \quad \left|\frac{\widetilde{A}(B_1)_{ii}}{a_{*,ii}} - 1\right| \le \widetilde{\eta} \text{ and hence } \frac{\widetilde{A}(B_1)_{ii}^{1/2}}{\sqrt{a_{*,ii}}} \ge \sqrt{1-\widetilde{\eta}}$$

$$\text{and} \quad \forall i \neq j, \quad \left|\frac{\widetilde{A}(B_1)_{ij}}{\sqrt{a_{*,ii}a_{*,jj}}} - \frac{a_{*,ij}}{\sqrt{a_{*,ii}a_{*,jj}}}\right| \le \lambda_{f,n}(1+o(1)) + \frac{|a_{*,ij}|}{\sqrt{a_{*,ii}a_{*,jj}}}\widetilde{\mu}.$$

We have for all i, j, on event $\mathcal{A}_1$,

$$\begin{aligned}
\left|\widehat{\Gamma}_{ij}(A_0) - \rho_{ij}(A_0)\right| &:= \left|\frac{\widetilde{A}_{ij}(B_1)}{\widetilde{A}_{ii}^{1/2}(B_1)\widetilde{A}_{jj}^{1/2}(B_1)} - \rho_{ij}(A_0)\right| \\
&= \left|\frac{\widetilde{A}_{ij}^{1/2}(B_1)/\sqrt{a_{*,ii}a_{*,jj}} - \rho_{ij}(A_0)}{(\widetilde{A}_{ii}^{1/2}(B_1)/\sqrt{a_{*,ii}})(\widetilde{A}_{jj}^{1/2}(B_1)/\sqrt{a_{*,jj}})}\right| \\
&\quad + \left|\frac{\rho_{ij}(A_0)}{(\widetilde{A}_{ii}^{1/2}(B_1)/\sqrt{a_{*,ii}})(\widetilde{A}_{jj}^{1/2}(B_1)/\sqrt{a_{*,jj}})} - \rho_{ij}(A_0)\right| \\
&\le \frac{\lambda_{f,n}(1+o(1)) + \widetilde{\mu}\,|\rho_{ij}(A_0)|}{1-\widetilde{\eta}} + |\rho_{ij}(A_0)|\left|\frac{1}{1-\widetilde{\eta}} - 1\right| \\
&= \frac{\lambda_{f,n}(1+o(1))}{1-\widetilde{\eta}} + |\rho_{ij}(A_0)|\,\frac{\widetilde{\eta}+\widetilde{\mu}}{1-\widetilde{\eta}} \\
&= \frac{\lambda_{f,n}(1+o(1))}{1-\widetilde{\eta}} + |\rho_{ij}(A_0)|\,\frac{\lambda_{f,n}(1+o(1)) + 2\widetilde{\mu}}{1-\widetilde{\eta}} \\
&= \frac{2}{1-\widetilde{\eta}}\left(\lambda_{f,n}(1+o(1)) + |\rho_{ij}(A_0)|\,\widetilde{\mu}\right) \le \frac{2\widetilde{\eta}}{1-\widetilde{\eta}}. \qquad \square
\end{aligned}$$

17.4. *Proof of Theorem 6.4 and Lemma 17.5.* In order to prove Theorem 6.4, we first state the bound for the entry-wise errors for the sample covariance matrix as defined in Step 3 in Lemma 17.5.

LEMMA 17.5. *Suppose all conditions in Theorem 6.4 hold. Let $\widehat{A}_\rho = \widehat{A}_\rho(B_1)$ and A_1 be defined as in Step 2. Suppose we choose λ_{B_1} as in (38). Then on event $\mathcal{A}_1 \cap \mathcal{E}_2$, for $\widetilde{B}(A_1)$ as defined in (29), and η as in (37),*

$$\left|\left(\widetilde{B}(A_1) - B_*\right)_{ij}\right| \le \sqrt{b_{*,ii}b_{*,jj}}\lambda_{m,n}(1+o(1)) + |b_{*,ij}|\,\widetilde{\xi} \tag{86}$$

$$\textit{where} \quad \widetilde{\xi} = \lambda_{B_1}\left|\widehat{A}_\rho^{-1}\right|_{1,\text{off}}/m + \frac{\widetilde{\eta}}{1-\widetilde{\eta}}\left|\widehat{A}_\rho^{-1}\right|_1/m \le \xi$$

where $\widetilde{\eta}$ is as defined in Theorem 6.2 and ξ is defined in (41).

