SUPPLEMENTARY MATERIALS FOR ROCKET: ROBUST CONFIDENCE INTERVALS VIA KENDALL'S TAU FOR TRANSELLIPTICAL GRAPHICAL MODELS

BY RINA FOYGEL BARBER* AND MLADEN KOLAR†

University of Chicago

APPENDIX A: GAUSSIAN VECTORS AND THE SIGN-SUBGAUSSIAN PROPERTY

In this section we prove Lemma 4.5, which shows that that a centered Gaussian vector $Z \sim N(0, \Sigma)$ satisfies the sign-subgaussianity property, that is, the sign vector $\text{sign}(Z)$ is itself subgaussian.

LEMMA 4.5. *Let $Z \sim N(0, \Sigma)$ for some $\Sigma \in \mathbb{R}^{p\times p}$. Then $\text{sign}(Z)$ is $\mathsf{C}(\Sigma)$-subgaussian.*

PROOF OF LEMMA 4.5. Without loss of generality, rescale so that $\lambda_{\min}(\Sigma) = 1$ and then $\mathsf{C}(\Sigma) = \lambda_{\max}(\Sigma)$. Write $\Sigma = AA^\top + \mathbf{I}_p$ for some matrix $A \in \mathbb{R}^{n\times n}$. Then we can write $Z = X + AY$, where $X, Y \overset{iid}{\sim} N(0, \mathbf{I}_p)$. Then, for any fixed vector $v \in \mathbb{R}^p$,

$$\mathbb{E}\left[e^{v^\top \text{sign}(Z)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{v^\top \text{sign}(Z)} \mid Y\right]\right] = \mathbb{E}\left[\mathbb{E}\left[e^{v^\top \text{sign}(X+AY)} \mid Y\right]\right]$$
$$= \mathbb{E}\left[\prod_i \mathbb{E}\left[e^{v_i \,\text{sign}(X_i+(AY)_i)} \mid Y\right]\right],$$

where the last step holds because, conditional on Y, each of the terms $\text{sign}(X_i + (AY)_i)$ depends on X_i only, and therefore these terms are conditionally independent. Next, observe that

$$\mathbb{E}\left[\text{sign}(X_i + (AY)_i) \mid Y\right] = \mathbb{E}\left[\text{sign}\left(N(0,1) + (AY)_i\right) \mid Y\right]$$
$$= \Phi\left((AY)_i\right) - \Phi\left(-(AY)_i\right) = \psi\left((AY)_i\right),$$

*Department of Statistics, University of Chicago
†Booth School of Business, University of Chicago

where we define $\psi(z) = \Phi(z) - \Phi(-z)$ for $z \in \mathbb{R}$. Then, for each i,

$$\mathbb{E}\left[e^{v_i \operatorname{sign}(X_i + (AY)_i)} \mid Y\right] = \mathbb{E}\left[e^{v_i (\operatorname{sign}(X_i + (AY)_i) - \psi((AY)_i))} \mid Y\right] \cdot e^{v_i \psi((AY)_i)} \leqslant e^{v_i^2/2} \cdot e^{v_i \psi((AY)_i)} ,$$

where the inequality is proved by applying Hoeffding's Lemma (see, for example, Massart (2007, Lemma 2.6)) to the bounded mean-zero random variable $[v_i (\operatorname{sign}(X_i + (AY)_i) - \psi((AY)_i))]$. Combining the calculations so far, we have

$$\mathbb{E}\left[e^{v^\top \operatorname{sign}(Z)}\right] \leqslant e^{||v||_2^2/2} \cdot \mathbb{E}\left[e^{v^\top \psi(AY)}\right] ,$$

where $\psi(AY)$ applies the function $\psi(\cdot)$ elementwise to the vector AY.

Next we show that $y \mapsto v^\top \psi(Ay)$ is Lipschitz over $y \in \mathbb{R}^n$. Note that $z \mapsto \psi(z)$ is 1-Lipschitz over $z \in \mathbb{R}$ since the density of the standard normal distribution is bounded uniformly as $\phi(z) \leqslant \frac{1}{\sqrt{2\pi}} \leqslant \frac{1}{2}$. For any $y, y' \in \mathbb{R}^n$, we have

$$\left|v^\top \psi(Ay) - v^\top \psi(Ay')\right| \leqslant \sum_i |v_i| \cdot \left|\psi((Ay)_i) - \psi\left((Ay')_i\right)\right| \leqslant \sum_i |v_i| \cdot \left|(Ay)_i - (Ay')_i\right| \leqslant ||v||_2 \cdot ||A(y-y')||_2 \leqslant ||v||_2 \cdot \sqrt{\lambda_{\max}(\Sigma) - 1} \cdot ||y - y'||_2 ,$$

where the last step is true because

$$||A||_{\mathsf{op}} = \sqrt{||\Sigma - \mathbf{I}_p||_{\mathsf{op}}} = \sqrt{\mathsf{C}(\Sigma) - 1} .$$

Therefore, $y \mapsto v^\top \psi(Ay)$ is $\left(||v||_2 \cdot \sqrt{\mathsf{C}(\Sigma) - 1}\right)$-Lipschitz in y. Furthermore, $\psi(z) = -\psi(-z)$ for all $z \in \mathbb{R}$, and so for any $y \in \mathbb{R}^n$,

$$v^\top \psi(Ay) = -v^\top \psi(A \cdot (-y)) \quad \Rightarrow \quad \mathbb{E}[\psi(AY)] = 0 \text{ since } Y \overset{\mathcal{D}}{=} -Y .$$

We can now apply standard concentration results for Lipschitz functions of a Gaussian: by Massart (2007, Proposition 3.5), $\mathbb{E}\left[e^{v^\top \psi(AY)}\right] \leqslant e^{||v||_2^2 (\mathsf{C}(\Sigma) - 1)/2}$. Therefore,

$$\mathbb{E}\left[e^{v^\top \operatorname{sign}(Z)}\right] \leqslant \mathbb{E}\left[e^{||v||_2^2/2 + v^\top \psi(AY)}\right] \leqslant e^{||v||_2^2/2 + ||v||_2^2(\mathsf{C}(\Sigma) - 1)/2} = e^{||v||_2^2 \cdot \mathsf{C}(\Sigma)/2} .$$

□

APPENDIX B: PROOF OF MAIN RESULT

B.1. Preliminaries. We first compute bounds on $||\gamma_c||_2$ and $||\gamma_c||_1$ for each $c = a, b$, which we will use many times in the proofs below. First, for $c = a, b$ note that

$$||\gamma_c||_2 = ||\Sigma_I^{-1}\Sigma_{Ic}||_2 \leqslant ||\Sigma_I^{-1}|| \cdot ||\Sigma_{Ic}||_2 \leqslant [\lambda_{\min}(\Sigma)]^{-1} \cdot \lambda_{\max}(\Sigma) \leqslant C_{\mathsf{cov}} \tag{B.1}$$

by Assumption 3.1. Next,

$$\begin{aligned} ||\gamma_c||_1 &= ||\Sigma_I^{-1}\Sigma_{Ic}||_1 \\ &= ||-\Omega_{I,ab}\Theta_{ab,c}||_1 \qquad \text{(by matrix blockwise inversion)} \\ &= \sum_{j\in I} |\Omega_{j,ab}\Theta_{ab,c}| \leqslant \sum_{j\in I} ||\Omega_{j,ab}||_1 ||\Theta_{ab,c}||_\infty \\ &\leqslant C_{\mathsf{cov}} \sum_{j\in I} ||\Omega_{j,ab}||_1 \end{aligned} \tag{B.2}$$

since

$$||\Theta||_\infty \leqslant \lambda_{\max}(\Theta) = (\lambda_{\min}(\Omega_{ab,ab}))^{-1} \leqslant (\lambda_{\min}(\Omega))^{-1} = \lambda_{\max}(\Sigma) \leqslant \mathsf{C}_{\mathsf{cov}}.$$

Therefore,

$$||\gamma_c||_1 \leqslant C_{\mathsf{cov}} \left(||\Omega_a||_1 + ||\Omega_b||_1\right) \leqslant 2C_{\mathsf{cov}}C_{\mathsf{sparse}}\sqrt{k_n} \tag{B.3}$$

by applying Assumption 3.2.

B.2. Proof of Theorem 4.1: asymptotic normality of the oracle estimator.

THEOREM 4.1. *Suppose that Assumptions 3.1, 3.2, and 3.4 hold. Then there exist constants $C_{\mathsf{normal}}, C_{\mathsf{variance}}$ depending on $C_{\mathsf{cov}}, C_{\mathsf{sparse}}, C_{\mathsf{kernel}}$ but not on (n, p_n, k_n), such that*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \sqrt{n} \cdot \frac{\widetilde{\Theta}_{ab} - \Theta_{ab}}{S_{ab} \cdot \det(\Theta)} \leqslant t \right\} - \Phi(t) \right| \leqslant C_{\mathsf{normal}} \cdot \frac{k_n \log(p_n)}{\sqrt{n}} + \frac{1}{2p_n},$$

where S_{ab} is defined in the proof and satisfies $S_{ab} \cdot \det(\Theta) \geqslant C_{\mathsf{variance}} > 0$.

PROOF OF THEOREM 4.1. We first show that the error $\widetilde{\Theta}_{ab} - \Theta_{ab}$ can be approximated by a linear function of the Kendall's tau estimator $\widehat{T}$. Define vectors $u, v \in \mathbb{R}^{p_n}$ with entries

$$u_a = 1, u_b = 0, u_I = -\gamma_a \text{ and } v_a = 0, v_b = 1, v_I = -\gamma_b\ .$$

Then by definition, we have $\widetilde{\Theta}_{ab} = u^\top \widehat{\Sigma} v$ and $\Theta_{ab} = u^\top \Sigma v$, that is, the error is given by

$$\widetilde{\Theta}_{ab} - \Theta_{ab} = u^\top (\widehat{\Sigma} - \Sigma) v \ .$$

Next, since $\widehat{\Sigma} = \sin\left(\frac{\pi}{2}\widehat{T}\right)$ and $\Sigma = \sin\left(\frac{\pi}{2}T\right)$, we take a second-order Taylor expansion of $\sin(\cdot)$ to see that, for some $t \in [0,1]$,

$$\text{(B.4)} \quad \widetilde{\Theta}_{ab} - \Theta_{ab} = u^\top \Bigg[\frac{\pi}{2} \cos\left(\frac{\pi}{2}T\right) \circ (\widehat{T} - T) - \frac{1}{2} \cdot \left(\frac{\pi}{2}\right)^2 \cdot \sin\left(\frac{\pi}{2}(t \cdot T + (1-t) \cdot \widehat{T})\right) \circ (\widehat{T} - T) \circ (\widehat{T} - T) \Bigg] v \ .$$

Next, we rewrite this linear term. We have

$$L := u^\top \left[\cos\left(\frac{\pi}{2}T\right) \circ \widehat{T}\right] v = \frac{1}{\binom{n}{2}} \sum_{i<i'} \mathrm{sign}(X_i - X_{i'})^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \mathrm{sign}(X_i - X_{i'}) \ ,$$

which is a U-statistic of order 2 with respect to the data $(X_1, \dots, X_n)$. Note that

$$L - \mathbb{E}[L] = u^\top \left[\frac{\pi}{2}\cos\left(\frac{\pi}{2}T\right) \circ \widehat{T}\right] v - u^\top \left[\frac{\pi}{2}\cos\left(\frac{\pi}{2}T\right) \circ \mathbb{E}[\widehat{T}]\right] v = u^\top \left[\frac{\pi}{2}\cos\left(\frac{\pi}{2}T\right) \circ (\widehat{T} - T)\right] v \ .$$

Define the kernel $g(X, X') = \mathrm{sign}(X - X')^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \mathrm{sign}(X - X')$, and let $g_1(X) = \mathbb{E}\left[g(X, X') \mid X\right]$, where $X, X' \overset{iid}{\sim} \mathsf{TE}(\Sigma, \xi; f_1, \dots, f_p)$. Let $\nu_{g_1}^2 = \mathsf{Var}(g_1(X))$ and $\eta_g^3 = \mathbb{E}\left[|g(X, X')|^3\right]$. By Callaert and Janssen (1978, Section 2), we have

$$\text{(B.5)} \qquad \sup_{t \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\sqrt{n}(L - \mathbb{E}[L])}{2\nu_{g_1}} \leqslant t \right\} - \Phi(t) \right| \leqslant C_{\mathsf{Ustat}} \cdot \frac{\eta_g^3}{\nu_{g_1}^3} \cdot \frac{1}{\sqrt{n}} \ ,$$

for a universal constant C_{Ustat}. Next we bound the ratio $\frac{\eta_g^3}{\nu_{g_1}^3}$ in the following lemma, which is proved in Appendix D.8.

LEMMA B.1. *Suppose that Assumptions 3.1, 3.2 and 3.4 hold. Let $g(X, X')$ and $g_1(X)$ be defined as in the proof of Theorem 4.1. Then*

$$\nu_{g_1}^2 = \mathsf{Var}(g_1(X)) \geqslant \frac{1}{\pi^2} C_{\mathsf{variance}}^2$$

and

$$\nu_{g_1}^3 \leqslant \eta_g^3 = \mathbb{E}\left[|g(X, X')|^3\right] \leqslant C_{\mathsf{moment}}$$

where $C_{\mathsf{variance}}, C_{\mathsf{moment}}$ are constants depending only on $C_{\mathsf{cov}}, C_{\mathsf{kernel}}$ and not on (n, p_n, k_n).

In particular, this lemma implies that $S_{ab} := \pi\nu_{g_1}(\det(\Theta))^{-1} \geqslant C_{\mathsf{variance}} \cdot (\det(\Theta))^{-1}$.

Finally, the linear term L analysed here provides only an approximation to $\widetilde{\Theta}_{ab} - \Theta_{ab}$. Define

$$\Delta = \widetilde{\Theta}_{ab} - \Theta_{ab} - \frac{\pi}{2}(L - \mathbb{E}[L]) \ .$$

Then we have

$$\begin{aligned}
|\Delta| &= \left| u^\top \left[\frac{1}{2} \cdot \left(\frac{\pi}{2}\right)^2 \cdot \sin\left(\frac{\pi}{2}(t \cdot T + (1-t) \cdot \widehat{T})\right) \circ (\widehat{T} - T) \circ (\widehat{T} - T) \right] v \right| \\
&\leqslant ||u||_1 ||v||_1 \left\| \frac{1}{2} \cdot \left(\frac{\pi}{2}\right)^2 \cdot \sin\left(\frac{\pi}{2}(t \cdot T + (1-t) \cdot \widehat{T})\right) \circ (\widehat{T} - T) \circ (\widehat{T} - T) \right\|_\infty \\
&\leqslant \frac{\pi^2}{8} ||u||_1 ||v||_1 ||\widehat{T} - T||_\infty^2
\end{aligned}$$

(B.6)
$$\leqslant \frac{\pi^2}{8} \cdot k_n \cdot (1 + 2C_{\mathsf{cov}} C_{\mathsf{sparse}})^2 \cdot ||\widehat{T} - T||_\infty^2 \ ,$$

where the last inequality holds by (B.3).

Finally, the next lemma is proved in de la Pena and Giné (1999).

LEMMA B.2 ((de la Pena and Giné, 1999, Theorem 4.1.8)). *For any $\delta > 0$, with probability at least $1 - \delta$,*

$$||\widehat{T} - T||_\infty \leqslant \sqrt{\frac{4 \log\left(2\binom{p_n}{2}/\delta\right)}{n}} \ .$$

Applying this lemma with $\delta = \frac{1}{2p_n}$, we have $||\widehat{T} - T||_\infty^2 \leqslant \frac{4\log(2p_n^3)}{n} \leqslant \frac{16\log(p_n)}{n}$ with probability at least $1 - \frac{1}{2p_n}$.

To summarize the computations so far, we have $\widetilde{\Theta}_{ab} - \Theta_{ab} = \frac{\pi}{2}(L - \mathbb{E}[L]) + \Delta$, where (B.5) gives an asymptotic normality result for the linear term $(L - \mathbb{E}[L])$, while (B.6) gives a bound on Δ. To prove therefore that $\widetilde{\Theta}_{ab} - \Theta_{ab}$ is asymptotically normal, we will use the following lemma (proved in Appendix D.1):

LEMMA B.3. *Let A, B, C be random variables such that*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}\{A \leqslant t\} - \Phi(t)| \leqslant \epsilon_A \quad \text{and} \quad \mathbb{P}\{|B| \leqslant \delta_B, |C| \leqslant \delta_C\} \geqslant 1 - \epsilon_{BC} \ ,$$

where $\epsilon_A, \epsilon_{BC}, \delta_B, \delta_C \in (0,1)$. *Then the variable* $(A+B)\cdot(1+C)^{-1}$ *converges to a standard normal distribution with rate*

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left\{(A+B)\cdot(1+C)^{-1} \leqslant t\right\} - \Phi(t)\right| \leqslant \delta_B + \frac{\delta_C}{1-\delta_C} + \epsilon_A + \epsilon_{BC} \ .$$

We apply this lemma with $A = \frac{\pi}{2}\cdot\sqrt{n}\cdot\frac{L-\mathbb{E}[L]}{S_{ab}\cdot\det(\Theta)}$ and $B = \sqrt{n}\cdot\frac{\Delta}{S_{ab}\cdot\det(\Theta)}$ and $C = 0$. We have

$$\sup_{t\in\mathbb{R}} |\mathbb{P}\{A \leqslant t\} - \Phi(t)| \leqslant C_{\mathsf{Ustat}} \cdot \frac{C_{\mathsf{moment}}}{\left(\frac{1}{\pi^2}C^2_{\mathsf{variance}}\right)^{1.5}} \cdot \frac{1}{\sqrt{n}} =: \epsilon_A$$

by (B.5) and Lemma B.1. Furthermore,

$$\mathbb{P}\left\{|B| \leqslant \sqrt{n}\cdot\frac{\frac{\pi^2}{8}\cdot k_n\cdot(1+2C_{\mathsf{cov}}C_{\mathsf{sparse}})^2\cdot\frac{16\log(p_n)}{n}}{C_{\mathsf{variance}}} =: \delta_B\right\}$$
$$\leqslant \mathbb{P}\left\{||\widehat{T}-T||^2_\infty \leqslant \frac{16\log(p_n)}{n}\right\} \geqslant 1 - \frac{1}{2p_n} =: 1-\epsilon_{BC}$$

by (B.6) and Lemmas B.1 and B.2. Noting that $\sqrt{n}\cdot\frac{\widetilde{\Theta}_{ab}-\Theta_{ab}}{S_{ab}} = A+B$, and defining

$$C_{\mathsf{normal}} = \frac{2\pi^2(1+2C_{\mathsf{cov}}C_{\mathsf{sparse}})^2}{C_{\mathsf{variance}}} + C_{\mathsf{Ustat}}\cdot\frac{C_{\mathsf{moment}}}{\left(\frac{1}{\pi^2}C^2_{\mathsf{variance}}\right)^{1.5}} \ ,$$

we have proved the desired result.

□

B.3. Proof of Theorem 4.2: gap between the estimator and the oracle estimator, and estimation of the variance.