Proof of Theorem 6.4. Given Lemma 17.5, the proof follows exactly the same arguments as Theorem 6.2, and hence is omitted. □

In order to prove Lemma 17.5, we need the following auxiliary results. Let $\widehat{A}_\rho$ be as defined in Lemma 17.5. Let

$$\widetilde{W}_1 = \mathrm{diag}(\widetilde{A}(B_1))^{1/2}, \;\; A_1 = \widetilde{W}_1 \widehat{A}_\rho \widetilde{W}_1, \text{ and } \; \check{A}_1 := A_1 - A_*. \tag{87}$$

Claim 17.6. *Suppose all conditions in Lemma 17.5 hold. Then on $\mathcal{A}_1$,*

$$\lambda_{B_1} \left|\widehat{A}_\rho^{-1}\right|_{1,\mathrm{off}} - \frac{\widetilde{\eta}}{1-\widetilde{\eta}} \left|\widehat{A}_\rho^{-1}\right|_1 \le \mathrm{tr}(\check{A}_1 A_1^{-1}) \le \lambda_{B_1} \left|\widehat{A}_\rho^{-1}\right|_{1,\mathrm{off}} + \frac{\widetilde{\eta}}{1-\widetilde{\eta}} \left|\widehat{A}_\rho^{-1}\right|_1 .$$

Corollary 17.7. *Suppose all conditions in Lemma 17.5 hold. Then on $\mathcal{A}_1$, for ξ as defined in (41),*

$$\frac{\left|\mathrm{tr}(\check{A}_1 A_1^{-1})\right|}{m} \;\; \le \;\; \widetilde{\xi} \le \lambda_{B_1} \frac{\left|\rho(A_0)^{-1}\right|_{1,\mathrm{off}}}{m} + \frac{\widetilde{\eta}}{1-\widetilde{\eta}} \frac{\left|\rho(A_0)^{-1}\right|_1}{m} + o\left(\lambda_{B_1}\right) \le \xi$$

Proof of Lemma 17.5. Throughout this proof, we assume that (A1) (A2), and (A3) hold. Let

$$\begin{aligned} \widehat{R}_B &= \left[\mathrm{vec}\left\{ \widehat{\mathrm{S}}_{\mathrm{n}}^{11} \right\} \ldots \mathrm{vec}\left\{ \widehat{\mathrm{S}}_{\mathrm{n}}^{1\mathrm{m}} \right\} \ldots \mathrm{vec}\left\{ \widehat{\mathrm{S}}_{\mathrm{n}}^{\mathrm{mm}} \right\}\right], \\ R_B &= [a_{11}\mathrm{vec}\{\mathrm{B}_0\} \ldots \mathrm{a}_{1\mathrm{m}}\mathrm{vec}\{\mathrm{B}_0\} \ldots \mathrm{a}_{\mathrm{mm}}\mathrm{vec}\{\mathrm{B}_0\}] = \mathrm{vec}\{\mathrm{B}_0\} \otimes (\mathrm{vec}\{\mathrm{A}_0^{\mathrm{T}}\})^{\mathrm{T}}. \end{aligned}$$

For $\check{A}_1$ as in (87), we have

$$\Delta^1 = A_1^{-1} - A_*^{-1} = -A_1^{-1}(A_1 - A_*)A_*^{-1} = -A_1^{-1}\check{A}_1 A_*^{-1}. \tag{88}$$

First we write

$$\begin{aligned} & \tfrac{1}{m}\mathrm{vec}\left\{ \widetilde{\mathrm{B}}(\mathrm{A}_1) \right\} - \tfrac{1}{\mathrm{m}}\mathrm{R_B}\mathrm{vec}\{\mathrm{A}_*^{-1}\} = \tfrac{1}{\mathrm{m}}\widehat{\mathrm{R}}_\mathrm{B}\mathrm{vec}\{\mathrm{A}_1^{-1}\} - \mathrm{vec}\{\mathrm{B}_*\} \\ &= \tfrac{1}{m}(\widehat{R}_B - R_B)\mathrm{vec}\{\mathrm{A}_*^{-1}\} + \tfrac{1}{\mathrm{m}}\mathrm{R_B}\mathrm{vec}\{\Delta^1\} + \tfrac{1}{\mathrm{m}}(\widehat{\mathrm{R}}_\mathrm{B} - \mathrm{R_B})\mathrm{vec}\{\Delta^1\} \\ &:= V_1 + V_2 + V_3, \end{aligned} \tag{89}$$

where the matrix correspondent of each summand will be denoted by M_1, M_2, and M_3 respectively. Now for the first summand on the RHS, we have

$$\begin{aligned} V_1 &= \tfrac{1}{m}(\widehat{R}_B - R_B)\mathrm{vec}\{\mathrm{A}_*^{-1}\} = \mathrm{vec}\left\{ \widetilde{\mathrm{B}}(\mathrm{A}_*) - \mathrm{B}_* \right\} \\ &= \tfrac{1}{f}\sum_{k=1}^f \sum_{j=1}^f \mathrm{vec}\left\{ \widehat{\mathrm{S}}_{\mathrm{n}}^{\mathrm{kj}} - \mathrm{a}_{\mathrm{kj}}\mathrm{B}_0 \right\} (\mathrm{A}_*^{-1})_{\mathrm{jk}} \end{aligned}$$

and $M_1 = \widetilde{B}(A_*) - B_* = \frac{1}{m}\sum_{k=1}^{f}\sum_{\ell=1}^{f}\widehat{S}_n^{\ell k}(A_*^{-1})_{k\ell} - B_*$.

By Lemma 17.1, we have on event $\mathcal{E}_0$,

$$|M_{1,ij}| = \left|\widetilde{B}_{ij}(A_*) - B_{*,ij}\right| \quad\leq\quad \sqrt{b_{*,ii}b_{*,jj}}\lambda_{m,n}.$$

We now examine the second summand on the RHS of (89), where recall that $\check{A}_1 = A_1 - A_*$. Now by (88),

$$\begin{aligned} V_2 &= \tfrac{1}{m}R_B \mathrm{vec}\left\{-\mathrm{A}_1^{-1}\check{\mathrm{A}}_1\mathrm{A}_*^{-1}\right\} = \tfrac{1}{\mathrm{m}}\mathrm{vec}\left\{\mathrm{B}_0\right\}\mathrm{vec}\left\{\mathrm{A}_0^{\mathrm{T}}\right\}^{\mathrm{T}}(\mathrm{A}_*^{-\mathrm{T}}\otimes\mathrm{A}_1^{-1})\mathrm{vec}\left\{-\check{\mathrm{A}}_1\right\} \\ &= \tfrac{1}{m}\mathrm{vec}\left\{\mathrm{B}_0\right\}\mathrm{tr}(-\mathrm{A}_0\mathrm{A}_1^{-1}\check{\mathrm{A}}_1\mathrm{A}_*^{-1}) = \tfrac{1}{\mathrm{m}}\mathrm{vec}\left\{\mathrm{B}_*\right\}\mathrm{tr}(-\mathrm{A}_1^{-1}\check{\mathrm{A}}_1). \end{aligned}$$

Hence $M_2 = B_*\mathrm{tr}(-A_1^{-1}\check{A}_1)/m$. From Claim 17.6, we have on event $\mathcal{A}_1$,

$$|M_{2,ij}| = \frac{|b_{*,ij}|}{m}\left|\mathrm{tr}(-\check{A}_1A_1^{-1})\right| \leq \frac{|b_{*,ij}|}{m}\left(\lambda_{B_1}\left|\widehat{A}_\rho^{-1}\right|_{1,\mathrm{off}} + \frac{\widetilde{\eta}}{1-\widetilde{\eta}}\left|\widehat{A}_\rho^{-1}\right|_1\right) = |b_{*,ij}|\,\widetilde{\xi}.$$