THEOREM 4.2. *Suppose that Assumptions 3.1, 3.2, and 3.3 hold. Then there exists a constant* C_{oracle}, *depending on* $C_{\mathsf{cov}}, C_{\mathsf{sparse}}, C_{\mathsf{est}}$ *but not on* (n, p_n, k_n), *such that, if*[1] $n \geqslant 15k_n\log(p_n)$, *then, with probability at least* $1 - \frac{1}{2p_n} - \delta_n$,

$$||\check{\Theta} - \widetilde{\Theta}||_\infty \leqslant C_{\mathsf{oracle}} \cdot \frac{k_n\log(p_n)}{n}$$

and

$$\left|\check{S}_{ab}\cdot\det(\check{\Theta}) - S_{ab}\cdot\det(\Theta)\right| \leqslant C_{\mathsf{oracle}}\cdot\sqrt{\frac{k_n^2\log(p_n)}{n}} \ .$$

[1]Note that the additional condition $n \geqslant 15k_n\log(p_n)$ can be assumed to hold in our main result Theorem 3.5, since if this inequality does not hold, then the claim in Theorem 3.5 is trivial.

The first part of Theorem 4.2, which bounds the distance between our estimator $\breve{\Theta}$ of Θ and the oracle estimator $\widetilde{\Theta}$, is established using bounds on $\widehat{\Sigma} - \Sigma$ in Section 4.3. Details are given in Appendix B.3.1. The second part of Theorem 4.2, which bounds the error in estimating variance, $\left|\breve{S}_{ab} \cdot \det(\breve{\Theta}) - S_{ab} \cdot \det(\Theta)\right|$, is treated in Appendix B.3.2.

B.3.1. *Bounds on* $\breve{\Theta} - \widetilde{\Theta}$. We use our bounds on the covariance error, $\widehat{\Sigma} - \Sigma$, to derive a bound on the difference between our empirical estimator $\breve{\Theta}$ and the oracle estimator $\widetilde{\Theta}$ of Θ. The bounds we give here are deterministic (given that our initial assumptions hold).

The following lemma is proved in Appendix D.7:

LEMMA B.4. *If Assumptions 3.1, 3.2, and 3.3 hold, then with probability at least* $1 - \delta_n$,

$$||\breve{\Theta} - \widetilde{\Theta}||_\infty \leqslant C_{\mathsf{submatrix}} \left(\frac{k_n \log(p_n)}{n} + ||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_{k_n}} \cdot \sqrt{\frac{k_n \log(p_n)}{n}} \right),$$

where $C_{\mathsf{submatrix}}$ *is a constant depending on* C_{cov}, C_{est}, *and* C_{sparse} *but not on* (n, p_n, k_n).

From this point on, we combine Corollary 4.8 and Lemma B.4 to obtain our probabilistic bound on $||\breve{\Theta} - \widetilde{\Theta}||_\infty$ (Theorem 4.2). Looking first at Corollary 4.8, and setting $\delta_1 = \delta_2 = \frac{1}{6p_n}$, we see that by the assumption $p_n \geqslant 2, k_n \geqslant 1$ and the assumption $n \geqslant 15 k_n \log(p_n)$ stated in Theorem 4.2, the conditions of Corollary 4.8 must hold. Then, with probability at least $1 - \delta_1 - \delta_2 = 1 - \frac{1}{3p_n}$,

$$\text{(B.7)} \quad ||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_{k_n}} \leqslant \frac{\pi^2}{8} \cdot k_n \cdot \frac{4 \log\left(12 p_n \binom{p_n}{2}\right)}{n} + 2\pi \cdot 16(1 + \sqrt{5}) C_{\mathsf{cov}} \cdot \sqrt{\frac{\log(12p_n) + 2k_n \log(12p_n)}{n}} \leqslant C_{\mathsf{cov}} \cdot C' \cdot \sqrt{\frac{k_n \log(p_n)}{n}},$$

where we choose the universal constant $C' = 3\pi^2 + 2\pi \cdot 16(1 + \sqrt{5})\sqrt{15}$ which guarantees that the last inequality holds (using the assumptions $n \geqslant k_n \log(p_n)$, $p_n \geqslant 2$, and $k_n \geqslant 1$).

Now combining this result with Lemma B.4, we obtain

$$||\breve{\Theta} - \widetilde{\Theta}||_\infty \leqslant \frac{k_n \log(p_n)}{n} \cdot C_{\mathsf{submatrix}} \left(1 + C_{\mathsf{cov}} \cdot C'\right).$$

Taking $C_{\mathsf{oracle}} \geqslant C_{\mathsf{submatrix}}(1 + C_{\mathsf{cov}} \cdot C')$, we have proved that the first bound in Theorem 4.2 holds with probability at least $1 - \frac{1}{3p_n} - \delta_n$.

B.3.2. *Variance estimate.* For the second part of the theorem, that is, bounding the error in the variance estimate $\breve{S}_{ab}$, we state this bound as a lemma and defer the proof to Appendix D.11, since we need to develop some additional technical results before treating this bound.

LEMMA B.5. *Under the assumptions and definitions of Theorem 4.2, with probability at least $1 - \frac{1}{6p_n}$, if $n \geqslant k_n^2 \log(p_n)$, on the event that the bounds (3.1) in Assumption 3.3 hold,*

$$\left|\breve{S}_{ab} \cdot \det(\breve{\Theta}) - S_{ab} \cdot \det(\Theta)\right| \leqslant C_{\mathsf{oracle}} \cdot \sqrt{\frac{k_n^2 \log(p_n)}{n}} .$$

Combining this lemma with the work above, and using Assumption 3.3, we have proved that both bounds stated in Theorem 4.2 hold with probability at least $1 - \frac{1}{2p_n} - \delta_n$, as desired.

B.4. Proof of Theorem 3.5: main result. We now prove our main result, Theorem 3.5.

PROOF OF THEOREM 3.5. Recall that our goal is to prove that $\frac{\sqrt{n}(\breve{\Omega}_{ab} - \Omega_{ab})}{\breve{S}_{ab}}$ converges to the $N(0,1)$ distribution. Recalling that $\Theta = (\Omega_{ab,ab})^{-1}$ and using the formula for a 2×2 matrix inverse, we separate this random variable into several terms:

$$\begin{aligned}
&\frac{\sqrt{n}(\breve{\Omega}_{ab} - \Omega_{ab})}{\breve{S}_{ab}} \\
&= \frac{\sqrt{n}\left(\frac{-\breve{\Theta}_{ab}}{\det(\breve{\Theta})} - \frac{-\Theta_{ab}}{\det(\Theta)}\right)}{\breve{S}_{ab}} = \frac{\sqrt{n}\left(-\breve{\Theta}_{ab} + \Theta_{ab} \cdot \frac{\det(\breve{\Theta})}{\det(\Theta)}\right)}{\breve{S}_{ab} \cdot \det(\breve{\Theta})} \\
&= \frac{\sqrt{n}\left(\Theta_{ab} - \widetilde{\Theta}_{ab} + \widetilde{\Theta}_{ab} - \breve{\Theta}_{ab} - \Theta_{ab} \cdot \left(1 - \frac{\det(\breve{\Theta})}{\det(\Theta)}\right)\right)}{\breve{S}_{ab} \cdot \det(\breve{\Theta})} \\
&= \left[-\frac{\sqrt{n}\left(\widetilde{\Theta}_{ab} - \Theta_{ab}\right)}{S_{ab} \cdot \det(\Theta)} + \frac{\sqrt{n}\left(\widetilde{\Theta}_{ab} - \breve{\Theta}_{ab}\right)}{S_{ab} \cdot \det(\Theta)} + \frac{\sqrt{n} \cdot \Omega_{ab} \cdot \left(\det(\Theta) - \det(\breve{\Theta})\right)}{S_{ab} \cdot \det(\Theta)}\right] \\
&\qquad \times \left[1 + \frac{\breve{S}_{ab} \cdot \det(\breve{\Theta}) - S_{ab} \cdot \det(\Theta)}{S_{ab} \cdot \det(\Theta)}\right]^{-1} .
\end{aligned}$$

To show that $\frac{\sqrt{n}(\check{\Omega}_{ab}-\Omega_{ab})}{\check{S}_{ab}}$ converges to the standard normal distribution, we will can apply Lemma B.3 (stated in Appendix B.2). In order to apply this lemma and obtain the desired result, we assemble the following pieces:

First, the variable $A := -\frac{\sqrt{n}(\widetilde{\Theta}_{ab}-\Theta_{ab})}{S_{ab}\cdot\det(\Theta)}$ satisfies $\sup_{t\in\mathbb{R}} |\mathbb{P}\{A \leqslant t\} - \Phi(t)| \leqslant C_{\mathsf{normal}} \cdot \frac{k_n \log(p_n)}{\sqrt{n}} + \frac{1}{2p_n} =: \epsilon_A$, as shown in Theorem 4.1.

Second, we define variables $B := \frac{\sqrt{n}(\widetilde{\Theta}_{ab}-\check{\Theta}_{ab})}{S_{ab}\cdot\det(\Theta)} + \frac{\sqrt{n}\cdot\Omega_{ab}\cdot(\det(\Theta)-\det(\check{\Theta}))}{S_{ab}\cdot\det(\Theta)}$ and $C := \frac{\check{S}_{ab}\cdot\det(\check{\Theta})-S_{ab}\cdot\det(\Theta)}{S_{ab}\cdot\det(\Theta)}$, and set

$$\delta_B = \frac{k_n \log(p_n)}{\sqrt{n}} \cdot \left(\frac{C_{\mathsf{oracle}} + 4C_{\mathsf{cov}}^2 C_{\mathsf{oracle}} + 2C_{\mathsf{cov}}C_{\mathsf{oracle}}^2}{C_{\mathsf{variance}}}\right)$$

and

$$\delta_C = \frac{C_{\mathsf{oracle}}}{C_{\mathsf{variance}}} \cdot \sqrt{\frac{k_n^2 \log(p_n)}{n}} .$$

We now show that, by Theorem 4.2, with probability at least $1 - \frac{1}{2p_n} - \delta_n =: 1 - \epsilon_{BC}$ it holds that $|B| \leqslant \delta_B$ and $|C| \leqslant \delta_C$. For the variable C, this is a trivial consequence of the bound on $\left|\check{S}_{ab} \cdot \det(\check{\Theta}) - S_{ab} \cdot \det(\Theta)\right|$ in Theorem 4.2 combined with the lower bound $S_{ab} \cdot \det(\Theta) \geqslant C_{\mathsf{variance}}$ from Theorem 4.1.

Now we turn to the bound on B. To prove this bound, observe that $||\check{\Theta} - \widetilde{\Theta}||_\infty \leqslant C_{\mathsf{oracle}} \cdot \frac{k_n \log(p_n)}{n}$ by Theorem 4.2 (with the stated probability). We also have

$$|B| = \left|\frac{\sqrt{n}\left(\widetilde{\Theta}_{ab} - \check{\Theta}_{ab}\right)}{S_{ab} \cdot \det(\Theta)}\right| \leqslant \sqrt{n}\cdot\frac{1}{S_{ab} \cdot \det(\Theta)}\cdot||\check{\Theta}-\widetilde{\Theta}||_\infty \leqslant \frac{\sqrt{n}}{C_{\mathsf{variance}}}\cdot||\check{\Theta}-\widetilde{\Theta}||_\infty ,$$

where the last step follows from Theorem 4.1. And,

$$\begin{aligned}\left|\det(\check{\Theta}) - \det(\Theta)\right| &= \left|\left(\check{\Theta}_{aa}\check{\Theta}_{bb} - \check{\Theta}_{ab}^2\right) - \left(\Theta_{aa}\Theta_{bb} - \Theta_{ab}^2\right)\right| \\ &\leqslant 4||\Theta||_\infty||\check{\Theta} - \Theta||_\infty + 2||\check{\Theta} - \Theta||_\infty^2\end{aligned}$$

and

$$|\Omega_{ab}| \leqslant \lambda_{\max}(\Omega) = (\lambda_{\min}(\Sigma))^{-1} \leqslant C_{\mathsf{cov}} .$$

Therefore,

$$\begin{aligned}\left|\frac{\sqrt{n} \cdot \Omega_{ab} \cdot \left(\det(\Theta) - \det(\check{\Theta})\right)}{S_{ab} \cdot \det(\Theta)}\right| &\leqslant \sqrt{n}\cdot\frac{|\Omega_{ab}|}{S_{ab} \cdot \det(\Theta)}\cdot\left(4||\Theta||_\infty||\check{\Theta} - \Theta||_\infty + 2||\check{\Theta} - \Theta||_\infty^2\right) \\ &\leqslant \sqrt{n} \cdot \frac{C_{\mathsf{cov}}}{C_{\mathsf{variance}}} \cdot \left(4\mathsf{C}_{\mathsf{cov}}||\check{\Theta} - \Theta||_\infty + 2||\check{\Theta} - \Theta||_\infty^2\right) ,\end{aligned}$$

where the last step follows from Theorem 4.1 along with the fact that

$$||\Theta||_\infty \leqslant \lambda_{\max}(\Theta) = (\lambda_{\min}(\Omega_{ab,ab}))^{-1} \leqslant (\lambda_{\min}(\Omega))^{-1} = \lambda_{\max}(\Sigma) \leqslant \mathsf{C}_{\mathsf{cov}} \;.$$

Combining everything, we have

$$\begin{aligned}|B| &\leqslant \sqrt{n}\cdot\frac{1}{C_{\mathsf{variance}}}\cdot||\check{\Theta}-\widetilde{\Theta}||_\infty+\sqrt{n}\cdot\frac{C_{\mathsf{cov}}}{C_{\mathsf{variance}}}\cdot\left(4\mathsf{C}_{\mathsf{cov}}||\check{\Theta}-\Theta||_\infty+2||\check{\Theta}-\Theta||_\infty^2\right)\\ &\leqslant \frac{k_n\log(p_n)}{\sqrt{n}}\left[\frac{C_{\mathsf{oracle}}+4C_{\mathsf{cov}}^2C_{\mathsf{oracle}}+2C_{\mathsf{cov}}C_{\mathsf{oracle}}^2\cdot\frac{k_n\log(p_n)}{n}}{C_{\mathsf{variance}}}\right]\;.\end{aligned}$$

If $n < k_n \log(p_n)$, then the main result in Theorem 3.5 holds trivially. Assuming then that $n \geqslant k_n \log(p_n)$, we have proved the desired bound on $|B|$.

Given these convergence results, we apply Lemma B.3 to obtain the following result:

$$\begin{aligned}\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{n}(\check{\Omega}_{ab}-\Omega_{ab})}{\check{S}_{ab}}\leqslant t\right\}-\Phi(t)\right| &\leqslant \delta_B+\frac{\delta_C}{1-\delta_C}+\epsilon_A+\epsilon_{BC}\\ &= \frac{k_n\log(p_n)}{\sqrt{n}}\cdot\left(\frac{C_{\mathsf{oracle}}+4C_{\mathsf{cov}}^2C_{\mathsf{oracle}}+2C_{\mathsf{cov}}C_{\mathsf{oracle}}^2}{C_{\mathsf{variance}}}\right)+\\ &\quad \frac{\frac{C_{\mathsf{oracle}}}{C_{\mathsf{variance}}}\cdot\sqrt{\frac{k_n^2\log(p_n)}{n}}}{1-\frac{C_{\mathsf{oracle}}}{C_{\mathsf{variance}}}\cdot\sqrt{\frac{k_n^2\log(p_n)}{n}}}+C_{\mathsf{normal}}\cdot\frac{k_n\log(p_n)}{\sqrt{n}}+\frac{1}{2p_n}+\frac{1}{2p_n}+\delta_n\;.\end{aligned}$$

If $\frac{C_{\mathsf{oracle}}}{C_{\mathsf{variance}}}\cdot\sqrt{\frac{k_n^2\log(p_n)}{n}} > \frac{1}{2}$, then the result of Theorem 3.5 holds trivially, and so assuming that this is not the case, we have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{n}(\check{\Omega}_{ab}-\Omega_{ab})}{\check{S}_{ab}}\leqslant t\right\}-\Phi(t)\right| \leqslant C_{\mathsf{converge}}\cdot\sqrt{\frac{k_n^2\log^2(p_n)}{n}}+\frac{1}{p_n}+\delta_n\;,$$

where

$$C_{\mathsf{converge}} := \frac{C_{\mathsf{oracle}}+4C_{\mathsf{cov}}^2C_{\mathsf{oracle}}+2C_{\mathsf{cov}}C_{\mathsf{oracle}}^2}{C_{\mathsf{variance}}}+\frac{2C_{\mathsf{oracle}}}{C_{\mathsf{variance}}}+C_{\mathsf{normal}}\;.$$

□

APPENDIX C: ACCURACY OF THE INITIAL LASSO ESTIMATOR

COROLLARY 3.8. *Suppose that Assumption 3.1 holds. Assume additionally that the columns Ω_a, Ω_b of the true inverse covariance $\Omega = \Sigma^{-1}$ are k_n-sparse. Then there exist constants $C_{\mathsf{sample}}, C_{\mathsf{Lasso}}$, depending on C_{cov} but not on (n, k_n, p_n), such that if $n \geqslant C_{\mathsf{sample}} k_n \log(p_n)$ then, with probability at least $1 - \frac{1}{2p_n}$, any local minimizer $\breve{\gamma}_a$ of the objective function*

$$\frac{1}{2}\gamma^\top \widehat{\Sigma}_I \gamma - \gamma^\top \widehat{\Sigma}_{Ia} + \lambda ||\gamma||_1$$

over the set $\{\gamma \in \mathbb{R}^I : ||\gamma||_1 \leqslant C_{\mathsf{cov}}\sqrt{2k_n}\}$ satisfies

$$||\breve{\gamma}_a - \gamma_a||_2 \leqslant 3\sqrt{2} C_{\mathsf{cov}} \lambda \sqrt{k_n} \text{ and } ||\breve{\gamma}_a - \gamma_a||_1 \leqslant 24 C_{\mathsf{cov}} \lambda \sqrt{k_n} \ .$$

where we choose $\lambda = C_{\mathsf{Lasso}} \cdot \sqrt{\frac{\log(p_n)}{n}}$. The same result holds for estimating γ_b.

PROOF OF COROLLARY 3.8. Define

$$A = \widehat{\Sigma}_I, \ z = \widehat{\Sigma}_{Ia}, \ x^\star = \gamma_a, \ p = p_n - 2, \ k = k_n \ .$$

Now we apply Theorem 3.7 to this sparse recovery problem. In order to do so, we need to check that the conditions (3.3), (3.4), and (3.5) hold, and that γ_a is feasible under the condition $||\gamma||_1 \leqslant R$. Once these conditions are satisfied, the result of Theorem 3.7 can be applied to this setting.

Feasibility of γ_a. Define $R = C_{\mathsf{cov}}\sqrt{2k_n}$. As proved in (B.1), $||\gamma_a||_2 \leqslant C_{\mathsf{cov}}$, and furthermore $||\gamma_a||_0 \leqslant ||\Omega_a||_0 + ||\Omega_b||_0 \leqslant 2k_n$ (this is true because $\gamma_a = -\Omega_{I,ab}\Theta_{ab,a}$ by (B.3)). Therefore, $||\gamma_a||_1 \leqslant C_{\mathsf{cov}}\sqrt{2k_n} = R$.

Condition (3.3) (restricted strong convexity). Now we need to check that the restricted strong convexity conditions (3.3) hold for our matrix $A = \widehat{\Sigma}_I$. We will show that Corollary 4.8 implies that there exists a constant C_{RSC} depending only on C_{cov}, such that if $n \geqslant 16\log(p_n)$, then with probability at least $1 - \frac{1}{8p_n}$, for all $v \in \mathbb{R}^{p_n}$,

$$\left| v^\top \left(\widehat{\Sigma}_I - \Sigma_I \right) v \right| \leqslant \frac{1}{2C_{\mathsf{cov}}} \left(||v||_2^2 + ||v||_1^2 \cdot \frac{C_{\mathsf{RSC}} \log(p_n)}{n} \right) \ .$$

To see this why this holds, set $k = \frac{n}{C_{\mathsf{RSC}} \log(p_n)}$, and apply Lemma 4.9 to obtain

$$\left| v^\top \left(\widehat{\Sigma}_I - \Sigma_I \right) v \right| \leqslant (||v||_2 + ||v||_1/\sqrt{k})^2 ||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_k} \leqslant 2(||v||_2^2 + ||v||_1^2/k) ||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_k}.$$

Then, applying Corollary 4.8 with this value of k and with $\delta_1 = \delta_2 = \frac{1}{16p_n}$, we see that with probability at least $1 - \frac{1}{8p_n}$,

$$||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_k} \leqslant \frac{1}{4C_{\mathsf{cov}}},$$

as long as we set the constant C_{RSC} large enough.

Then, if this event holds, for all $v \in \mathbb{R}^I$ we have

$$v^\top \widehat{\Sigma}_I v \geqslant v^\top \Sigma_I v - \left| v^\top \left(\widehat{\Sigma}_I - \Sigma_I \right) v \right| \geqslant C_{\mathsf{cov}}^{-1} \cdot ||v||_2^2 - \left| v^\top \left(\widehat{\Sigma}_I - \Sigma_I \right) v \right|$$
$$\geqslant \frac{1}{2C_{\mathsf{cov}}} \cdot ||v||_2^2 - \frac{C_{\mathsf{RSC}}}{2C_{\mathsf{cov}}} \cdot \frac{\log(p_n)}{n} \cdot ||v||_1^2 .$$

Therefore, with probability at least $1 - \frac{1}{8p_n}$, the restricted strong convexity condition (3.3) holds with

$$\alpha_1 = \frac{1}{2C_{\mathsf{cov}}} \text{ and } \tau_1 = \frac{C_{\mathsf{RSC}}}{2C_{\mathsf{cov}}} .$$

Condition (3.5) (penalty parameter). Below, we will prove that, with probability at least $1 - \frac{3}{8p_n}$,

(C.1)
$$||Ax^\star - z||_\infty = ||\widehat{\Sigma}_I \gamma_a - \widehat{\Sigma}_{Ia}||_\infty \leqslant \frac{\pi}{2} C_{\mathsf{feasible}} \sqrt{\frac{\log(p_n)}{n}} + \sqrt{\frac{\log(p_n)}{n}} \cdot \left[\frac{1.5\sqrt{3}\pi^2 \sqrt{1 + C_{\mathsf{cov}}^2}}{\sqrt{C_{\mathsf{sample}}}} \right] ,$$

for a constant C_{feasible} depending only on C_{cov}, as long as we set

$$C_{\mathsf{sample}} \geqslant \left[16(1 + \sqrt{5}) C_{\mathsf{cov}} \sqrt{1 + C_{\mathsf{cov}}^2} \right]^2 .$$

Given that this is true, we now require that condition (3.5) holds, that is,

$$\max \left\{ 4||Ax^\star - z||_\infty, 4\alpha_1 \sqrt{\frac{\log(p)}{n}} \right\} \leqslant \lambda \leqslant \frac{\alpha_1}{6R} .$$

Define

$$C_{\mathsf{Lasso}} = \max \left\{ 4 \left[\frac{\pi}{2} C_{\mathsf{feasible}} + \frac{1.5\sqrt{3}\pi^2 \sqrt{1 + C_{\mathsf{cov}}^2}}{\sqrt{C_{\mathsf{sample}}}} \right], \frac{2}{C_{\mathsf{cov}}} \right\} ,$$

Plugging in the bound (C.1), we see that the lower bound on λ is satisfied for $\lambda = C_{\mathsf{Lasso}} \sqrt{\frac{\log(p_n)}{n}}$. To check the upper bound, we only need

$$\lambda = C_{\mathsf{Lasso}} \sqrt{\frac{\log(p_n)}{n}} \leqslant \frac{\alpha_1}{6R} = \frac{1}{2C_{\mathsf{cov}}^2 \sqrt{2k_n}}.$$

Assuming that

$$n \geqslant 8C_{\mathsf{Lasso}}^2 C_{\mathsf{cov}}^4 \cdot k_n \log(p_n) \ , \tag{C.2}$$

then this follows directly. Therefore, (3.5) is satisfied with probability at least $1 - \frac{3}{8p_n}$.

Condition (3.4) *(sample size).* To satisfy (3.4), by plugging in the definitions of R, α_1, and τ_1 above, we see that it is sufficient to require

$$n \geqslant 64C_{\mathsf{cov}}^2 C_{\mathsf{RSC}} \max\{1, 2C_{\mathsf{RSC}}\} \cdot k_n \log(p_n) \ . \tag{C.3}$$

Conclusion. Combining all of our work above, we see that the conditions (3.3), (3.4), and (3.5), and the feasibility of γ_a, are all satisfied with probability at least $1 - \frac{1}{2p_n}$, as long as

$$n \geqslant C_{\mathsf{sample}} k_n \log(p_n)$$

for

$$C_{\mathsf{sample}} := \max\left\{16, \left[16(1+\sqrt{5})C_{\mathsf{cov}}\sqrt{1+C_{\mathsf{cov}}^2}\right]^2, 8C_{\mathsf{Lasso}}^2 C_{\mathsf{cov}}^4, 64C_{\mathsf{cov}}^2 C_{\mathsf{RSC}} \max\{1, 2C_{\mathsf{RSC}}\}\right\} \ .$$

Therefore, applying Theorem 3.7, if these high probability events hold, then then for any $\breve{\gamma}_a$ that is a local minimizer of

$$L(x) = \frac{1}{2}\gamma^\top \widehat{\Sigma}_I \gamma - \gamma^\top \widehat{\Sigma}_{Ia} + \lambda ||\gamma||_1$$

over the set $\{\gamma \in \mathbb{R}^I : ||\gamma||_1 \leqslant 2C_{\mathsf{cov}} C_{\mathsf{sparse}} \sqrt{k_n}\}$, it holds that

$$||\breve{\gamma}_a - \gamma_a||_2 \leqslant \frac{1.5\lambda \cdot \sqrt{2k_n}}{\alpha_1} = 3\sqrt{2} C_{\mathsf{cov}} \lambda \sqrt{k_n} \text{ and } ||\breve{\gamma}_a - \gamma_a||_1 \leqslant \frac{6\lambda \cdot 2k_n}{\alpha_1} = 24 C_{\mathsf{cov}} \lambda \sqrt{k_n} \ .$$

By the same arguments, the same results hold for estimating γ_b.

Proving (C.1). Now we consider the term $||Ax^\star - z||_\infty = ||\widehat{\Sigma}_I \gamma_a - \widehat{\Sigma}_{Ia}||_\infty$. Since $\gamma_a = \Sigma_I^{-1} \Sigma_{Ia}$, we have

$$||\widehat{\Sigma}_I \gamma_a - \widehat{\Sigma}_{Ia}||_\infty = ||(\widehat{\Sigma}_I - \Sigma_I)\gamma_a - (\widehat{\Sigma}_{Ia} - \Sigma_{Ia})||_\infty = ||(\widehat{\Sigma} - \Sigma)u||_\infty \ ,$$

where $u \in \mathbb{R}^{p_n}$ is the fixed vector with

$$u_a = 1, u_b = 0, u_I = -\gamma_a \ .$$

By the Taylor expansion of $\widehat{\Sigma}-\Sigma$ (calculated as in (B.4)), we have

$$||(\widehat{\Sigma}-\Sigma)u||_\infty = \max_j \left|\mathbf{e}_j^\top(\widehat{\Sigma}-\Sigma)u\right|$$
$$\leqslant \frac{\pi}{2}\max_j\left|\mathbf{e}_j^\top\left(\cos\left(\frac{\pi}{2}T\right)\circ(\widehat{T}-T)\right)u\right| + \frac{\pi^2}{8}\left|\mathbf{e}_j^\top\left(\sin\left(\frac{\pi}{2}\overline{T}\right)\circ(\widehat{T}-T)\circ(\widehat{T}-T)\right)u\right|$$
(C.4)
$$\leqslant \frac{\pi}{2}\max_j\left|\mathbf{e}_j^\top\left(\cos\left(\frac{\pi}{2}T\right)\circ(\widehat{T}-T)\right)u\right| + \frac{\pi^2}{8}||u||_1||\widehat{T}-T||_\infty^2 .$$

Next we bound each term in this final expression (C.4) separately. Beginning with the second term, by (B.1), we know that $||u||_1 \leqslant \sqrt{||u||_0}||u||_2 \leqslant \sqrt{1+2k_n}\cdot\sqrt{1+C_{\mathsf{cov}}^2} \leqslant \sqrt{k_n}\cdot\sqrt{3(1+C_{\mathsf{cov}}^2)}$, where to bound $||u||_0$ we use the calculation $||\gamma_a||_0 \leqslant 2k_n$ from before. Furthermore, by Lemma B.2, with probability at least $1-\frac{1}{8p_n}$,

$$||\widehat{T}-T||_\infty \leqslant \sqrt{\frac{12\log(8p_n)}{n}} \leqslant \sqrt{\frac{48\log(p_n)}{n}} ,$$

using $p_n \geqslant 2$. Therefore, the second term in (C.4) is bounded as

(C.5)
$$\frac{\pi^2}{8}||u||_1||\widehat{T}-T||_\infty^2 \leqslant \frac{\pi^2}{8}\sqrt{k_n}\cdot\sqrt{3(1+C_{\mathsf{cov}}^2)}\frac{48\log(p_n)}{n} \leqslant \sqrt{\frac{\log(p_n)}{n}}\cdot\left[\frac{6\sqrt{3}\pi^2\sqrt{1+C_{\mathsf{cov}}^2}}{\sqrt{C_{\mathsf{sample}}}}\right] ,$$

where we use the assumption $n \geqslant C_{\mathsf{sample}}\log(p_n)$.

Next we turn to the first term in (C.4). In order to bound this term, we begin by stating two lemmas (proved in Appendix D.9):

LEMMA C.1. *There exist vectors $a_1, a_2, \dots$ and $b_1, b_2, \dots$ with $||a_r||_\infty, ||b_r||_\infty \leqslant 1$ for all $r \geqslant 1$, and a sequence $t_1, t_2, \dots \geqslant 0$ with $\sum_r t_r = 4$, such that $\cos\left(\frac{\pi}{2}T\right) = \sum_{r\geqslant 1} t_r a_r b_r^\top$.*

LEMMA C.2. *For fixed u, v with $||u||_2, ||v||_2 \leqslant 1$, for any $|t| \leqslant \frac{n}{4(1+\sqrt{5})C_{\mathsf{cov}}}$,*

$$\mathbb{E}\left[\exp\left(t\cdot u^\top(\widehat{T}-T)v\right)\right] \leqslant \exp\left(\frac{\left[4(1+\sqrt{5})\right]^2 t^2\cdot C_{\mathsf{cov}}^2}{n}\right) .$$

By Lemma C.1, we can write

$$\cos\left(\frac{\pi}{2}T\right) = \sum_{r\geqslant 1} t_r\cdot a_r b_r^\top ,$$

where $t_r \geqslant 0$, $\sum_r t_r = 4$, and $||a_r||_\infty, ||b_r||_\infty \leqslant 1$. Then

$$\mathbf{e}_j^\top \left(\cos\left(\frac{\pi}{2}T\right) \circ (\widehat{T} - T)\right) u = \left\langle \cos\left(\frac{\pi}{2}T\right) \circ \mathbf{e}_j u^\top, \widehat{T} - T\right\rangle = \sum_{r\geqslant 1} t_r \cdot (a_r \circ \mathbf{e}_j)^\top (\widehat{T} - T)(b_r \circ u) .$$

Note that

$$||a_r \circ \mathbf{e}_j||_2 \leqslant ||a_r||_\infty \cdot ||\mathbf{e}_j||_2 \leqslant 1$$

and, by (B.1),

$$||b_r \circ u||_2 \leqslant ||b_r||_\infty \cdot ||u||_2 \leqslant \sqrt{1 + C_{\mathsf{cov}}^2} .$$

Then for any $|t| \leqslant \frac{n}{16(1+\sqrt{5})C_{\mathsf{cov}}\sqrt{1+C_{\mathsf{cov}}^2}}$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{t \cdot \mathbf{e}_j^\top \left(\cos\left(\frac{\pi}{2}T\right) \circ (\widehat{T} - T)\right) u\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{\sum_{r\geqslant 1} t_r \left[t \cdot (a_r \circ \mathbf{e}_j)^\top (\widehat{T} - T)(b_r \circ u)\right]\right\}\right] \\
&\leqslant \sum_r \frac{t_r}{4} \mathbb{E}\left[\exp\left\{4\left[t \cdot (a_r \circ \mathbf{e}_j)^\top (\widehat{T} - T)(b_r \circ u)\right]\right\}\right] \quad \text{(by Jensen's inequality)} \\
&\leqslant \sum_r \frac{t_r}{4} \exp\left(\frac{\left[4(1+\sqrt{5})\right]^2 16t^2 \cdot C_{\mathsf{cov}}^2(1 + C_{\mathsf{cov}}^2)}{n}\right) \quad \text{(by Lemma C.2)} \\
&= \exp\left(\frac{\left[4(1+\sqrt{5})\right]^2 16t^2 \cdot C_{\mathsf{cov}}^2(1 + C_{\mathsf{cov}}^2)}{n}\right) .
\end{aligned}$$

Observe that we can set $t = \pm\sqrt{n\log(p_n)}$, which satisfies $|t| \leqslant \frac{n}{16(1+\sqrt{5})C_{\mathsf{cov}}\sqrt{1+C_{\mathsf{cov}}^2}}$ as long as we set $C_{\mathsf{sample}} \geqslant \left[16(1+\sqrt{5})C_{\mathsf{cov}}\sqrt{1 + C_{\mathsf{cov}}^2}\right]^2$, due to the assumption $n \geqslant C_{\mathsf{sample}} \log(p_n)$. Then, we see that for any $C > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{\left|\mathbf{e}_j^\top \left(\cos\left(\frac{\pi}{2}T\right) \circ (\widehat{T} - T)\right) u\right| > C\sqrt{\frac{\log(p_n)}{n}}\right\} \\
&\qquad \leqslant \mathbb{E}\left[e^{\sqrt{n\log(p_n)}\cdot \mathbf{e}_j^\top \left(\cos\left(\frac{\pi}{2}T\right)\circ(\widehat{T}-T)\right)u - \sqrt{n\log(p_n)}\cdot C\sqrt{\frac{\log(p_n)}{n}}}\right] \\
&\qquad + \mathbb{E}\left[e^{-\sqrt{n\log(p_n)}\cdot \mathbf{e}_j^\top \left(\cos\left(\frac{\pi}{2}T\right)\circ(\widehat{T}-T)\right)u - \sqrt{n\log(p_n)}\cdot C\sqrt{\frac{\log(p_n)}{n}}}\right] \\
&\leqslant 2\exp\left(\frac{\left[4(1+\sqrt{5})\right]^2 16(\sqrt{n\log(p_n)})^2 \cdot C_{\mathsf{cov}}^2(1 + C_{\mathsf{cov}}^2)}{n}\right) \cdot \exp\left\{-\sqrt{n\log(p_n)} \cdot C\sqrt{\frac{\log(p_n)}{n}}\right\} \\
&\qquad \leqslant 2p_n^{-\left(C - \left[4(1+\sqrt{5})\right]^2 \cdot 16C_{\mathsf{cov}}^2(1+C_{\mathsf{cov}}^2)\right)} = 2p_n^{-5} \leqslant \frac{1}{4p_n^2} ,
\end{aligned}$$

where we set $C = C_{\textsf{feasible}} := 5 + \left[4(1+\sqrt{5})\right]^2 \cdot 16C_{\textsf{cov}}^2(1+C_{\textsf{cov}}^2)$. Therefore,

$$\text{(C.6)}\quad \mathbb{P}\left\{\max_j \left|\mathbf{e}_j^\top \left(\cos\left(\frac{\pi}{2}T\right)\circ(\widehat{T}-T)\right)u\right| > C_{\textsf{feasible}}\sqrt{\frac{\log(p_n)}{n}}\right\} \leqslant \frac{1}{4p_n}\ .$$

Combining (C.5) and (C.6), and returning to (C.4), we have

$$||\widehat{\Sigma}_I\gamma_a - \widehat{\Sigma}_{Ia}||_\infty \leqslant \frac{\pi}{2}C_{\textsf{feasible}}\sqrt{\frac{\log(p_n)}{n}} + \sqrt{\frac{\log(p_n)}{n}}\cdot\left[\frac{1.5\sqrt{3}\pi^2\sqrt{1+C_{\textsf{cov}}^2}}{\sqrt{C_{\textsf{sample}}}}\right]\ ,$$

with probability at least $1-\frac{3}{8p_n}$. This proves (C.1). □

APPENDIX D: PROOFS OF LEMMAS

D.1. Proof of the normal convergence lemma.

LEMMA B.3. *Let A, B, C be random variables such that*

$$\sup_{t\in\mathbb{R}}|\mathbb{P}\{A\leqslant t\}-\Phi(t)|\leqslant\epsilon_A \quad \textit{and} \quad \mathbb{P}\{|B|\leqslant\delta_B, |C|\leqslant\delta_C\}\geqslant 1-\epsilon_{BC}\ ,$$

where $\epsilon_A,\epsilon_{BC},\delta_B,\delta_C\in[0,1)$. Then the variable $(A+B)\cdot(1+C)^{-1}$ converges to a standard normal distribution with rate

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left\{(A+B)\cdot(1+C)^{-1}\leqslant t\right\}-\Phi(t)\right|\leqslant\delta_B+\frac{\delta_C}{1-\delta_C}+\epsilon_A+\epsilon_{BC}\ .$$

PROOF OF LEMMA B.3. First, define truncated versions of B and C:

$$\widetilde{B}=\text{sign}(B)\cdot\min\{|B|,\delta_B\},\ \widetilde{C}=\text{sign}(C)\cdot\min\{|C|,\delta_C\}\ .$$

Then, for any $t\in\mathbb{R}$,

$$\left|\mathbb{P}\left\{(A+B)\cdot(1+C)^{-1}\leqslant t\right\}-\mathbb{P}\left\{(A+\widetilde{B})\cdot(1+\widetilde{C})^{-1}\leqslant t\right\}\right| \leqslant \mathbb{P}\left\{B\neq\widetilde{B}\text{ or }C\neq\widetilde{C}\right\}\leqslant\epsilon_{BC}\ .$$

Note that $|\widetilde{B}|\leqslant\delta_B$ and $|\widetilde{C}|\leqslant\delta_C$ with probability 1.