Finally, we bound the third summand. Let Δ^1 be as in (88). By definition,

$$V_3 \quad=\quad \tfrac{1}{m}(\widehat{R}_B - R_B)\mathrm{vec}\left\{\Delta^1\right\} = \tfrac{1}{\mathrm{m}}\sum_{\mathrm{k}=1}^{\mathrm{m}}\sum_{\mathrm{j}=1}^{\mathrm{m}}\mathrm{vec}\left\{\widehat{\mathrm{S}}_{\mathrm{n}}^{\mathrm{kj}} - \mathrm{a}_{\mathrm{kj}}\mathrm{B}_0\right\}\Delta_{\mathrm{jk}}^1$$

and $M_3 = \frac{1}{m}\sum_{k=1}^{m}\sum_{j=1}^{m}\widehat{S}_n^{kj}\Delta_{jk}^1 - \frac{\mathrm{tr}(A_0\Delta^1)}{m}B_0$. Define event $\mathcal{E}_2$ as

$$\left\{\left\|\mathrm{diag}(B_0)^{-1/2}\left(\sum_{k=1}^{m}\sum_{j=1}^{m}\Delta_{kj}^1\widehat{S}_n^{jk}\right)\mathrm{diag}(B_0)^{-1/2} - \mathrm{tr}(A_0\Delta^1)\rho(B_0)\right\|_{\max} \leq t\right\}$$

where $t = 2\tau_0\sqrt{m}\left\|A_0\right\|_2\left\|\Delta^1\right\|_2$. Then under event A_1, we have by proof of Theorem 13.1, $\mathbb{P}(\mathcal{E}_2|\mathcal{A}_1) \geq 1 - \frac{2}{(m\vee f)^2}$. Under $\mathcal{E}_2\cap\mathcal{A}_1$,

$$|M_{3,ij}| \quad\leq\quad \sqrt{b_{ii}b_{jj}}t/m \asymp \sqrt{b_{ii}b_{jj}}\lambda_{mn}\left\|A_0\right\|_2\left\|\Delta^1\right\|_2$$

where by Corollary 17.2, on event $\mathcal{A}_1$

$$\left\|\Delta^1\right\|_2 \quad=\quad \left\|A_1^{-1} - A_*^{-1}\right\|_2 \leq \frac{2C\lambda_{B_1}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}}\vee 1}}{a_{*,\min}\varphi_{\min}^2(\rho(A_0))} = o\left(\frac{1}{a_{*,\min}\varphi_{\min}^2(\rho(A_0))}\right)$$

where for $0 < \varepsilon_1 < 1$,

$$\lambda_{B_1} = \frac{2\widetilde{\eta}}{\varepsilon_1(1-\widetilde{\eta})} \asymp \lambda_{f,n} + \lambda_{m,n} = O(\lambda_{B_0}), \;\text{ and hence }\; \lambda_{B_1}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}}} = o(1).$$

The lemma is thus proved by plugging in the bounds on $|M_{1,ij}|$, $|M_{2,ij}|$, and $|M_{3,ij}|$. On $\mathcal{A}_1 \cap \mathcal{E}_2$, we have

$$\begin{aligned}\left|\left(\widetilde{B}(A_1) - B_*\right)_{ij}\right| &\leq \left|\left(\widetilde{B}(A_*) - B_*\right)_{ij}\right| + \left|b_{*,ij}\frac{\operatorname{tr}(\check{A}_1 A_*^{-1})}{m}\right| + |M_{3,ij}| \\ &\leq \sqrt{b_{*,ii}b_{*,jj}}\lambda_{m,n}\left(1 + o\left(\|A_0\|_2 / (a_{\min}\varphi^2_{\min}(\rho(A_0)))\right)\right) + |b_{*,ij}|\,\widetilde{\xi}\end{aligned}$$

by (89), where $\widetilde{\xi}$ is bounded by ξ as shown in Corollary 17.7. □

Proof of Claim 17.6. Throughout this proof, we assume that event $\mathcal{A}_1$ holds. On $\mathcal{A}_1$, we have by (33),

$$\forall i \neq j, \quad \left|\frac{\widetilde{A}_{ij}(B_1)}{\sqrt{a_{*,ii}a_{*,jj}}} - \rho_{ij}(A_0)\right| \leq \lambda_{f,n}(1 + o(1)) + |\rho_{ij}(A_0)|\,\widetilde{\mu}, \tag{90a}$$