Next, fix any $t\geqslant 0$ and suppose that $A\leqslant t(1-\delta_C)-\delta_B$. Then

$$(A+\widetilde{B})\cdot(1+\widetilde{C})^{-1}\leqslant((t(1-\delta_C)-\delta_B)+\delta_B)\cdot(1-\delta_C)^{-1}=t\ ,$$

and so

$$\begin{aligned}
&\mathbb{P}\left\{(A+\widetilde{B})\cdot(1+\widetilde{C})^{-1}\leqslant t\right\}\geqslant \mathbb{P}\left\{A\leqslant t(1-\delta_C)-\delta_B\right\}\\
&\geqslant \Phi\left(t(1-\delta_C)-\delta_B\right)-\epsilon_A\\
&=\Phi(t)-\mathbb{P}\left\{t(1-\delta_C)-\delta_B<N(0,1)<t(1-\delta_C)\right\}-\mathbb{P}\left\{t(1-\delta_C)<N(0,1)<t\right\}-\epsilon_A\\
&\geqslant \Phi(t)-\delta_B-\mathbb{P}\left\{t(1-\delta_C)<N(0,1)<t\right\}-\epsilon_A \quad \left(\text{since the density of } N(0,1) \text{ is} \leqslant \frac{1}{\sqrt{2\pi}}\leqslant 1\right)\\
&\geqslant \Phi(t)-\delta_B-\frac{\delta_C}{1-\delta_C}-\epsilon_A. \quad (\text{by applying Lemma D.1 stated below})
\end{aligned}$$

To prove the reverse bound, suppose that $(A+\widetilde{B})\cdot(1+\widetilde{C})^{-1}\leqslant t$. Then

$$A=(A+\widetilde{B})\cdot(1+\widetilde{C})^{-1}\cdot(1+\widetilde{C})-\widetilde{B}\leqslant t(1+\delta_C)+\delta_B\ .$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\left\{(A+\widetilde{B})\cdot(1+\widetilde{C})^{-1}\leqslant t\right\}\leqslant \mathbb{P}\left\{A\leqslant t(1+\delta_C)+\delta_B\right\}\\
&\leqslant \Phi\left(t(1+\delta_C)+\delta_B\right)+\epsilon_A\\
&=\Phi(t)+\mathbb{P}\left\{t(1+\delta_C)<N(0,1)<t(1+\delta_C)+\delta_B\right\}+\mathbb{P}\left\{t<N(0,1)<t(1+\delta_C)\right\}+\epsilon_A\ .\\
&\leqslant \Phi(t)+\delta_B+\mathbb{P}\left\{t<N(0,1)<t(1+\delta_C)\right\}+\epsilon_A \quad \left(\text{since the density of } N(0,1) \text{ is} \leqslant \frac{1}{\sqrt{2\pi}}\leqslant 1\right)\\
&\leqslant \Phi(t)+\delta_B+\delta_C+\epsilon_A. \quad (\text{by applying Lemma D.1 stated below})
\end{aligned}$$

Therefore, for all $t\geqslant 0$,

$$\left|\mathbb{P}\left\{(A+\widetilde{B})\cdot(1+\widetilde{C})^{-1}\leqslant t\right\}-\Phi(t)\right|\leqslant \delta_B+\frac{\delta_C}{1-\delta_C}+\epsilon_A\ .$$

By identical arguments, we can prove the same for $t\leqslant 0$. □

LEMMA D.1. *For any* $0\leqslant a\leqslant b$,

$$\mathbb{P}\left\{a<N(0,1)<b\right\}\leqslant\left(\frac{b}{a}-1\right)\cdot\frac{1}{\sqrt{2\pi e}}\leqslant\left(\frac{b}{a}-1\right)\ .$$

PROOF.

$$\begin{aligned}
\mathbb{P}\left\{a<N(0,1)<b\right\}&=\int_{t=a}^{b}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,\mathsf{d}t\\
&\leqslant (b-a)\cdot\frac{1}{\sqrt{2\pi}}e^{-a^2/2}\\
&=\left(\frac{b}{a}-1\right)\cdot\frac{1}{\sqrt{2\pi}}\cdot a\cdot e^{-a^2/2}\\
&\leqslant\left(\frac{b}{a}-1\right)\cdot\frac{1}{\sqrt{2\pi e}}\ ,
\end{aligned}$$

where the last step holds because $\sup_{t>0}\{t \cdot e^{-t^2/2}\} = \frac{1}{\sqrt{e}}$. □

D.2. Sign vector of a transelliptical distribution.

LEMMA 4.4. *Let $X, X' \stackrel{iid}{\sim} \mathsf{TE}(\Sigma, \xi; f_1, \dots, f_p)$. Suppose that Σ is positive definite, and that $\xi > 0$ with probability 1. Then $\mathrm{sign}(X - X')$ is equal in distribution to $\mathrm{sign}(Z)$, where $Z \sim N(0, \Sigma)$.*

PROOF OF LEMMA 4.4. First, since the f_j's are strictly monotone, we see that $\mathrm{sign}(X - X')$ has the same distribution regardless of the choice of the f_j's. Therefore it suffices to consider the case that the f_j's are each the identity function, and so $X, X' \stackrel{iid}{\sim} \mathsf{E}(\mathbf{0}, \Sigma, \xi)$, that is, a zero-mean elliptical distribution. In this case, by Lindskog, McNeil and Schmock (2003, Lemma 1), $X - X' \sim \mathsf{E}(\mathbf{0}, \Sigma, \zeta)$ where the distribution of the random variable $\zeta \geqslant 0$ obeys $\varphi_\zeta(t) = \varphi_\xi(t)^2$, where φ_ζ and φ_ξ are the characteristic functions of ζ and ξ, respectively. Note that for two independent copies $\xi_1, \xi_2 \stackrel{iid}{\sim} \xi$, we have $\varphi_{\xi_1+\xi_2} = \varphi_{\xi_1} \cdot \varphi_{\xi_2} = \varphi_\xi^2 = \varphi_\zeta$, and therefore, $\zeta \stackrel{\mathcal{D}}{=} \xi_1 + \xi_2$. Since $\xi > 0$ with probability 1, this proves that $\zeta > 0$ with probability 1.

Next take $Z \sim N(0, \Sigma)$. Then $\frac{\Sigma^{-1/2}Z}{||\Sigma^{-1/2}Z||_2}$ is uniformly distributed on the unit sphere, and so

$$\zeta \cdot \Sigma^{1/2} \cdot \frac{\Sigma^{-1/2}Z}{||\Sigma^{-1/2}Z||_2} \sim \mathsf{E}(\mathbf{0}, \Sigma, \zeta) ,$$

which is the distribution of $X - X'$. Using the fact that $\zeta > 0$ with probability 1, we see that $\mathrm{sign}(X - X')$ is equal in distribution to

$$\mathrm{sign}\left(\zeta \cdot \Sigma^{1/2} \cdot \frac{\Sigma^{-1/2}Z}{||\Sigma^{-1/2}Z||_2}\right) = \mathrm{sign}(\Sigma^{1/2} \cdot \Sigma^{-1/2}Z) = \mathrm{sign}(Z) ,$$

as desired. □

D.3. Proof of Lemma 4.7.

LEMMA 4.7. *The following bound holds deterministically: for any $k \geqslant 1$,*

$$||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_k} \leqslant \frac{\pi^2}{8} \cdot k||\widehat{T} - T||_\infty^2 + 2\pi||\widehat{T} - T||_{\mathcal{S}_k} .$$

PROOF OF LEMMA 4.7. By Taylor's theorem,

$$\widehat{\Sigma} = \Sigma + \frac{\pi}{2}\cos\left(\frac{\pi}{2}T\right) \circ \left(\widehat{T} - T\right) - \frac{\pi^2}{8}\sin\left(\frac{\pi}{2}\bar{T}\right) \circ \left(\widehat{T} - T\right) \circ \left(\widehat{T} - T\right)$$

where $\bar{T}$ has entries $\bar{\tau}_{ab} = (1-t_{ab})\tau_{ab} + t_{ab}\widehat{\tau}_{ab}$, with $t_{ab} \in [0,1]$ for each a, b. Taking any $u, v \in \mathcal{S}_k$, then,

$$\left|u^\top(\widehat{\Sigma}-\Sigma)v\right| \leqslant \frac{\pi}{2}\left|u^\top\left[\cos\left(\frac{\pi}{2}T\right)\circ\left(\widehat{T}-T\right)\right]v\right| + \frac{\pi^2}{8}\left|u^\top\left[\sin\left(\frac{\pi}{2}\bar{T}\right)\circ\left(\widehat{T}-T\right)\circ\left(\widehat{T}-T\right)\right]v\right| .$$

First, to bound the $\sin(\cdot)$ matrix term, note that

$$\begin{aligned}&\left|u^\top\left[\sin\left(\frac{\pi}{2}\bar{T}\right)\circ\left(\widehat{T}-T\right)\circ\left(\widehat{T}-T\right)\right]v\right|\\ &\leqslant ||u||_1||v||_1\left\|\sin\left(\frac{\pi}{2}\bar{T}\right)\circ\left(\widehat{T}-T\right)\circ\left(\widehat{T}-T\right)\right\|_\infty \leqslant ||u||_1||v||_1||\widehat{T}-T||^2_\infty ,\end{aligned}$$

where the last step holds since the $\sin(\cdot)$ function lies in $[-1,1]$. Furthermore, $||u||_1, ||v||_1 \leqslant \sqrt{k}$ for all $u, v \in \mathcal{S}_k$ by definition.

Next, we bound the $\cos(\cdot)$ matrix term. By Lemma C.1, we can express $\cos\left(\frac{\pi}{2}T\right)$ as a convex combination,

$$\cos\left(\frac{\pi}{2}T\right) = \sum_{r\geqslant 1} t_r \cdot a_r b_r^\top ,$$

where $a_r, b_r \in \mathbb{R}^p$ satisfy $||a_r||_\infty, ||b_r||_\infty \leqslant 1$ for all r, and $t_r \geqslant 0$ satisfy $\sum_r t_r = 4$. Furthermore, for $u, v \in \mathcal{S}_k$ and for each r, note that $u \circ a_r, v \circ b_r \in \mathcal{S}_k$ due to the bound on $||a_r||_\infty, ||b_r||_\infty$. Then

$$\begin{aligned}&\left|u^\top\left[\cos\left(\frac{\pi}{2}T\right)\circ\left(\widehat{T}-T\right)\right]v\right| \leqslant \sum_{r\geqslant 1} t_r \left|u^\top\left[a_r b_r^\top \circ \left(\widehat{T}-T\right)\right]v\right|\\ &= \sum_{r\geqslant 1} t_r\left|(u\circ a_r)^\top\left(\widehat{T}-T\right)(v\circ b_r)\right| \leqslant 4\sup_{u',v'\in\mathcal{S}_k}\left|u'^\top(\widehat{T}-T)v'\right| = 4||\widehat{T}-T||_{\mathcal{S}_k} .\end{aligned}$$

Using the definition $||\widehat{\Sigma}-\Sigma||_{\mathcal{S}_k} = \max_{u,v\in\mathcal{S}_k} u^\top(\widehat{\Sigma}-\Sigma)v$, this proves the lemma. ☐

D.4. Proof of Lemma 4.6.

LEMMA 4.6. *Suppose that $k \geqslant 1$ and $\delta \in (0,1)$ satisfy $\log(2/\delta) + 2k\log(12p) \leqslant n$. Then with probability at least $1-\delta$ it holds that*

$$||\widehat{T}-T||_{\mathcal{S}_k} \leqslant 32(1+\sqrt{5})\mathsf{C}(\Sigma)\cdot\sqrt{\frac{\log(2/\delta)+2k\log(12p)}{n}} .$$

PROOF OF LEMMA 4.6. This lemma is a straightforward combination of Lemma C.2 (stated in Appendix C) together with the following result:

LEMMA D.2 (Adapted from Lemma 5.1 and Theorem 5.2 of Baraniuk et al. (2008)). *Let A be a random matrix satisfying*

(D.1)
$$\exp\left\{t \cdot u^\top A u\right\} \leqslant \exp\left\{\frac{c_1 t^2}{n}\right\} \text{ for all } |t| \leqslant c_0 n \text{ and all unit vectors } u \in \mathbb{R}^p$$

for some constants c_0, c_1. Then for any $k \geqslant 1$ and any $\delta \in (0,1)$ satisfying

$$\log(2/\delta) + k\log(12p) \leqslant n c_0^2 c_1 \ ,$$

with probability at least $1-\delta$ it holds that

(D.2)
$$|u^\top A u| \leqslant \sqrt{\frac{16 c_1}{n}\left(\log(2/\delta) + k\log(12p)\right)} \text{ for all } k\text{-sparse unit vectors } u \in \mathbb{R}^p.$$

Combined, Lemma D.2 (applied with $2k$ in place of k, with $c_0 = \frac{1}{4(1+\sqrt{5})C_{\mathsf{cov}}}$, and $c_1 = c_0^{-1}$) and Lemma C.2 immediately yield the bound

$$\sup_{u \in \mathcal{S}_{2k}} \left|u^\top (\widehat{T} - T) u\right| \leqslant 16(1+\sqrt{5})\mathsf{C}(\Sigma) \cdot \sqrt{\frac{\log(2/\delta) + 2k\log(12p)}{n}} \ ,$$

with probability at least $1-\delta$, as long as $\log(2/\delta) + 2k\log(12p) \leqslant n c_0^2 c_1 = n$. Next take any $u, v \in \mathcal{S}_k$. Then $\frac{u+v}{2}, \frac{u-v}{2} \in \mathcal{S}_{2k}$.

$$\begin{aligned} u^\top(\widehat{T} - T)v &= \left(\frac{u+v}{2}\right)^\top (\widehat{T} - T)\left(\frac{u+v}{2}\right) - \left(\frac{u-v}{2}\right)^\top (\widehat{T} - T)\left(\frac{u+v}{2}\right) \\ &\leqslant 16(1+\sqrt{5})\mathsf{C}(\Sigma) \cdot \sqrt{\frac{\log(2/\delta) + 2k\log(12p)}{n}} \ . \end{aligned}$$

This proves Lemma 4.6, as desired. □

We next turn to the proof of Lemma D.2.

PROOF OF LEMMA D.2. (Adapted from Lemma 5.1 and Theorem 5.2 of Baraniuk et al. (2008).) First fix any $S \subset [p]$ with $|S| = k$. Let $\epsilon = \sqrt{\frac{16c_1}{n}\left(\log(2/\delta) + k\log(12p)\right)}$. Following the same arguments as in Baraniuk et al. (2008, Lemma 5.1), we can take a set $\mathcal{U} \subset \mathbb{R}^S$ of unit vectors, with $|\mathcal{U}| \leqslant 12^k$, such that

$$\sup_{\text{unit } u \in \mathbb{R}^S} \left|u^\top A u\right| \leqslant 2 \sup_{\widetilde{u} \in \mathcal{U}} \left|\widetilde{u}^\top A \widetilde{u}\right| \ .$$

Furthermore, for any fixed $\widetilde{u} \in \mathcal{U}$, for any $0 < t \leqslant c_0 n$,

$$\begin{aligned}\mathbb{P}\left\{\widetilde{u}^\top A\widetilde{u} > \epsilon/2\right\} &\leqslant \mathbb{E}\left[t \cdot \widetilde{u}^\top A\widetilde{u} - t \cdot \epsilon/2\right] \\ &\leqslant \exp\left(\frac{c_1 t^2}{n} - t \cdot \epsilon/2\right) \\ &= \exp\left(-\frac{n\epsilon^2}{16c_1}\right) ,\end{aligned}$$

where for the last step we set $t = \frac{n\epsilon}{4c_1} \leqslant c_0 n$. Similarly,

$$\mathbb{P}\left\{\widetilde{u}^\top A\widetilde{u} < -\epsilon/2\right\} \leqslant \exp\left(-\frac{n\epsilon^2}{16c_1}\right) .$$

Therefore,

$$\mathbb{P}\left\{\sup_{\widetilde{u}\in\mathcal{U}} \left|\widetilde{u}^\top A\widetilde{u}\right| > \epsilon/2\right\} \leqslant 2 \cdot 12^k \cdot \exp\left(-\frac{n\epsilon^2}{16c_1}\right) ,$$

and so

$$\mathbb{P}\left\{\sup_{\text{unit } u\in\mathbb{R}^S} \left|u^\top Au\right| > \epsilon\right\} \leqslant 2 \cdot 12^k \cdot \exp\left(-\frac{n\epsilon^2}{16c_1}\right) .$$

Finally, taking all $\binom{p}{k} \leqslant p^k$ choices for S, we see that

$$\mathbb{P}\left\{\sup_{u\in\mathcal{S}_k} \left\{u^\top Au\right\} \leqslant \epsilon\right\} \geqslant 1 - 2(12p)^k \cdot \exp\left(-\frac{n\epsilon^2}{16c_1}\right) = 1 - \delta .$$

□

D.5. Proof of Corollary 4.8.

COROLLARY 4.8. *Take any $\delta_1, \delta_2 \in (0,1)$ and any $k \geqslant 1$ such that $\log(2/\delta_2) + 2k\log(12p) \leqslant n$. Then, with probability at least $1 - \delta_1 - \delta_2$, the following bound on $\widehat{\Sigma} - \Sigma$ holds:*

$$||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_k} \leqslant$$
$$\frac{\pi^2}{8} \cdot k \cdot \frac{4\log\left(2\binom{p}{2}/\delta_1\right)}{n} + 2\pi \cdot 32(1+\sqrt{5})\mathsf{C}(\Sigma) \cdot \sqrt{\frac{\log(2/\delta_2) + 2k\log(12p)}{n}} .$$

Proof of Corollary 4.8. This proof is a straightforward combination of Lemmas B.2, 4.6, and 4.7. We have

$$\begin{aligned}
&||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_k}\\
&\leqslant \frac{\pi^2}{8}\cdot k||\widehat{T}-T||_\infty^2 + 2\pi||\widehat{T}-T||_{\mathcal{S}_k} \quad \text{by Lemma 4.7}\\
&\leqslant \frac{\pi^2}{8}\cdot k\frac{4\log\left(2\binom{p_n}{2}/\delta_1\right)}{n} + 2\pi\cdot 32(1+\sqrt{5})\mathsf{C}(\Sigma)\cdot\sqrt{\frac{\log(2/\delta_2)+2k\log(12p)}{n}},
\end{aligned}$$

where the last step holds by applying Lemma B.2 with $\delta = \delta_1$ and Lemma 4.6 with $\delta = \delta_2$. □

D.6. Proof of Lemma 4.9.

Lemma 4.9 (Based on Proposition 5 of Sun and Zhang (2012)). *For any fixed matrix $M \in \mathbb{R}^{p\times p}$ and vectors $u, v \in \mathbb{R}^p$, and any $k \geqslant 1$,*

$$|u^\top Mv| \leqslant \left(||u||_2 + ||u||_1/\sqrt{k}\right)\cdot\left(||v||_2 + ||v||_1/\sqrt{k}\right)\cdot \sup_{u',v'\in\mathcal{S}_k}|u'^\top Mv'|.$$

Proof of Lemma 4.9. We first introduce two vector norms used in Proposition 5 of Sun and Zhang (2012),

$$||w||_{(2,k)} = \sup_{|S|\leqslant k}||w_S||_2$$

and its dual norm $||w||^*_{(2,k)}$. Note that

$$||w||_{(2,k)} = \sup_{z\in\mathcal{S}_k}|z^\top w|,$$

from the norm definition. We then have

$$|u^\top Mv| \leqslant ||u||^*_{(2,k)}||Mv||_{(2,k)}$$

by definition of the dual norm, and

$$||Mv||_{2,k} = \sup_{w\in\mathcal{S}_k}|w^\top Mv| \leqslant \sup_{w\in\mathcal{S}_k}||v||^*_{(2,k)}||M^\top w||_{(2,k)} = \sup_{w\in\mathcal{S}_k}||v||^*_{(2,k)}\left(\sup_{z\in\mathcal{S}_k}|z^\top M^\top w|\right),$$

and so combining everything,

$$|u^\top Mv| \leqslant ||u||^*_{(2,k)}||v||^*_{(2,k)}\cdot\sup_{w,z\in\mathcal{S}_k}|w^\top Mz| = ||u||^*_{(2,k)}||v||^*_{(2,k)}\cdot||M||_{\mathcal{S}_k}.$$

Finally, Proposition 5(ii) of Sun and Zhang (2012) (applied with $q = 2$ and $m = k$, in their notation), proves that, for any w,

$$||w||^*_{2,k} \leqslant ||w||_{(2,4k)} + ||w||_1/\sqrt{k} \leqslant ||w||_2 + ||w||_1/\sqrt{k},$$

where the last inequality follows trivially from the norm definition. This proves the lemma. □

D.7. Proof of Lemma B.4.

LEMMA B.4. *If Assumptions 3.1, 3.2, and 3.3 hold, then with probability at least* $1 - \delta_n$,

$$||\breve{\Theta} - \widetilde{\Theta}||_\infty \leqslant C_{\mathsf{submatrix}} \left(\frac{k_n \log(p_n)}{n} + ||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_{k_n}} \cdot \sqrt{\frac{k_n \log(p_n)}{n}} \right),$$

where $C_{\mathsf{submatrix}}$ *is a constant depending on* C_{cov}, C_{est}, *and* C_{sparse} *but not on* (n, p_n, k_n).

PROOF OF LEMMA B.4. Choose any $c, d \in \{a, b\}$; we will bound the (c, d)th entry of the error, that is, $\left|\breve{\Theta}_{cd} - \widetilde{\Theta}_{cd}\right|$. Write $\Delta_c = \breve{\gamma}_c - \gamma_c$ for each $c = a, b$. First, with probability at least $1 - \delta_n$, the bounds on Δ_c, Δ_d in Assumption 3.3 all hold; assume that this event occurs from this point on.

We have

$$\begin{aligned}
&\left|\breve{\Theta}_{cd} - \widetilde{\Theta}_{cd}\right| \\
&= \left|\left(\widehat{\Sigma}_{cd} - \breve{\gamma}_c^\top \widehat{\Sigma}_{Id} - \widehat{\Sigma}_{Ic}^\top \breve{\gamma}_d + \breve{\gamma}_c^\top \widehat{\Sigma}_I \breve{\gamma}_d\right) - \left(\widehat{\Sigma}_{cd} - \gamma_c^\top \widehat{\Sigma}_{Id} - \widehat{\Sigma}_{Ic}^\top \gamma_d + \gamma_c^\top \widehat{\Sigma}_I \gamma_d\right)\right| \\
&= \left|\breve{\gamma}_c^\top \widehat{\Sigma}_I \breve{\gamma}_d - \gamma_c^\top \widehat{\Sigma}_I \gamma_d - \Delta_c^\top \widehat{\Sigma}_{Id} - \widehat{\Sigma}_{Ic}^\top \Delta_d\right| \\
&= \left|\Delta_c^\top \widehat{\Sigma}_I \gamma_d + \gamma_c^\top \widehat{\Sigma}_I \Delta_d + \Delta_c^\top \widehat{\Sigma}_I \Delta_d - \Delta_c^\top \widehat{\Sigma}_{Id} - \widehat{\Sigma}_{Ic}^\top \Delta_d\right|
\end{aligned}$$

(D.3)

$$\begin{aligned}
&\leqslant \left|\Delta_c^\top \Sigma_I \gamma_d + \gamma_c^\top \Sigma_I \Delta_d + \Delta_c^\top \Sigma_I \Delta_d - \Delta_c^\top \Sigma_{Id} - \Sigma_{Ic}^\top \Delta_d\right| \\
&\quad + \left|\Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\gamma_d + \gamma_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d + \Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d - \Delta_c^\top (\widehat{\Sigma}_{Id} - \Sigma_{Id}) - (\widehat{\Sigma}_{Ic} - \Sigma_{Ic})^\top \Delta_d\right| .
\end{aligned}$$

Now we bound each of these terms. To bound the first term on the right-hand side of (D.3), we have

$$\begin{aligned}
&\left|\Delta_c^\top \Sigma_I \gamma_d + \gamma_c^\top \Sigma_I \Delta_d + \Delta_c^\top \Sigma_I \Delta_d - \Delta_c^\top \Sigma_{Id} - \Sigma_{Ic}^\top \Delta_d\right| \\
&= \left|\Delta_c^\top \Sigma_I \Sigma_I^{-1} \Sigma_{Id} + \Sigma_{Ic}^\top \Sigma_I^{-1} \Sigma_I \Delta_d + \Delta_c^\top \Sigma_I \Delta_d - \Delta_c^\top \Sigma_{Id} - \Sigma_{Ic}^\top \Delta_d\right| \\
&= \left|\Delta_c^\top \Sigma_I \Delta_d\right| \\
&\leqslant ||\Delta_c||_2 \cdot ||\Delta_d||_2 \cdot ||\Sigma||_{\mathsf{op}} \\
&\leqslant ||\Delta_c||_2 \cdot ||\Delta_d||_2 \cdot \mathsf{C}(\Sigma) \\
&\leqslant C_{\mathsf{cov}} (C_{\mathsf{est}})^2 \frac{k_n \log(p_n)}{n} \ ,
\end{aligned}$$

where the next-to-last step holds because

$$||\Sigma_I|| \leqslant ||\Sigma|| = \lambda_{\min}(\Sigma) \cdot \mathsf{C}(\Sigma) \ ,$$

and we must have $\lambda_{\min}(\Sigma) \leqslant 1$ because $\mathrm{diag}(\Sigma) = \mathbf{1}$, while the last step holds by Assumptions 3.1 and 3.3.

Finally, to bound the second term on the right-hand side of (D.3), we have

$$\begin{aligned}
&\left|\Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\gamma_d + \gamma_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d + \Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d - \Delta_c^\top (\widehat{\Sigma}_{Id} - \Sigma_{Id}) - (\widehat{\Sigma}_{Ic} - \Sigma_{Ic})^\top \Delta_d\right| \\
&\leqslant \left|\Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\gamma_d\right| + \left|\gamma_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d\right| + \left|\Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d\right| + \left|\Delta_c^\top (\widehat{\Sigma}_{Id} - \Sigma_{Id})\right| + \left|(\widehat{\Sigma}_{Ic} - \Sigma_{Ic})^\top \Delta_d\right| \\
&= \left|\Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\gamma_d\right| + \left|\gamma_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d\right| + \left|\Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\Delta_d\right| + \left|\Delta_c^\top (\widehat{\Sigma}_I - \Sigma_I)\mathbf{e}_d\right| + \left|\mathbf{e}_c^\top (\widehat{\Sigma}_I - \Sigma_I)^\top \Delta_d\right| ,
\end{aligned}$$

where $\mathbf{e}_c$ and $\mathbf{e}_d$ are the basis vectors in $\mathbb{R}^I$ corresponding to nodes c and d. Now, for each vector $w \in \{\gamma_c, \gamma_d, \mathbf{e}_c, \mathbf{e}_d\}$ and each vector $z \in \{\Delta_c, \Delta_d\}$, we have

$$||w||_2 + ||w||_1/\sqrt{k} \leqslant C_{\mathsf{cov}} + 2C_{\mathsf{cov}} C_{\mathsf{sparse}} + 2$$

by (B.1) and (B.3), and

$$||z||_2 + ||z||_1/\sqrt{k} \leqslant 2C_{\mathsf{est}} \sqrt{\frac{k_n \log(p_n)}{n}}$$

by Assumption 3.3, and so

$$\left|w^\top (\widehat{\Sigma}_I - \Sigma_I) z\right| \leqslant ||\widehat{\Sigma}_I - \Sigma_I||_{\mathcal{S}_{k_n}} \cdot (C_{\mathsf{cov}} + 2C_{\mathsf{cov}} C_{\mathsf{sparse}} + 2) \cdot 2C_{\mathsf{est}} \sqrt{\frac{k_n \log(p_n)}{n}}$$

by applying Lemma 4.9. Noting that $||\widehat{\Sigma}_I - \Sigma_I||_{\mathcal{S}_{k_n}} \leqslant ||\widehat{\Sigma} - \Sigma||_{\mathcal{S}_{k_n}}$ trivially, the desired result of the lemma follows trivially from these bounds by setting $C_{\mathsf{submatrix}}$ appropriately. □

D.8. Proof of Lemma B.1.

LEMMA B.1. *Suppose that Assumptions 3.1, 3.2 and 3.4 hold. Let $g(X, X')$ and $g_1(X)$ be defined as in the proof of Theorem 4.1. Then*

$$\nu_{g_1}^2 = \mathsf{Var}(g_1(X)) \geqslant \frac{1}{\pi^2} C_{\mathsf{variance}}^2$$

and

$$\nu_{g_1}^3 \leqslant \eta_g^3 = \mathbb{E}\left[|g(X, X')|^3\right] \leqslant C_{\mathsf{moment}}$$

where $C_{\mathsf{variance}}, C_{\mathsf{moment}}$ are constants depending only on $C_{\mathsf{cov}}, C_{\mathsf{kernel}}$ and not on (n, p_n, k_n).

PROOF OF LEMMA B.1. First, we have

$$\begin{aligned} g_1(X) =& \mathbb{E}\left[\text{sign}(X - X')^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \text{sign}(X - X') \mid X\right] \\ &= \mathbb{E}\left[\left(\text{sign}(X - X') \otimes \text{sign}(X - X')\right)^\top \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \mid X\right] \\ &= \mathbb{E}\left[\left(\text{sign}(X - X') \otimes \text{sign}(X - X')\right) \mid X\right]^\top \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \\ &= h_1(X)^\top \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \ , \end{aligned}$$

where $h_1(X)$ is defined in Assumption 3.4, and has variance Σ_{h_1}. Therefore,

$$\begin{aligned} \nu_{g_1}^2 &= \mathsf{Var}(g_1(X)) \\ &= \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)^\top \cdot \Sigma_{h_1} \cdot \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \\ &\geqslant C_{\mathsf{kernel}} \cdot \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)^\top \cdot \Sigma_h \cdot \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \quad \text{(by Assumption 3.4)} \\ &= C_{\mathsf{kernel}} \cdot \mathsf{Var}\left(\mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)^\top h(X, X')\right) \\ &= C_{\mathsf{kernel}} \cdot \mathsf{Var}\left(\text{sign}(X - X')^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \text{sign}(X - X')\right) \\ &= C_{\mathsf{kernel}} \cdot \mathsf{Var}\left(\text{sign}(Z)^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \text{sign}(Z)\right) \end{aligned}$$

(D.4)

$$\geqslant C_{\mathsf{kernel}} \cdot C_{\mathsf{signs}} \cdot \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)_{ab}^2 \ ,$$

where for the next-to-last step, we take $Z \sim N(0, \Sigma)$ and apply Lemma 4.4 to see that $\text{sign}(X - X')$ has the same distribution as $\text{sign}(Z)$, and for the last step, we apply the following lemma (proved in Appendix D.10).

LEMMA D.3. *Take any positive definite $\Sigma \in \mathbb{R}^{p\times p}$, any distinct $a, b \in \{1,\dots,p\}$, and any matrix $M \in \mathbb{R}^{p\times p}$ with $M_{ja} = 0$ for all j. Let $Z \sim N(0,\Sigma)$. Then there exists a constant C_{signs} depending on $\mathsf{C}(\Sigma)$ only, such that*

$$\mathsf{Var}\left(\text{sign}(Z)^\top M \,\text{sign}(Z)\right) \geqslant C_{\mathsf{signs}} \cdot M_{ab}^2 \,.$$

Applying this lemma with $M = uv^\top \circ \cos\left(\frac{\pi}{2}T\right)$ yields the lower bound (D.4) (since $v_a = 0$ so that $M_{ja} = 0$ for all j).

Finally, we have

$$\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)_{ab}^2 = u_a^2 v_b^2 \cos\left(\frac{\pi}{2}T_{ab}\right)^2 \geqslant (C_{\mathsf{cov}})^{-2} \,,$$

where the last step holds because $u_a = v_b = 1$ and

$$\cos\left(\frac{\pi}{2}T_{ab}\right) = \sqrt{1 - \sin\left(\frac{\pi}{2}T_{ab}\right)^2} = \sqrt{1-\Sigma_{ab}^2} \geqslant \lambda_{\min}\left(\Sigma_{ab,ab}\right) \geqslant (C_{\mathsf{cov}})^{-1} \,.$$

To summarize, we have

$$\nu_{g_1}^2 \geqslant \frac{C_{\mathsf{kernel}} C_{\mathsf{signs}}}{C_{\mathsf{cov}}^2} =: \frac{1}{\pi^2} C_{\mathsf{variance}}^2 \,.$$

Next, we give an upper bound on $\nu_{g_1}^2$:

$$\begin{aligned}
\nu_{g_1}^2 &= \mathsf{Var}(g_1(X)) \\
&= \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)^\top \cdot \Sigma_{h_1} \cdot \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \\
&\leqslant \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)^\top \cdot \Sigma_{h} \cdot \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \\
&= \mathsf{Var}\left(\text{sign}(X-X')^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)\text{sign}(X-X')\right) \quad \text{(as for the lower bound earlier)} \\
&= \mathsf{Var}\left(g(X,X')\right) \leqslant \mathbb{E}\left[|g(X,X')|^2\right] \leqslant \mathbb{E}\left[|g(X,X')|^3\right]^{2/3} = \eta_g^2 \,.
\end{aligned}$$

Finally, we compute an upper bound on $\eta_g^3 = \mathbb{E}\left[|g(X,X')|^3\right]$. By Lemma C.1, there exists a decomposition

$$\cos\left(\frac{\pi}{2}T\right) = \sum_{r\geqslant 1} t_r a_r b_r^\top$$

where $t_r \geqslant 0$, $\sum_r t_r \leqslant 4$, and $||a_r||_\infty, ||b_r||_\infty \leqslant 1$. Note that, by (B.1),

$$||u||_2 = \sqrt{1 + ||\gamma_a||_2^2} \leqslant \sqrt{1 + C_{\mathsf{cov}}^2}$$

and similarly $||v||_2 \leqslant \sqrt{1 + C_{\mathsf{cov}}^2}$. Then for each r,

$$||u \circ a_r||_2 \vee ||v \circ b_r||_2 \leqslant \sqrt{1 + C_{\mathsf{cov}}^2}\ .$$

Then we have

$$\begin{aligned}
&\mathbb{E}\left[|g(X,X')|^3\right] = \mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)\mathrm{sign}(X-X')\right|^3\right] \\
&= \mathbb{E}\left[\left|\sum_{r\geqslant 1} t_r \cdot \mathrm{sign}(X-X')^\top \left(uv^\top \circ a_r b_r^\top\right)\mathrm{sign}(X-X')\right|^3\right] \\
&\leqslant \sum_{r\geqslant 1}\frac{t_r}{4}\cdot \mathbb{E}\left[\left|4\,\mathrm{sign}(X-X')^\top \left(uv^\top \circ a_r b_r^\top\right)\mathrm{sign}(X-X')\right|^3\right] \quad \text{(by Jensen's inequality)} \\
&\leqslant 64 \cdot \max_r \mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top \left(uv^\top \circ a_r b_r^\top\right)\mathrm{sign}(X-X')\right|^3\right] \\
&= 64 \cdot \max_r \mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top (u\circ a_r)\right|^3 \cdot \left|\mathrm{sign}(X-X')^\top (v\circ b_r)\right|^3\right] \\
&= 64 \cdot \max_r \sqrt{\mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top (u\circ a_r)\right|^6\right]}\cdot\sqrt{\mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top (v\circ b_r)\right|^6\right]} \\
&\leqslant 64||u\circ a_r||_2^3 \cdot ||v\circ b_r||_2^3 \cdot \max_{||w||_2=1}\mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top w\right|^6\right] \\
&\leqslant 64(1+C_{\mathsf{cov}}^2)^3 \cdot \max_{||w||_2=1}\mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top w\right|^6\right] \\
&\leqslant 64(1+C_{\mathsf{cov}}^2)^3 \cdot C_{\mathsf{cov}}^3 \cdot 6! \cdot 2\sqrt{e} =: C_{\mathsf{moment}}\ ,
\end{aligned}$$

where the last inequality holds because $\mathrm{sign}(X-X')$ is C_{cov}-subgaussian by Lemmas 4.4 and 4.5. □

D.9. Proofs of lemmas for the initial estimators.

LEMMA C.1. *There exist vectors $a_1, a_2, \dots$ and $b_1, b_2, \dots$ with $||a_r||_\infty, ||b_r||_\infty \leqslant 1$ for all $r \geqslant 1$, and a sequence $t_1, t_2, \dots \geqslant 0$ with $\sum_r t_r = 4$, such that $\cos\left(\frac{\pi}{2}T\right) = \sum_{r\geqslant 1} t_r a_r b_r^\top$.*

PROOF OF LEMMA C.1. We will use the matrix max norm, defined for a matrix $M \in \mathbb{R}^{d_1\times d_2}$ as

$$||M||_{\max} = \min\left\{\max_{1\leqslant i\leqslant d_1}||A_{(i)}||_2 \cdot \max_{1\leqslant j\leqslant d_2}||B_{(j)}||_2 \ :\ r\geqslant 1, A\in\mathbb{R}^{d_1\times r}, B\in\mathbb{R}^{d_2\times r} \text{ s.t. } M = A\cdot B^\top\right\}\ ,$$

where $A_{(i)}$ and $B_{(j)}$ denote the ith row of A and the jth row of B, respectively. The matrix max norm satisfies several key properties that we will use here (Srebro and Shraibman, 2005): first,

$$W \succeq 0 \;\Rightarrow ||W||_{\max} \leqslant \max_i W_{ii} \; ; \tag{D.5}$$

second,

$$||W||_{\max} \leqslant 1 \;\Rightarrow\; \frac{W}{2} \in \mathsf{ConvexHull}\left\{ab^\top : ||a||_\infty, ||b||_\infty \leqslant 1\right\} \; ; \tag{D.6}$$

and finally,

(D.7)

$$||W \circ (uv^\top)||_* \leqslant ||W||_{\max} \text{ for all unit vectors } u, v \text{ and all matrices } W \; ,$$

where recall that $||\cdot||_*$ is the matrix nuclear norm (the sum of the singular values).