$$\text{and } \frac{1}{\sqrt{1+\widetilde{\eta}}} \leq \frac{\sqrt{a_{*,ii}}}{\widetilde{W}_{1,ii}} \leq \frac{1}{\sqrt{1-\widetilde{\eta}}}, \text{ where } \widetilde{W}_1^2 = \operatorname{diag}(\widetilde{A}(B_1)) \tag{90b}$$

as $\left|\widetilde{W}^2_{1,ii}/a_{*,ii} - 1\right| \leq \widetilde{\eta} = \lambda_{f,n}(1 + o(1)) + \widetilde{\mu} < \eta$. By the KKT conditions we obtain for $A_1 = \widetilde{W}_1\widehat{A}_\rho\widetilde{W}_1$ and $\widetilde{A}(B_1) = \widetilde{W}_1\widehat{\Gamma}(A_0)\widetilde{W}_1$, where $\widetilde{W}_1 \succ 0$,

$$\begin{aligned}\left|\widehat{A}_{\rho,ij} - \widehat{\Gamma}_{ij}(A_0)\right| &\leq \lambda_{B_1}, \ \forall \widehat{A}^{-1}_{\rho,ij} = 0 \ (\text{ hence } A^{-1}_{1,ij} = 0) \\ \widehat{A}_{\rho,ij} - \widehat{\Gamma}_{ij}(A_0) &= \lambda_{B_1}, \ \forall \widehat{A}^{-1}_{\rho,ij} > 0 \ (\text{ hence } A^{-1}_{1,ij} > 0) \\ \text{and } \widehat{A}_{\rho,ij} - \widehat{\Gamma}_{ij}(A_0) &= -\lambda_{B_1} \ \forall \widehat{A}^{-1}_{\rho,ij} < 0 \ (\text{ hence } A^{-1}_{1,ij} < 0),\end{aligned}$$

and hence for all i, j,

$$\begin{aligned}A_{1,ij} - \widetilde{A}_{ij}(B_1) &= \widetilde{W}_{1,ii}\left(\widehat{A}_{\rho,ij} - \widehat{\Gamma}_{ij}(A_0)\right)\widetilde{W}_{1,jj} \\ &= \begin{cases} 0 & \text{if } i = j \\ \widetilde{W}_{1,ii}\lambda_{B_1}\widetilde{W}_{1,jj} & \text{if } A^{-1}_{1,ij} > 0 \\ -\widetilde{W}_{1,ii}\lambda_{B_1}\widetilde{W}_{1,jj} & \text{if } A^{-1}_{1,ij} < 0 \\ \in [-\widetilde{W}_{1,ii}\lambda_{B_1}\widetilde{W}_{1,jj}, \widetilde{W}_{1,ii}\lambda_{B_1}\widetilde{W}_{1,jj}] & \text{if } A^{-1}_{1,ij} = 0 \end{cases}.\end{aligned}$$

Thus we have

$$\operatorname{tr}\left((A_1 - \widetilde{A}(B_1))A_1^{-1}\right) = \sum_{i \neq j}\widetilde{W}_{1,ii}\lambda_{B_1}\widetilde{W}_{1,jj}\left|\widetilde{W}^{-1}_{1,ii}\widehat{A}^{-1}_{\rho,ij}\widetilde{W}^{-1}_{1,jj}\right| = \lambda_{B_1}\left|\widehat{A}^{-1}_\rho\right|_{1,\text{off}}.$$

The claim is proved if we show that

$$\left|\operatorname{tr}\left((\widetilde{A}(B_1) - A_*)A_1^{-1}\right)\right| \leq \frac{\widetilde{\eta}}{1-\widetilde{\eta}}\left|\widehat{A}^{-1}_\rho\right|_1 \tag{91}$$

so that for $\check{A}_1 = A_1 - A_*$, we have

$$\begin{aligned}
\lambda_{B_1} \left|\widehat{A}_\rho^{-1}\right|_{1,\text{off}} /m - \frac{\widetilde{\eta}}{1-\widetilde{\eta}} \left|\widehat{A}_\rho^{-1}\right|_1 /m &\le \operatorname{tr}(\check{A}_1 A_1^{-1})/m \\
&= \operatorname{tr}\left((A_1 - \widetilde{A}(B_1))A_1^{-1}\right)/m + \operatorname{tr}\left((\widetilde{A}(B_1) - A_*)A_1^{-1}\right)/m \\
&\le \lambda_{B_1} \left|\widehat{A}_\rho^{-1}\right|_{1,\text{off}} /m + \frac{\widetilde{\eta}}{1-\widetilde{\eta}} \left|\widehat{A}_\rho^{-1}\right|_1 /m.
\end{aligned}$$

Now to show (91), one follows the same line of arguments as in the proof of Claim 17.3, given (90a) and (90b). □

Proof of Corollary 17.7. Throughout this proof, we assume that event $\mathcal{A}_1$ holds. Let $\Delta_{A_1} = \widehat{A}_\rho^{-1} - \rho(A_0)^{-1}$. We have

$$|\text{diag}(\Delta_{A_1})|_1 \le \sqrt{m} \left\|\Delta_{A_1}\right\|_F \le \sqrt{m} 9(1+\varepsilon_1)\lambda_{B_1} \frac{\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}}{2\varphi^2_{\min}(\rho(A_0))} = o(\sqrt{m})$$

where $\lambda_{B_1}\sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1} = o(1)$ by (A1), given that $\lambda_{B_1} \asymp \frac{\eta}{1-\eta} = O(\lambda_{B_0})$. For the non-diagonal part of Δ_{A_1}, we have by Corollary 14.1, for $\lambda_{B_1} = 2\widetilde{\eta}/\varepsilon_1(1-\widetilde{\eta})$ where $0 < \varepsilon_1 < 1$,

$$\begin{aligned}
|\Delta_{A_1}|_{1,\text{off}} &\le \sqrt{|A_0^{-1}|_{0,\text{off}}} \frac{1+\varepsilon_1}{1-\varepsilon_1} 9\lambda_{B_1} \sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}/(\varphi^2_{\min}(\rho(A_0))) \\
&\le \sqrt{|A_0^{-1}|_{0,\text{off}}} \frac{1+\varepsilon_1}{(1-\varepsilon_1)\varepsilon_1} \frac{18\widetilde{\eta}}{1-\widetilde{\eta}} \sqrt{|A_0^{-1}|_{0,\text{off}} \vee 1}/(\varphi^2_{\min}(\rho(A_0))) \\
&= o\left(\sqrt{|A_0^{-1}|_{0,\text{off}}}\right)
\end{aligned}$$

where $\frac{1+\varepsilon_1}{(1-\varepsilon_1)\varepsilon_1}$ is bounded so long as ε_1 is bounded away from 0 and 1. Hence

$$\begin{aligned}
\tfrac{1}{m}\left|\Delta_{A_1}\right|_1 &= o(1/\sqrt{m}) + o\left(\tfrac{1}{m}\sqrt{|A_0^{-1}|_{0,\text{off}}}\right) = o\left(\tfrac{1}{m}\sqrt{|A_0^{-1}|_{0,\text{off}} + m}\right) = o(1) \\
\left|\widehat{A}_\rho^{-1}\right|_{1,\text{off}} /m &\le \left|\rho(A_0)^{-1}\right|_{1,\text{off}} /m + o\left(\tfrac{1}{m}\sqrt{|A_0^{-1}|_{0,\text{off}}}\right) = \left|\rho(A_0)^{-1}\right|_{1,\text{off}} /m + o(1) \\
\left|\widehat{A}_\rho^{-1}\right|_1 /m &\le \left|\rho(A_0)^{-1}\right|_1 /m + \left|\Delta_{A_1}\right|_1 /m = \left|\rho(A_0)^{-1}\right|_1 /m + o(1)
\end{aligned}$$

where clearly $\frac{\sqrt{|A_0^{-1}|_{0,\text{off}}+m}}{m} \le 1$. We now insert the inequalities above in $\widetilde{\xi}$,