For our matrix $\cos\left(\frac{\pi}{2}T\right)$, Wegkamp and Zhao (2013) show that

$$\cos\left(\frac{\pi}{2}T\right) = \sum_{r\geqslant 0} \binom{1/2}{r} (-1)^r \Sigma \circ_{2r} \Sigma, \text{ and } \Sigma \circ_{2r} \Sigma \succeq 0 \text{ for all } r \; ,$$

where $\Sigma \circ_{2r} \Sigma$ is the matrix with entries given by elementwise powers of Σ, that is, $(\Sigma \circ_{2r} \Sigma)_{jk} = (\Sigma_{jk})^{2r}$. Then for each $r \geqslant 0$, applying (D.5),

$$||\Sigma \circ_{2r} \Sigma||_{\max} \leqslant \max_i (\Sigma \circ_{2r} \Sigma)_{ii} = \max_i (\Sigma_{ii})^{2r} = 1 \; ,$$

since Σ is a correlation matrix. Then

$$\begin{aligned}\left\|\cos\left(\frac{\pi}{2}T\right)\right\|_{\max} &= \left\|\sum_{r\geqslant 0} \binom{1/2}{r} (-1)^r \Sigma \circ_{2r} \Sigma\right\|_{\max} \\ &\leqslant \sum_{r\geqslant 0} \left|\binom{1/2}{r}\right| \cdot ||\Sigma \circ_{2r} \Sigma||_{\max} \leqslant \sum_{r\geqslant 0} \left|\binom{1/2}{r}\right| = 2 \; ,\end{aligned}$$

where the last identity comes from Wegkamp and Zhao (2013). Finally, by (D.6), we have

$$\frac{\cos\left(\frac{\pi}{2}T\right)}{4} \in \mathsf{ConvexHull}\left\{ab^\top : ||a||_\infty, ||b||_\infty \leqslant 1\right\}$$

and so $\cos\left(\frac{\pi}{2}T\right)$ can be expressed as a convex combination as stated in the lemma. □

LEMMA C.2. *For fixed u, v with $||u||_2, ||v||_2 \leqslant 1$, for any $|t| \leqslant \frac{n}{4(1+\sqrt{5})C_{\mathsf{cov}}}$,*

$$\mathbb{E}\left[\exp\left(t \cdot u^\top(\widehat{T} - T)v\right)\right] \leqslant \exp\left(\frac{\left[4(1+\sqrt{5})\right]^2 t^2 \cdot C_{\mathsf{cov}}^2}{n}\right) .$$

PROOF OF LEMMA C.2. We start with a simple observation that

$$u^T\left(\widehat{T} - T\right)v = \frac{1}{4}(u+v)^\top(\widehat{T} - T)(u+v) - \frac{1}{4}(u-v)^\top(\widehat{T} - T)(u-v),$$

which gives us (via Cauchy-Schwartz)

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t \cdot u^\top(\widehat{T} - T)v\right)\right] \\
&= \mathbb{E}\left[\exp\left(t \cdot \frac{1}{4}(u+v)^\top(\widehat{T} - T)(u+v) - t \cdot \frac{1}{4}(u-v)^\top(\widehat{T} - T)(u-v)\right)\right] \\
&\leqslant \sqrt{\mathbb{E}\left[\exp\left(t \cdot \frac{1}{2}(u+v)^\top(\widehat{T} - T)(u+v)\right)\right]} \cdot \sqrt{\mathbb{E}\left[\exp\left(-t \cdot \frac{1}{2}(u-v)^\top(\widehat{T} - T)(u-v)\right)\right]} .
\end{aligned}$$

Note that $||\frac{1}{2}(u+v)||_2 \vee ||\frac{1}{2}(u-v)||_2 \leqslant 1$. Therefore, it will be sufficient to show that for any $|t| \leqslant \frac{n}{4(1+\sqrt{5})C_{\mathsf{cov}}}$ and any unit vector w,

$$\mathbb{E}\left[\exp\left(2t \cdot w^\top(\widehat{T} - T)w\right)\right] \leqslant \exp\left(\frac{\left[4(1+\sqrt{5})\right]^2 t^2 \cdot C_{\mathsf{cov}}^2}{n}\right) . \tag{D.8}$$

We will prove (D.8) using the Chernoff bounding technique. To that end, denote S_n the group of permutations of $[n]$, and for any i, let $X_{(i)}$ denote the i-th row of X. For a fixed $w \in \mathbb{R}^p$, for each $i \in [n/2]$ and $\sigma \in S_n$, define

$$Z_{\sigma,i} = w^\top\left(\text{sign}\left((X_{(\sigma(i))} - X_{(\sigma(i+n/2))})(X_{(\sigma(i))} - X_{(\sigma(i+n/2))})^\top\right) - T\right)w .$$

Observe that

$$w^T\left(\widehat{T} - T\right)w = \frac{1}{n!}\sum_{\sigma \in S_n}\frac{2}{n}\sum_{i \in [n/2]} Z_{\sigma,i}, \tag{D.9}$$

and that for any fixed $\sigma \in S_n$, the $Z_{\sigma,i}$'s are i.i.d. for $i = 1, \ldots, n/2$, and are identically distributed as

$$\widetilde{Z} = w^\top\left(\text{sign}\left((X_{(1)} - X_{(1+n/2)})(X_{(1)} - X_{(1+n/2)})^\top\right) - T\right)w .$$

Using Lemma 4.4 and Lemma 4.5, for any fixed unit vector $w \in \mathbb{R}^p$, $w^T\,\text{sign}\left(X_{(i)} - X_{(i+n/2)}\right)$ is a C_{cov}-subgaussian random variable, and

$$\widetilde{Z} = \left(w^T\,\text{sign}\left(X_{(1)} - X_{(1+n/2)}\right)\right)^2 - \mathbb{E}\left[\left(w^T\,\text{sign}\left(X_{(1)} - X_{(1+n/2)}\right)\right)^2\right] .$$

Applying Lemma D.4 (stated below), for any $|t| \leqslant \frac{1}{2(1+\sqrt{5})C_{\textsf{cov}}}$,

$$\mathbb{E}\left[\exp\left(t\widetilde{Z}\right)\right] \leqslant \exp\left(\frac{32t^2C_{\textsf{cov}}^2}{1-4tC_{\textsf{cov}}}\right) \leqslant \exp\left(\frac{32t^2C_{\textsf{cov}}^2}{1-\frac{2}{1+\sqrt{5}}}\right) = \exp\left(8(1+\sqrt{5})^2t^2C_{\textsf{cov}}^2\right) \ .$$

Then, referring back to (D.9), for $0 < t \leqslant \frac{n}{4(1+\sqrt{5})C_{\textsf{cov}}}$,

$$\begin{aligned}\mathbb{E}\left[\exp\left(tw^T\left(\widehat{T}-T\right)w\right)\right] &= \mathbb{E}\left[\exp\left(\frac{t}{n!}\sum_{\sigma\in S_n}\frac{2}{n}\sum_{i\in[n/2]}Z_{\sigma,i}\right)\right]\\ &\leqslant \frac{1}{n!}\sum_{\sigma\in S_n}\mathbb{E}\left[\exp\left(\frac{2t}{n}\sum_{i\in[n/2]}Z_{\sigma,i}\right)\right] \quad \text{(by Jensen's inequality)}\\ &= \frac{1}{n!}\sum_{\sigma\in S_n}\left(\mathbb{E}\left[\exp\left(\frac{2t}{n}\widetilde{Z}\right)\right]\right)^{n/2},\end{aligned}$$

since for any fixed σ, the $Z_{\sigma,i}$'s are i.i.d., and are each equal to $\widetilde{Z}$ in distribution. Continuing from this step, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(tw^T\left(\widehat{T}-T\right)w\right)\right] &= \left(\mathbb{E}\left[\exp\left(\frac{2t}{n}\widetilde{Z}\right)\right]\right)^{n/2}\\ &\leqslant \left(\exp\left(8(1+\sqrt{5})^2(2t/n)^2C_{\textsf{cov}}^2\right)\right)^{n/2}\\ &= \exp\left(\frac{\left[4(1+\sqrt{5})\right]^2t^2\cdot C_{\textsf{cov}}^2}{n}\right) \ .\end{aligned}$$

□

LEMMA D.4. *Suppose Z is C-subgaussian, that is, $\mathbb{E}\left[\exp\left(tZ\right)\right] \leqslant e^{Ct^2/2}$ for all $t \in \mathbb{R}$. Then*

$$\mathbb{E}\left[\exp\left\{t(Z^2-\mathbb{E}[Z^2])\right\}\right] \leqslant \exp\left(\frac{32t^2C^2}{1-4|t|C}\right)$$

for all $|t| < \frac{1}{4C}$.

We remark that it is well known that the square of a subgaussian random variable satisfies subgaussian tails near to its mean (see, for example, Lemmas 5.5, 5.14, 5.15 in Vershynin, 2012), but here we obtain small explicit constants.

PROOF. The first part of this proof follows the arguments in Vershynin (2012, Lemma 5.5). First, we bound $\mathbb{E}\left[Z^{2k}\right]$ for all integers $k \geqslant 1$. We have

$$\mathbb{E}\left[Z^{2k}\right] = \frac{C^k}{(2k)^k}\mathbb{E}\left[\left(\sqrt{\frac{2k}{C}}\cdot Z\right)^{2k}\right] \leqslant \frac{C^k}{(2k)^k}\mathbb{E}\left[(2k)!\cdot\exp\left\{\sqrt{\frac{2k}{C}}\cdot Z\right\}\right]$$
$$\leqslant \frac{C^k}{(2k)^k}(2k)!\cdot\exp\left\{\left(\sqrt{\frac{2k}{C}}\right)^2\cdot C/2\right\} = \frac{C^k(2k)!e^k}{(2k)^k}.$$

Then, for any $t > 0$,

$$\mathbb{E}\left[e^{tZ^2}\right] = 1 + t\mathbb{E}\left[Z^2\right] + \sum_{k\geqslant 2}\mathbb{E}\left[\frac{\left(tZ^2\right)^k}{k!}\right] \leqslant 1 + t\mathbb{E}\left[Z^2\right] + \sum_{k\geqslant 2}\frac{t^kC^ke^k}{(2k)^k}\cdot\frac{(2k)!}{k!}$$
$$\leqslant 1 + t\mathbb{E}\left[Z^2\right] + \sum_{k\geqslant 2}\frac{t^kC^ke^k}{(2k)^k}\cdot\frac{e\cdot(2k)^{2k+1/2}\cdot e^{-2k}}{\sqrt{2\pi}\cdot k^{k+1/2}\cdot e^{-k}} \quad \text{(by Stirling's approximation for } (2k)! \text{ and } k!)$$
$$= 1 + t\mathbb{E}\left[Z^2\right] + \frac{e}{\sqrt{\pi}}\sum_{k\geqslant 2}(2tC)^k$$
$$= 1 + t\mathbb{E}\left[Z^2\right] + \frac{e}{\sqrt{\pi}}\cdot\frac{4t^2C^2}{1-2tC}$$

(D.10)

$$\leqslant 1 + t\mathbb{E}\left[Z^2\right] + \frac{8t^2C^2}{1-2tC},$$

as long as $2tC < 1$ (we also use the fact $\frac{e}{\sqrt{\pi}} \leqslant 2$ to simplify). Next, trivially, for any $k \geqslant 2$, $\left|\mathbb{E}\left[\left(Z^2 - \mathbb{E}[Z^2]\right)^k\right]\right| \leqslant 2^k\mathbb{E}\left[Z^{2k}\right]$. Then we have, for $|t| < \frac{1}{4C}$,

$$\mathbb{E}\left[\exp\left(t(Z^2 - \mathbb{E}[Z^2])\right)\right]$$
$$= 1 + \sum_{k\geqslant 2}\frac{t^k}{k!}\mathbb{E}\left[\left(Z^2 - \mathbb{E}[Z^2]\right)^k\right] \leqslant 1 + \sum_{k\geqslant 2}\frac{|t|^k}{k!}\left|\mathbb{E}\left[\left(Z^2 - \mathbb{E}[Z^2]\right)^k\right]\right|$$
$$\leqslant 1 + \sum_{k\geqslant 2}\frac{2^k|t|^k}{k!}\mathbb{E}\left[Z^{2k}\right] = \left(\sum_{k\geqslant 0}\frac{2^k|t|^k}{k!}\mathbb{E}\left[Z^{2k}\right]\right) - 2|t|\mathbb{E}\left[Z^2\right]$$
$$= \mathbb{E}\left[\exp\left(2|t|Z^2\right)\right] - 2|t|\mathbb{E}\left[Z^2\right]$$
$$\leqslant 1 + 2|t|\mathbb{E}\left[Z^2\right] + \frac{32t^2C^2}{1-4|t|C} - 2|t|\mathbb{E}\left[Z^2\right] \quad \text{(by (D.10) with } 2|t| \text{ in place of } t)$$
$$\leqslant \exp\left\{\frac{32t^2C^2}{1-4|t|C}\right\}.$$

□

D.10. Lower bounds on variance for signs of a Gaussian.

LEMMA D.3. *Take any positive definite $\Sigma \in \mathbb{R}^{p\times p}$, any distinct $a, b \in \{1,\dots,p\}$, and any matrix $M \in \mathbb{R}^{p\times p}$ with $M_{ja} = 0$ for all j. Let $Z \sim N(0,\Sigma)$. Then there exists a constant C_{signs} depending on $\mathsf{C}(\Sigma)$ only, such that*

$$\mathsf{Var}\left(\mathrm{sign}(Z)^\top M\,\mathrm{sign}(Z)\right) \geqslant C_{\mathsf{signs}} \cdot M_{ab}^2\,.$$

PROOF OF LEMMA D.3. Let $(-a)$ denote the set $[p]\backslash\{a\}$. By the law of total variance,

$$\mathsf{Var}\left(\mathrm{sign}(Z)^\top M\,\mathrm{sign}(Z)\right) \geqslant \mathbb{E}\left[\mathsf{Var}\left(\mathrm{sign}(Z)^\top M\,\mathrm{sign}(Z) \mid Z_{(-a)}\right)\right]\,.$$

Let $M_{j,(-a)} \in \mathbb{R}^{p-1}$ denote the jth row of M with its ath entry removed, written as a column vector. Then, recalling that $M_{ja} = 0$ for all j, we have

$$\begin{aligned}
&\mathsf{Var}\left(\mathrm{sign}(Z)^\top M\,\mathrm{sign}(Z) \mid Z_{(-a)}\right)\\
&= \mathsf{Var}\left(\mathrm{sign}(Z_a)\cdot M_{a,(-a)}^\top\,\mathrm{sign}(Z_{(-a)}) + \sum_{j\neq a}\mathrm{sign}(Z_j)\cdot M_{j,(-a)}^\top\,\mathrm{sign}(Z_{(-a)}) \mid Z_{(-a)}\right)\\
&= \mathsf{Var}\left(\mathrm{sign}(Z_a)\cdot M_{a,(-a)}^\top\,\mathrm{sign}(Z_{(-a)}) \mid Z_{(-a)}\right)\\
&= \mathsf{Var}\left(\mathrm{sign}(Z_a) \mid Z_{(-a)}\right)\cdot\left(M_{a,(-a)}^\top\,\mathrm{sign}(Z_{(-a)})\right)^2\\
&= \mathsf{Var}\left(\mathrm{sign}(Z_{(-a)}^\top\beta_a + N(0,\nu_a^2))\right)\cdot\left(M_{a,(-a)}^\top\,\mathrm{sign}(Z_{(-a)})\right)^2,
\end{aligned}$$

where the last step holds since the distribution of Z_a conditional on $Z_{(-a)}$ is given by $Z_{(-a)}^\top\beta_a + N(0,\nu_a^2)$ where $\beta_a = \Sigma_{(-a)}^{-1}\Sigma_{(-a),a}$ and $\nu_a^2 = \Sigma_{aa} - \Sigma_{(-a),a}^\top\Sigma_{(-a)}^{-1}\Sigma_{(-a),a}$. Continuing from this step, we have

$$\begin{aligned}
&\mathsf{Var}\left(\mathrm{sign}(Z)^\top M\,\mathrm{sign}(Z) \mid Z_{(-a)}\right)\\
&= \left(1 - \mathbb{E}\left[\mathrm{sign}(Z_{(-a)}^\top\beta_a + N(0,\nu_a^2))\right]^2\right)\cdot\left(M_{a,(-a)}^\top\,\mathrm{sign}(Z_{(-a)})\right)^2\\
&= \left(1 - \psi\left(\frac{Z_{(-a)}^\top\beta_a}{\nu_a}\right)^2\right)\cdot\left(M_{a,(-a)}^\top\,\mathrm{sign}(Z_{(-a)})\right)^2\,,
\end{aligned}$$

where $\psi(z) = \Phi(z) - \Phi(-z) \in (-1,1)$.

Now we will give a lower bound on the expectation of this quantity. First consider the term $\psi\left(\frac{Z_{(-a)}^\top\beta_a}{\nu_a}\right)$. Note that

$$\frac{Z_{(-a)}^\top\beta_a}{\nu_a} \sim N\left(0, \frac{\beta_a^\top\Sigma_{(-a)}\beta_a}{\nu_a^2}\right) = N\left(0, \frac{\Sigma_{(-a),a}^\top\Sigma_{(-a)}^{-1}\Sigma_{(-a),a}}{\Sigma_{aa} - \Sigma_{(-a),a}^\top\Sigma_{(-a)}^{-1}\Sigma_{(-a),a}}\right)$$

and this variance is bounded by $\mathsf{C}(\Sigma)$. Then, for any $c \in (0,1)$,

(D.11)
$$\mathbb{P}\left\{\left|\psi\left(\frac{Z_{(-a)}^\top \beta_a}{\nu_a}\right)\right| \leqslant \psi\left(\sqrt{\mathsf{C}(\Sigma)}\cdot\Phi^{-1}\left(1-c/2\right)\right)\right\} = \mathbb{P}\left\{\left|\frac{Z_{(-a)}^\top \beta_a}{\nu_a}\right| \leqslant \sqrt{\mathsf{C}(\Sigma)}\cdot\Phi^{-1}\left(1-c/2\right)\right\}$$
$$\geqslant \mathbb{P}\left\{|N(0,\mathsf{C}(\Sigma))| \leqslant \sqrt{\mathsf{C}(\Sigma)}\cdot\Phi^{-1}\left(1-c/2\right)\right\} = 1-c\;.$$

Next, note that $M_{a,(-a)}^\top \operatorname{sign}(Z_{(-a)})$ is $\left(||M_{a,(-a)}||_2^2\cdot\mathsf{C}(\Sigma)\right)$-subgaussian by Lemma 4.5, and

$$\mathbb{E}\left[\left(M_{a,(-a)}^\top \operatorname{sign}(Z_{(-a)})\right)^2\right] \geqslant ||M_{a,(-a)}||_2^2\cdot\lambda_{\min}(T)\;,$$

where $T = \mathbb{E}\left[\operatorname{sign}(Z)\operatorname{sign}(Z)^\top\right]$ (recall that $\Sigma = \sin\left(\frac{\pi}{2}T\right)$). Furthermore, by Wegkamp and Zhao (2013, Section 4.3), we have

$$T = \frac{2}{\pi}\sum_{k\geqslant 1} g(k)\Sigma\circ_k\Sigma\;,$$

where $g(k)\geqslant 0$ are nonnegative scalars, $g(1)=1$, and $\Sigma\circ_k\Sigma$ is the k-fold Hadamard product, that is, $(\Sigma\circ_k\Sigma)_{ij} = (\Sigma_{ij})^k$. Wegkamp and Zhao (2013, Section 4.3) show also that $\Sigma\circ_k\Sigma\succeq 0$ for all k. Therefore,

$$T = \frac{2}{\pi}\Sigma + \frac{2}{\pi}\sum_{k\geqslant 2} g(k)\Sigma\circ_k\Sigma \succeq \frac{2}{\pi}\Sigma\;,$$

and so $\lambda_{\min}(T) \geqslant \frac{2}{\pi}\lambda_{\min}(\Sigma) \geqslant \frac{2}{\pi}(\mathsf{C}(\Sigma))^{-1}$. Applying Lemma D.5 (stated below),

$$\mathbb{P}\left\{\left(M_{a,(-a)}^\top \operatorname{sign}(Z_{(-a)})\right)^2 \geqslant ||M_{a,(-a)}||_2^2\cdot\lambda_{\min}(T)/2\right\} \geqslant \frac{1}{16e^{2\mathsf{C}(\Sigma)/\lambda_{\min}(T)}}\;,$$

and so,

$$\mathbb{P}\left\{\left(M_{a,(-a)}^\top \operatorname{sign}(Z_{(-a)})\right)^2 \geqslant ||M_{a,(-a)}||_2^2\cdot\frac{1}{\pi\mathsf{C}(\Sigma)}\right\} \geqslant \frac{1}{16e^{\pi\mathsf{C}(\Sigma)^2}}\;.$$

Now set $c = \frac{1}{32e^{\pi\mathsf{C}(\Sigma)^2}}$ in (D.11). Then, we see that with probability at least $\frac{1}{32e^{\pi\mathsf{C}(\Sigma)^2}}$,

$$\left(1-\psi\left(\frac{Z_{(-a)}^\top\beta_a}{\nu_a}\right)^2\right)\cdot\left(M_{a,(-a)}^\top \operatorname{sign}(Z_{(-a)})\right)^2 \geqslant$$
$$\left(1-\psi\left(\sqrt{\mathsf{C}(\Sigma)}\cdot\Phi^{-1}\left(1-\frac{1}{64e^{\pi\mathsf{C}(\Sigma)^2}}\right)\right)^2\right)\cdot||M_{a,(-a)}||_2^2\cdot\frac{1}{\pi\mathsf{C}(\Sigma)}\;.$$

Therefore, combining everything,

$$\mathsf{Var}\left(\mathrm{sign}(Z)^\top M\,\mathrm{sign}(Z)\right) \geqslant \frac{1}{32e^{\pi\mathsf{C}(\Sigma)^2}}\cdot\left(1-\psi\left(\sqrt{\mathsf{C}(\Sigma)}\cdot\Phi^{-1}\left(1-\frac{1}{64e^{\pi\mathsf{C}(\Sigma)^2}}\right)\right)^2\right)\cdot\frac{||M_{a,(-a)}||_2^2}{\pi\mathsf{C}(\Sigma)}\ .$$

Noting that $||M_{a,(-a)}||_2^2 \geqslant M_{ab}^2$, this proves the desired result, where we define

$$C_{\mathsf{signs}} = \frac{1}{32e^{\pi\mathsf{C}(\Sigma)^2}}\cdot\left(1-\psi\left(\sqrt{\mathsf{C}(\Sigma)}\cdot\Phi^{-1}\left(1-\frac{1}{64e^{\pi\mathsf{C}(\Sigma)^2}}\right)\right)^2\right)\cdot\frac{1}{\pi\mathsf{C}(\Sigma)}\ .$$

□

LEMMA D.5. *Suppose that $W\in\mathbb{R}$ is a random variable with $\mathbb{E}[W]=0$, $\mathbb{E}[W^2]\geqslant C_0$, and $\mathbb{E}[e^{tW}]\leqslant e^{C_1t^2/2}$ for all $t\in\mathbb{R}$. Then*

$$\mathbb{P}\left\{W^2\geqslant C_0/2\right\}\geqslant\frac{1}{16e^{2C_1/C_0}}\ .$$

PROOF OF LEMMA D.5.

$$\begin{aligned}
C_0/2 &\leqslant \mathbb{E}[W^2]-C_0/2\\
&=\mathbb{E}[W^2\cdot\mathbb{1}\left\{W^2\geqslant C_0/2\right\}]+\mathbb{E}[W^2\cdot\mathbb{1}\left\{W^2<C_0/2\right\}]-C_0/2\\
&\leqslant \mathbb{E}[W^2\cdot\mathbb{1}\left\{W^2\geqslant C_0/2\right\}]\\
&\leqslant C_0\mathbb{E}[(e^{W/\sqrt{C_0}}+e^{-W/\sqrt{C_0}})\cdot\mathbb{1}\left\{W^2\geqslant C_0/2\right\}]\quad\left(\text{since } t^2\leqslant e^t+e^{-t}\text{ for all } t\in\mathbb{R}\right)\\
&=C_0\mathbb{E}[e^{W/\sqrt{C_0}}\cdot\mathbb{1}\left\{W^2\geqslant C_0/2\right\}]+C_0\mathbb{E}[e^{-W/\sqrt{C_0}}\cdot\mathbb{1}\left\{W^2\geqslant C_0/2\right\}]\\
&\leqslant C_0\sqrt{\mathbb{E}[(e^{W/\sqrt{C_0}})^2]\cdot\mathbb{E}[\mathbb{1}\left\{W^2\geqslant C_0/2\right\}^2]}+C_0\sqrt{\mathbb{E}[(e^{-W/\sqrt{C_0}})^2]\cdot\mathbb{E}[\mathbb{1}\left\{W^2\geqslant C_0/2\right\}^2]}\\
&=C_0\sqrt{\mathbb{E}[e^{2W/\sqrt{C_0}}]\cdot\mathbb{P}\{W^2\geqslant C_0/2\}}+C_0\sqrt{\mathbb{E}[e^{-2W/\sqrt{C_0}}]\cdot\mathbb{P}\{W^2\geqslant C_0/2\}}\\
&\leqslant C_0\sqrt{e^{C_1/C_0\cdot 2^2/2}\cdot\mathbb{P}\{W^2\geqslant C_0/2\}}+C_0\sqrt{e^{C_1/C_0\cdot 2^2/2}\cdot\mathbb{P}\{W^2\geqslant C_0/2\}}\ ,
\end{aligned}$$

and rearranging terms we have proved the lemma. □

D.11. Bounding the error in estimating the variance (Lemma B.5).

LEMMA B.5. *Under the assumptions and definitions of Theorem 4.2, with probability at least $1-\frac{1}{6p_n}$, if $n\geqslant k_n^2\log(p_n)$, on the event that the bounds (3.1) in Assumption 3.3 hold,*

$$\left|\check{S}_{ab}\cdot\det(\check{\Theta})-S_{ab}\cdot\det(\Theta)\right|\leqslant C_{\mathsf{oracle}}\cdot\sqrt{\frac{k_n^2\log(p_n)}{n}}\ .$$

PROOF OF LEMMA B.5. Recall from the proof of Theorem 4.1 that we have defined

$$g(X, X') = \text{sign}(X - X')^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \text{sign}(X - X') ,$$

and $g_1(X) = \mathbb{E}[g(X, X') \mid X]$, where

$$u_a = 1, u_b = 0, u_I = -\gamma_a \text{ and } v_a = 0, v_b = 1, v_I = -\gamma_b .$$

Recall from the proof of Lemma B.1, given in Appendix D.8, that we have

$$\nu_{g_1}^2 = \mathsf{Var}(g_1(X)) = \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right)^\top \cdot \Sigma_{h_1} \cdot \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) ,$$

where $\Sigma_{h_1} = \mathsf{Var}(h_1(X))$ for

$$h_1(X) = \mathbb{E}\left[\text{sign}(X - X') \otimes \text{sign}(X - X') \mid X\right] \in \mathbb{R}^{p_n^2} .$$

Note that $||\Sigma_{h_1}||_\infty = 1$.

To estimate this variance, define vectors $\breve{u}$ and $\breve{v}$ with entries

$$\breve{u}_a = 1, \breve{u}_b = 0, \breve{u}_I = -\breve{\gamma}_a \text{ and } \breve{v}_a = 0, \breve{v}_b = 1, \breve{v}_I = -\breve{\gamma}_b ,$$

and define

$$\widehat{\Sigma}_{h_1} = \frac{1}{n}\sum_i \left(\widehat{h}_1(X_i) - \frac{1}{n}\sum_{i'}\widehat{h}_1(X_{i'})\right)\left(\widehat{h}_1(X_i) - \frac{1}{n}\sum_{i'}\widehat{h}_1(X_{i'})\right)^\top ,$$

where abusing notation, we write

$$\widehat{h}_1(X_i) = \frac{1}{n-1}\sum_{i' \neq i} h(X_i, X_i') = \frac{1}{n-1}\sum_{i' \neq i} \text{sign}(X_i - X_{i'}) \otimes \text{sign}(X_i - X_{i'}) .$$

We then define

$$\breve{\nu}_{g_1}^2 = \mathsf{vec}\left(\breve{u}\breve{v}^\top \circ \cos\left(\frac{\pi}{2}\widehat{T}\right)\right)^\top \cdot \widehat{\Sigma}_{h_1} \cdot \mathsf{vec}\left(\breve{u}\breve{v}^\top \circ \cos\left(\frac{\pi}{2}\widehat{T}\right)\right) .$$

Writing

$$x = \mathsf{vec}\left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \text{ and } \breve{x} = \mathsf{vec}\left(\breve{u}\breve{v}^\top \circ \cos\left(\frac{\pi}{2}\widehat{T}\right)\right) ,$$

we then have

$$\nu_{g_1}^2 = x^\top \Sigma_{h_1} x \text{ and } \breve{\nu}_{g_1}^2 = \breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x} .$$

Define also

$$\overline{x} = \mathsf{vec}\left(\breve{u}\breve{v}^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) .$$

The following lemma, proved in Appendix D.12, carries out some elementary calculations on the norms of these vectors $x, \overline{x}, \breve{x}$.

LEMMA D.6. *Define $x, \overline{x}, \breve{x}$ as in the proof of Lemma B.5, and assume $n \geqslant k_n^2 \log(p_n)$. If the bounds (3.1) in Assumption 3.3 hold then for constants C_0, C_1, C_2, C_3 that depend only on $C_{\mathsf{cov}}, C_{\mathsf{sparse}}, C_{\mathsf{est}}$,*

$$||x||_1 \leqslant C_0 k_n,$$

and with probability at least $1 - \frac{1}{36p_n}$, the following bounds all hold as well:

$$||\breve{x} - \overline{x}||_1 \leqslant C_1 \sqrt{\frac{k_n^2 \log(p_n)}{n}}\ ,$$
$$||\breve{x} - x||_1 \leqslant C_2 \sqrt{\frac{k_n^3 \log(p_n)}{n}}\ ,$$
$$||\mathsf{mat}(\overline{x} - x)||_{\ell_1/\ell_2} \leqslant C_3 \sqrt{\frac{k_n^2 \log(p_n)}{n}}\ ,$$

where $\mathsf{mat}(\cdot)$ reshapes a vector in $\mathbb{R}^{p_n^2}$ into a $p_n \times p_n$ matrix, and where we define the matrix ℓ_1/ℓ_2 norm as $M_{\ell_1/\ell_2} := \sum_j ||M_j||_2$, where M_j is the jth column of M.

We now continue bounding error in estimating ν_{g_1}. We have:
(D.12)
$$\left|\breve{\nu}_{g_1}^2 - \nu_{g_1}^2\right| = \left|\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x} - x^\top \Sigma_{h_1} x\right| \leqslant \left|x^\top (\widehat{\Sigma}_{h_1} - \Sigma_{h_1}) x\right| + \left|\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x} - x^\top \widehat{\Sigma}_{h_1} x\right|\ .$$
We bound each term separately. For the first term in (D.12), we apply the following lemma (proved in Appendix D.12 below):

LEMMA D.7. *Under the same assumptions and notation as Lemmas B.1 and B.5, for a universal constant $C_{\mathsf{studentized}}$,*

$$\mathbb{P}\left\{\left|x^\top (\widehat{\Sigma}_{h_1} - \Sigma_{h_1}) x\right| \leqslant C_{\mathsf{studentized}} \sqrt{\frac{k_n^2 \log(p_n)}{n}}\right\} \geqslant 1 - \frac{1}{36p_n}\ .$$

For the second term in (D.12), since $\widehat{\Sigma}_{h_1} \succeq 0$ and so $y \mapsto \sqrt{y^\top \widehat{\Sigma}_{h_1} y}$ is a norm and must satisfy the triangle inequality,

(D.13)
$$\begin{aligned}
\left|\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x} - x^\top \widehat{\Sigma}_{h_1} x\right| &= \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right| \cdot \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} + \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right| \\
&\leqslant \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right|^2 + \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right| \cdot 2\sqrt{x^\top \widehat{\Sigma}_{h_1} x} \\
&\leqslant \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right|^2 + \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right| \cdot 2\sqrt{x^\top \Sigma_{h_1} x + \left|x^\top (\widehat{\Sigma}_{h_1} - \Sigma_{h_1}) x\right|}\ .
\end{aligned}$$

To bound the difference term $\left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right|$ which appears twice in the expression above, applying the triangle inequality several times, we have

$$\left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1} \breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right| \leqslant \sqrt{(\breve{x} - x)^\top \widehat{\Sigma}_{h_1} (\breve{x} - x)} \tag{D.14}$$

$$\leqslant \sqrt{(\breve{x} - x)^\top \Sigma_{h_1} (\breve{x} - x)} + \sqrt{\left|(\breve{x} - x)^\top (\widehat{\Sigma}_{h_1} - \Sigma_{h_1}) (\breve{x} - x)\right|} \tag{D.15}$$

$$\leqslant \sqrt{(\breve{x} - x)^\top \Sigma_{h_1} (\breve{x} - x)} + \sqrt{||\widehat{\Sigma}_{h_1} - \Sigma_{h_1}||_\infty \cdot ||\breve{x} - x||_1^2} \tag{D.16}$$

$$\leqslant \sqrt{(\breve{x} - \overline{x})^\top \Sigma_{h_1} (\breve{x} - \overline{x})} + \sqrt{(\overline{x} - x)^\top \Sigma_{h_1} (\overline{x} - x)} + \sqrt{||\widehat{\Sigma}_{h_1} - \Sigma_{h_1}||_\infty \cdot ||\breve{x} - x||_1^2} \tag{D.17}$$

$$\leqslant \sqrt{||\Sigma_{h_1}||_\infty ||\breve{x} - \overline{x}||_1^2} + \sqrt{(\overline{x} - x)^\top \Sigma_{h_1} (\overline{x} - x)} + \sqrt{||\widehat{\Sigma}_{h_1} - \Sigma_{h_1}||_\infty \cdot ||\breve{x} - x||_1^2} \tag{D.18}$$

$$\leqslant C_1 \sqrt{\frac{k_n^2 \log(p_n)}{n}} + \sqrt{(\overline{x} - x)^\top \Sigma_{h_1} (\overline{x} - x)} + \sqrt{||\widehat{\Sigma}_{h_1} - \Sigma_{h_1}||_\infty} \cdot C_2 \sqrt{\frac{k_n^3 \log(p_n)}{n}} \ . \tag{D.19}$$

Next, we state two lemmas, which are proved in Appendix D.12.