$$\begin{aligned}
\widetilde{\xi} = \lambda_{B_1} \tfrac{1}{m}\left|\widehat{A}_\rho^{-1}\right|_{1,\text{off}} + \tfrac{\widetilde{\eta}}{1-\widetilde{\eta}} \tfrac{1}{m} \left|\widehat{A}_\rho^{-1}\right|_1 &\le \lambda_{B_1} \left(\tfrac{1}{m}\left|\rho(A_0)^{-1}\right|_{1,\text{off}} + o\left(\tfrac{1}{m}\sqrt{|A_0^{-1}|_{0,\text{off}}}\right)\right) \\
&\quad + \frac{\widetilde{\eta}}{1-\widetilde{\eta}} \left(\left|\rho(A_0)^{-1}\right|_1 /m + o\left(\tfrac{1}{m}\sqrt{|A_0^{-1}|_{0,\text{off}} + m}\right)\right) \\
&\le \frac{\widetilde{\eta}}{1-\widetilde{\eta}} \left|\rho(A_0)^{-1}\right|_1 /m + \lambda_{B_1} \left|\rho(A_0)^{-1}\right|_{1,\text{off}} /m + o\left(\lambda_{B_1}\right).
\end{aligned}$$

The other bounds follow from the fact that $\widetilde{\eta} \le \eta$. $\square$

17.5. *Proof of Corollary 17.2*. Throughout this proof, we assume that event $\mathcal{A}_1$ holds. Let $\widehat{A}_\rho = \widehat{A}_\rho(B_1)$ where $\widehat{A}_\rho(B_1)$ is obtained as in Step 2.

Clearly event $\mathcal{T}(A_0)$ holds for sample correlation matrix $\widehat{\Gamma}(A_0)$ for $\delta_{n,f} = \frac{2\widetilde{\eta}}{1-\widetilde{\eta}} \le \frac{2\eta}{1-\eta} \asymp \lambda_{f,n}$ on event $\mathcal{A}_1$, where $\delta_{f,n}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1} = o(1)$ by (A1). We have by Theorem 4.5, on event $\mathcal{A}_1$,

$$\left\|\widehat{A}_\rho^{-1} - \rho(A_0)^{-1}\right\|_F \le 9(1+\varepsilon)\lambda_{B_1}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}/(2\varphi^2_{\min}(\rho(A_0))), \tag{92}$$

$$\text{and} \quad \left\|\widehat{A}_\rho - \rho(A_0)\right\|_F \le 9(1+\varepsilon)\kappa(\rho(A_0))^2\lambda_{B_1}\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}. \tag{93}$$

Let $\widetilde{W}_1 = \text{diag}(\widetilde{A}(B_1))^{1/2}$ and $W = \text{diag}(A_*)^{1/2} = \text{diag}(\sqrt{a_{*,11}}, \ldots \sqrt{a_{*,mm}}) = \sqrt{\frac{m}{\text{tr}(A_0)}}\text{diag}(A_0)^{1/2}$. We have for all i, by (33), $\left|\widetilde{W}^2_{1,ii} - a_{*,ii}\right| \le a_{*,ii}\widetilde{\eta}$. Then the following holds on $\mathcal{A}_1$,

$$\left\|\widetilde{W}_1 - W\right\|_2 \le (\sqrt{1+\widetilde{\eta}} - 1) \vee (1 - \sqrt{1-\widetilde{\eta}})a^{1/2}_{*,\max} \le \sqrt{a_{*,\max}}\widetilde{\eta}, \tag{94}$$

$$\left\|\widetilde{W}_1^{-1} - W^{-1}\right\|_2 \le \left(\frac{\sqrt{1+\widetilde{\eta}}-1}{\sqrt{1+\widetilde{\eta}}} \vee \frac{1-\sqrt{1-\widetilde{\eta}}}{\sqrt{1-\widetilde{\eta}}}\right)\frac{1}{\sqrt{a_{*,\min}}}$$

$$\le \frac{\widetilde{\eta}}{\sqrt{1-\widetilde{\eta}}}\frac{1}{\sqrt{a_{*,\min}}}. \tag{95}$$

By Proposition 15.2, (93), and (94), and for $\widetilde{\eta} < \lambda_{B_1}(1-\widetilde{\eta})/2$, $\eta < 1/4$, and $18 > C > 9$,

$$\begin{aligned}
\left\|\widehat{A}_* - A_*\right\|_2 &= \left\|\widetilde{W}_1\widehat{A}_\rho\widetilde{W}_1 - \text{diag}(A_*)^{1/2}\rho(A_0)\text{diag}(A_*)^{1/2}\right\|_2 \\
&\le (\widetilde{\eta}+2)\widetilde{\eta}a_{*,\max}\left\|\rho(A_0)\right\|_2 + C\lambda_{B_1}\kappa\left(\rho(A_0)\right)^2\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}a_{*,\max}(1+\widetilde{\eta})^2 \\
&\le \lambda_{B_1}\frac{(1-\widetilde{\eta})(\widetilde{\eta}+2)}{2}a_{*,\max}\left\|\rho(A_0)\right\|_2 + C\lambda_{B_1}\kappa\left(\rho(A_0)\right)^2\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}a_{*,\max}(1+\widetilde{\eta})^2 \\
&\le 2C\lambda_{B_1}a_{*,\max}\kappa\left(\rho(A_0)\right)^2\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1} \quad \text{and} \\
\left\|\widehat{A}_* - A_*\right\|_F &\le \lambda_{B_1}\sqrt{m}a_{*,\max}\left\|\rho(A_0)\right\|_2 + C\lambda_{B_1}a_{*,\max}\kappa\left(\rho(A_0)\right)^2\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee 1}(1+\widetilde{\eta})^2 \\
&\le 2Ca_{*,\max}\lambda_{B_1}\kappa\left(\rho(A_0)\right)^2\sqrt{\left|A_0^{-1}\right|_{0,\text{off}} \vee m}.
\end{aligned}$$

Similarly, by Proposition 15.2 (92), and (95), we have for $\widetilde{\eta} \le \eta < 1/4$, and

$9 > C > 9/2$,

$$\left\|\Delta^1\right\|_2 := \left\|\widehat{A}_*^{-1} - A_*^{-1}\right\|_2 = \left\|\widetilde{W}_1^{-1}\widehat{A}_\rho^{-1}\widetilde{W}_1^{-1} - \mathrm{diag}(A_*)^{-1/2}\rho(A_0)^{-1}\mathrm{diag}(A_*)^{-1/2}\right\|_2$$

$$\leq \frac{\widetilde{\eta}}{1-\widetilde{\eta}}\left(\widetilde{\eta} + 2\sqrt{1-\widetilde{\eta}}\right)/(\varphi_{\min}(\rho(A_0))a_{*,\min}) + \frac{C\lambda_{B_1}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee 1}}{\varphi_{\min}^2(\rho(A_0))a_{*,\min}}\left(1 + \frac{\widetilde{\eta}}{\sqrt{1-\widetilde{\eta}}}\right)^2$$

$$\leq 2C\lambda_{B_1}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee 1}/(\varphi_{\min}^2(\rho(A_0))a_{*,\min}) \quad \text{and}$$

$$\left\|\Delta^1\right\|_F := \left\|\widehat{A}_*^{-1} - A_*^{-1}\right\|_F \leq 2C\lambda_{B_1}\sqrt{\left|A_0^{-1}\right|_{0,\mathrm{off}} \vee m}/(\varphi_{\min}^2(\rho(A_0))a_{*,\min}). \qquad \square$$

REMARK 17.8. *Finally, we mention that for the $\widehat{B}_*$ which we obtain in Step 3, we achieve the same convergence bounds as in Corollary 10.1, except that we replace λ_{A_0} with λ_{A_1}, which will be chosen to dominate the maximum entry-wise error: assuming that $\zeta < 1/3$ and $\eta < 1/4$,*

$$\left\|\widehat{\Gamma}(B_0) - \rho(B_0)\right\|_{\max} \leq 3\lambda_{m,n} + 4\max_{i\neq j}|\rho_{ij}(B_0)|\, C_m\left(\lambda_{f,n} + C_f C_A \lambda_{m,n}/(1-\alpha)\right).$$

REFERENCES

[1] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106** 594-607.

[2] RAVIKUMAR, P., WAINWRIGHT, M., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics* **4** 935–980.

[3] ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse Permutation Invariant Covariance Estimation. *Electronic Journal of Statistics* **2** 494-515.

[4] RUDELSON, M. and VERSHYNIN, R. (2013). Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability* **18** 1–9.

[5] ZHOU, S. (2012). Gemini: Graph estimation with matrix variate normal instances Technical Report, Technical report 531, Department of Statistics, University of Michigan.

DEPARTMENT OF STATISTICS,
UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109
E-MAIL: shuhengz@umich.edu