LEMMA D.8. *With probability at least* $1 - \frac{1}{9p_n}$,

$$||\widehat{\Sigma}_{h_1} - \Sigma_{h_1}||_\infty \leqslant 100 \sqrt{\frac{\log(p_n)}{n}} \ .$$

LEMMA D.9. *Let* Σ_{h_1} *be defined as in Assumption 3.4. For every* $z \in \mathbb{R}^{p_n^2}$,

$$z^\top \Sigma_{h_1} z \leqslant \lambda_{\max}(\Sigma) \cdot ||\mathsf{mat}(z)||_{\ell_1/\ell_2}^2 \ ,$$

where $||\mathsf{mat}(z)||_{\ell_1/\ell_2}$ *is defined as in the statement of Lemma D.6.*

From this point on, we assume that the bounds derived in Lemmas D.6 and D.8 all hold (which the lemmas have shown to be true with probability at least $1 - \frac{1}{6p_n}$, on the event that the bounds (3.1) of Assumption 3.3 hold.) By Lemmas D.6 and D.9,

$$(\overline{x} - x)^\top \Sigma_{h_1} (\overline{x} - x) \leqslant C_{\mathsf{cov}} \cdot \left(C_3 \sqrt{\frac{k_n^2 \log(p_n)}{n}} \right)^2 \ .$$

Applying this bound, along with the high probability events of Lemmas D.7 and D.8, we return to (D.14) and obtain

$$\left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1}\breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right| \leqslant$$

$$C_1\sqrt{\frac{k_n^2\log(p_n)}{n}} + \sqrt{C_{\mathsf{cov}}\cdot\left(C_3\sqrt{\frac{k_n^2\log(p_n)}{n}}\right)^2} + \sqrt{100\sqrt{\frac{\log(p_n)}{n}}\cdot C_2\sqrt{\frac{k_n^3\log(p_n)}{n}}}$$

$$= \sqrt{\frac{k_n^2\log(p_n)}{n}}\cdot\left(C_1 + C_3\sqrt{C_{\mathsf{cov}}} + 10C_2\sqrt[4]{\frac{k_n^2\log(p_n)}{n}}\right) \leqslant \sqrt{\frac{k_n^2\log(p_n)}{n}}\cdot C_4\ ,$$

where for the last step we define $C_4 = C_1 + C_3\sqrt{C_{\mathsf{cov}}} + 10C_2$ and use the assumption $n \geqslant k_n^2\log(p_n)$. Next, returning to (D.13),

$$\left|\breve{x}^\top \widehat{\Sigma}_{h_1}\breve{x} - x^\top \widehat{\Sigma}_{h_1} x\right|$$

$$\leqslant \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1}\breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right|^2 + \left|\sqrt{\breve{x}^\top \widehat{\Sigma}_{h_1}\breve{x}} - \sqrt{x^\top \widehat{\Sigma}_{h_1} x}\right|\cdot 2\sqrt{x^\top \Sigma_{h_1} x + \left|x^\top(\widehat{\Sigma}_{h_1} - \Sigma_{h_1})x\right|}$$

$$\leqslant C_4^2\cdot\frac{k_n^2\log(p_n)}{n} + C_4\sqrt{\frac{k_n^2\log(p_n)}{n}}\cdot 2\sqrt{C_{\mathsf{moment}}^{2/3} + C_{\mathsf{studentized}}\sqrt{\frac{k_n^2\log(p_n)}{n}}}\ ,$$

where the last step applies the high probability event of Lemma D.7, and uses the fact that $x^\top\Sigma_{h_1}x = \nu_{g_1}^2 \leqslant C_{\mathsf{moment}}^{2/3}$ by Lemma B.1. Defining $C_5 = C_4^2 + C_4\cdot 2\sqrt{C_{\mathsf{moment}}^{2/3} + C_{\mathsf{studentized}}}$, and using the assumption $n \geqslant k_n^2\log(p_n)$, we have

$$\left|\breve{x}^\top \widehat{\Sigma}_{h_1}\breve{x} - x^\top \widehat{\Sigma}_{h_1} x\right| \leqslant C_5\sqrt{\frac{k_n^2\log(p_n)}{n}}\ .$$

Finally, returning to (D.12) and applying Lemma D.8, we see that

$$\left|\breve{\nu}_{g_1}^2 - \nu_{g_1}^2\right| \leqslant C_{\mathsf{studentized}}\cdot\sqrt{\frac{k_n^2\log(p_n)}{n}} + C_5\sqrt{\frac{k_n^2\log(p_n)}{n}}\ .$$

Next, we have

$$\left|\breve{\nu}_{g_1} - \nu_{g_1}\right| = \frac{\left|\breve{\nu}_{g_1}^2 - \nu_{g_1}^2\right|}{\breve{\nu}_{g_1} + \nu_{g_1}} \leqslant \frac{\left|\breve{\nu}_{g_1}^2 - \nu_{g_1}^2\right|}{\nu_{g_1}} \leqslant \frac{(C_{\mathsf{studentized}} + C_5)\sqrt{\frac{k_n^2\log(p_n)}{n}}}{\frac{1}{\pi}C_{\mathsf{variance}}}\ ,$$

where for the denominator we apply Lemma B.1. Finally, since we know that $S_{ab} = \pi\nu_{g_1}\cdot(\det(\Theta))^{-1}$ and $\breve{S}_{ab} = \pi\breve{\nu}_{g_1}\cdot(\det(\breve{\Theta}))^{-1}$, and then we have

$$\left|\breve{S}_{ab}\cdot\det(\breve{\Theta}) - S_{ab}\cdot\det(\Theta)\right| = \pi\cdot\left|\breve{\nu}_{g_1} - \nu_{g_1}\right| \leqslant \pi\cdot\frac{(C_{\mathsf{studentized}} + C_5)\sqrt{\frac{k_n^2\log(p_n)}{n}}}{\frac{1}{\pi}C_{\mathsf{variance}}}\ .$$

Defining

$$C_{\mathsf{oracle}} \geqslant \pi \cdot \frac{C_{\mathsf{studentized}} + C_5}{\frac{1}{\pi} C_{\mathsf{variance}}} ,$$

we see that

$$\left| \breve{S}_{ab} \cdot \det(\breve{\Theta}) - S_{ab} \cdot \det(\Theta) \right| \leqslant C_{\mathsf{oracle}} \cdot \sqrt{\frac{k_n^2 \log(p_n)}{n}} .$$

□

D.12. Calculations for the variance estimate (Lemma B.5).

LEMMA D.6. *Define $x, \overline{x}, \breve{x}$ as in the proof of Lemma B.5, and assume $n \geqslant k_n^2 \log(p_n)$. If the bounds (3.1) in Assumption 3.3 hold then for constants C_0, C_1, C_2, C_3 that depend only on $C_{\mathsf{cov}}, C_{\mathsf{sparse}}, C_{\mathsf{est}}$,*

$$||x||_1 \leqslant C_0 k_n,$$

and with probability at least $1 - \frac{1}{36 p_n}$, the following bounds all hold as well:

$$||\breve{x} - \overline{x}||_1 \leqslant C_1 \sqrt{\frac{k_n^2 \log(p_n)}{n}} ,$$
$$||\breve{x} - x||_1 \leqslant C_2 \sqrt{\frac{k_n^3 \log(p_n)}{n}} ,$$
$$||\mathsf{mat}(\overline{x} - x)||_{\ell_1/\ell_2} \leqslant C_3 \sqrt{\frac{k_n^2 \log(p_n)}{n}} ,$$

where $\mathsf{mat}(\cdot)$ reshapes a vector in $\mathbb{R}^{p_n^2}$ into a $p_n \times p_n$ matrix, and where we define the matrix ℓ_1/ℓ_2 norm as $M_{\ell_1/\ell_2} := \sum_j ||M_j||_2$, where M_j is the jth column of M.

PROOF OF LEMMA D.6. We calculate

$$||x||_1 = \left\| uv^\top \circ \cos\left(\frac{\pi}{2} T\right) \right\|_1 \leqslant ||uv^\top||_1 \cdot \left\| \cos\left(\frac{\pi}{2} T\right) \right\|_\infty \leqslant ||u||_1 ||v||_1$$
$$\leqslant k_n (1 + 2 C_{\mathsf{cov}} C_{\mathsf{sparse}})^2 =: C_0 k_n ,$$

where for the last inequality we apply (B.3).

Next,

$$||\breve{x} - \overline{x}||_1 = || \left| ||\breve{u}\breve{v}^\top \circ \left(\cos\left(\frac{\pi}{2} \widehat{T}\right) - \cos\left(\frac{\pi}{2} T\right) \right) \right\|_1$$
$$\leqslant (||u||_1 + ||\breve{u} - u||_1) \cdot (||v||_1 + ||\breve{v} - v||_1) \cdot \left\| \cos\left(\frac{\pi}{2} \widehat{T}\right) - \cos\left(\frac{\pi}{2} T\right) \right\|_\infty .$$

Applying (B.3) and Assumption 3.3, and the fact that $\cos(\cdot)$ is 1-Lipschitz, if the bounds in Assumption 3.3 hold, we then have

$$||\breve{x} - \overline{x}||_1 \leqslant \left(\sqrt{k_n}(1 + 2C_{\mathsf{cov}}C_{\mathsf{sparse}}) + C_{\mathsf{est}}\sqrt{\frac{k_n^2 \log(p_n)}{n}}\right)^2 \cdot \frac{\pi}{2}||\widehat{T} - T||_\infty.$$

Furthermore, applying Lemma B.2, with probability at least $1 - \frac{1}{36p_n}$,

$$||\breve{x} - \overline{x}||_1 \leqslant \left(\sqrt{k_n}(1 + 2C_{\mathsf{cov}}C_{\mathsf{sparse}}) + C_{\mathsf{est}}\sqrt{\frac{k_n^2 \log(p_n)}{n}}\right)^2 \cdot \frac{\pi}{2}\sqrt{\frac{4\log(36p_n^3)}{n}}.$$

Finally, since $4\log(36p_n^3) \leqslant 36\log(p_n) \leqslant 4n$, where the last step holds by assumption in Theorem 4.2,

$$\begin{aligned}||\breve{x} - \overline{x}||_1 &\leqslant \sqrt{\frac{k_n^2 \log(p_n)}{n}} \cdot ((1 + 2C_{\mathsf{cov}}C_{\mathsf{sparse}}) + C_{\mathsf{est}})^2 \cdot \pi \\ &= C_1\sqrt{\frac{k_n^2 \log(p_n)}{n}} \text{ for } C_1 := ((1 + 2C_{\mathsf{cov}}C_{\mathsf{sparse}}) + C_{\mathsf{est}})^2 \cdot \pi \ .\end{aligned}$$

Next,

$$\begin{aligned}&||\breve{x} - x||_1 \leqslant ||\breve{x} - \overline{x}||_1 + ||\overline{x} - x||_1 \\ &\leqslant C_1\sqrt{\frac{k_n^2 \log(p_n)}{n}} + ||\overline{x} - x||_1 \\ &= C_1\sqrt{\frac{k_n^2 \log(p_n)}{n}} + \left\|(\breve{u}\breve{v}^\top - uv^\top) \circ \cos\left(\frac{\pi}{2}T\right)\right\|_1 \\ &\leqslant C_1\sqrt{\frac{k_n^2 \log(p_n)}{n}} + \left\|\breve{u}(\breve{v} - v)^\top \circ \cos\left(\frac{\pi}{2}T\right)\right\|_1 + \left\|(\breve{u} - u)v^\top \circ \cos\left(\frac{\pi}{2}T\right)\right\|_1 \\ &\leqslant C_1\sqrt{\frac{k_n^2 \log(p_n)}{n}} + ||\breve{u}||_1||\breve{v} - v||_1 \left\|\cos\left(\frac{\pi}{2}T\right)\right\|_\infty + ||\breve{u} - u||_1||v||_1 \left\|\cos\left(\frac{\pi}{2}T\right)\right\|_\infty \\ &\leqslant C_1\sqrt{\frac{k_n^2 \log(p_n)}{n}} + ||\breve{u}||_1||\breve{v} - v||_1 + ||\breve{u} - u||_1||v||_1 \\ &\leqslant C_1\sqrt{\frac{k_n^2 \log(p_n)}{n}} + \left(\sqrt{k_n}(1 + 2C_{\mathsf{cov}}C_{\mathsf{sparse}}) + C_{\mathsf{est}}\sqrt{\frac{k_n^2 \log(p_n)}{n}}\right) \cdot C_{\mathsf{est}}\sqrt{\frac{k_n^2 \log(p_n)}{n}} \\ &\qquad + \sqrt{k_n}(1 + 2C_{\mathsf{cov}}C_{\mathsf{sparse}}) \cdot C_{\mathsf{est}}\sqrt{\frac{k_n^2 \log(p_n)}{n}} \\ &\leqslant C_2\sqrt{\frac{k_n^3 \log(p_n)}{n}} \ ,\end{aligned}$$

where the next-to-last step applies (B.3) (assuming the bounds (3.1) in Assumption 3.3 hold), and where for the last step, $C_2 = C_1 + 2(1+2C_{\mathsf{cov}}C_{\mathsf{sparse}}) \cdot C_{\mathsf{est}} + C_{\mathsf{est}}^2$ and we use the assumption $n \geqslant k_n^2 \log(p)$.

Finally, noting that $\overline{x} - x = \mathsf{vec}\left((\breve{u}\breve{v}^\top - uv^\top) \circ \cos\left(\frac{\pi}{2}T\right)\right)$, we calculate the ℓ_1/ℓ_2 norm of this matrix:

$$\begin{aligned}
||\mathsf{mat}(\overline{x} - x)||_{\ell_1/\ell_2} &= \left\|(\breve{u}\breve{v}^\top - uv^\top) \circ \cos\left(\frac{\pi}{2}T\right)\right\|_{\ell_1/\ell_2} \\
&= \sum_j \left\|\left[(\breve{u}\breve{v}^\top - uv^\top) \circ \cos\left(\frac{\pi}{2}T\right)\right]_j\right\|_2 \\
&\leqslant \sum_j \left\|\left[\breve{u}\breve{v}^\top - uv^\top\right]_j\right\|_2 \cdot \left\|\cos\left(\frac{\pi}{2}T\right)\right\|_\infty \\
&\leqslant \sum_j \left\|\left[\breve{u}\breve{v}^\top - uv^\top\right]_j\right\|_2 \\
&\leqslant \sum_j ||\breve{u} \cdot (\breve{v}_j - v_j)||_2 + ||(\breve{u} - u) \cdot v_j||_2 \\
&= \sum_j ||\breve{u}||_2 \cdot |\breve{v}_j - v_j| + ||\breve{u} - u||_2 \cdot |v_j| \\
&= ||\breve{u}||_2 ||\breve{v} - v||_1 + ||\breve{u} - u||_2 ||v||_1 \ .
\end{aligned}$$

Next, applying (B.1), if the bounds (3.1) in Assumption 3.3 hold, we then have

$$\begin{aligned}
&||\mathsf{mat}(\overline{x} - x)||_{\ell_1/\ell_2} \\
&\leqslant \left(\sqrt{1 + C_{\mathsf{cov}}^2} + C_{\mathsf{est}}\sqrt{\frac{k_n \log(p_n)}{n}}\right) \cdot C_{\mathsf{est}} \cdot \sqrt{\frac{k_n^2 \log(p_n)}{n}} + C_{\mathsf{est}} \cdot \sqrt{\frac{k_n \log(p_n)}{n}} \cdot \sqrt{k_n}\sqrt{1 + C_{\mathsf{cov}}} \\
&\leqslant C_3 \sqrt{\frac{k_n^2 \log(p_n)}{n}} \ ,
\end{aligned}$$

where we define $C_3 = 2\sqrt{1 + C_{\mathsf{cov}}^2} \cdot C_{\mathsf{est}} + C_{\mathsf{est}}^2$ and use the assumption that $n \geqslant k_n^2 \log(p_n)$. □

LEMMA D.7. *Under the same assumptions and notation as Lemmas B.1 and B.5, for a universal constant* $C_{\mathsf{studentized}}$,

$$\mathbb{P}\left\{\left|x^\top(\widehat{\Sigma}_{h_1} - \Sigma_{h_1})x\right| \leqslant C_{\mathsf{studentized}}\sqrt{\frac{k_n^2 \log(p_n)}{n}}\right\} \geqslant 1 - \frac{1}{36 p_n} \ .$$

PROOF OF LEMMA D.7. By definition, we have $\Sigma_{h_1} = \mathsf{Var}(h_1(X))$ for

$$h_1(X) = \mathbb{E}\left[\text{sign}(X - X') \otimes \text{sign}(X - X') \mid X\right] \in \mathbb{R}^{p_n^2} .$$

Therefore, since x is fixed,

$$x^\top \Sigma_{h_1} x = x^\top \mathsf{Var}(h_1(X)) x = \mathsf{Var}(x^\top h_1(X)) = \mathsf{Var}(g_1(X)) = \nu_{g_1}^2 ,$$

where we recall that $g_1(X) = \mathbb{E}\left[g(X, X') \mid X\right]$ where we define the kernel

$$g(X, X') = \text{sign}(X - X')^\top \left(uv^\top \circ \cos\left(\frac{\pi}{2}T\right)\right) \text{sign}(X - X') = x^\top h(X, X') .$$

Define

$$\gamma(X, X', X'') = \frac{g(X, X')g(X, X'') + g(X', X)g(X', X'') + g(X'', X)g(X'', X')}{3} .$$

Note that $\gamma(X, X', X'')$ is a U-statistic of order 3, and that

$$\sup_{X, X'} |g(X, X')| \leqslant ||x||_1 \sup_{X, X'} ||h(X, X')||_\infty = ||x||_1 \leqslant C_0 k_n,$$

where the last step applies Lemma D.6. So,

$$||\gamma||_\infty := \sup_{X, X', X''} |\gamma(X, X', X'')| \leqslant \sup_{X, X'} |g(X, X')|^2 \leqslant C_0^2 k_n^2.$$

And,

$$\begin{aligned} \mathsf{Var}(\gamma) := \mathsf{Var}(\gamma(X, X', X'')) &\leqslant \mathbb{E}\left[\gamma(X, X', X'')^2\right] \leqslant \mathbb{E}\left[|g(X, X')|^4\right] \\ &\leqslant \mathbb{E}\left[|g(X, X')|^3\right] \cdot C_0 k_n \leqslant C_{\mathsf{moment}} \cdot C_0 k_n , \end{aligned}$$

where we use Lemma B.1 for the last bound.

Next, we have

$$\begin{aligned} \mathbb{E}\left[\gamma(X, X', X'')\right] &= \mathbb{E}\left[g(X, X')g(X, X'')\right] = \mathbb{E}\left[\mathbb{E}\left[g(X, X')g(X, X'')|X\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[g(X, X')|X\right]\mathbb{E}\left[g(X, X'')|X\right]\right] = \mathbb{E}\left[g_1(X)^2\right] . \end{aligned}$$

Therefore,

$$x^\top \Sigma_{h_1} x = \nu_{g_1}^2 = \mathsf{Var}(g_1(X)) = \mathbb{E}\left[\gamma(X, X', X'')\right] - \mathbb{E}\left[g_1(X)\right]^2 = \mathbb{E}\left[\gamma(X, X', X'')\right] - \mathbb{E}\left[g(X, X')\right]^2 .$$

Next, examining the definition of $\widehat{\Sigma}_{h_1}$, we obtain

$$x^\top \widehat{\Sigma}_{h_1} x = \frac{1}{n(n-1)^2}\left[\sum_{i \neq i' \neq i''} \gamma(X_i, X_{i'}, X_{i''}) + \sum_{i \neq i'} g(X_i, X_{i'})^2\right] - \left(\frac{1}{\binom{n}{2}} \sum_{i < i'} g(X_i, X_{i'})\right)^2 .$$

Therefore, using the fact that $|g(X,X')| \leqslant C_0 k_n$ always,

$$\left|x^\top \widehat{\Sigma}_{h_1} x - x^\top \Sigma_{h_1} x\right| \leqslant \left|\frac{1}{\binom{n}{3}} \sum_{i<i'<i''} \gamma(X_i, X_{i'}, X_{i''}) - \mathbb{E}[\gamma(X,X',X'')]\right| + \frac{C_0^2 k_n^2}{n-1} + \left|\left(\frac{1}{\binom{n}{2}} \sum_{i<i'} g(X_i, X_{i'})\right)^2 - \mathbb{E}[g(X,X')]^2\right| .$$

Now, using Bernstein's inequality for U-statistics (Peel, Anthoine and Ralaivola (2010, Theorem 2)), for any $\delta > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{\binom{n}{3}} \sum_{i<i'<i''} \gamma(X_i, X_{i'}, X_{i''}) - \mathbb{E}[\gamma(X,X',X'')]\right| > \sqrt{\frac{2\mathsf{Var}(\gamma)\log(2/\delta)}{(n/3)}} + \frac{2||\gamma||_\infty \log(2/\delta)}{3(n/3)}\right\} \leqslant \delta .$$

Therefore, with probability at least $1 - \frac{1}{72p_n}$,

$$\left|\frac{1}{\binom{n}{3}} \sum_{i<i'<i''} \gamma(X_i, X_{i'}, X_{i''}) - \mathbb{E}[\gamma(X,X',X'')]\right| \leqslant \sqrt{\frac{2C_{\mathsf{moment}} \cdot C_0 k_n \log(2 \cdot 72 p_n)}{(n/3)}} + \frac{2C_0^2 k_n^2 \log(2 \cdot 72 p_n)}{3(n/3)} \leqslant \sqrt{\frac{k_n^2 \log(p_n)}{n}} \cdot C' ,$$

where

$$C' = \sqrt{6C_{\mathsf{moment}} C_0 \log_2(144)} + 2C_0^2 \log_2(144) ,$$

and we use the assumption $n \geqslant k_n^2 \log(p_n)$ and $p_n \geqslant 2$. And, again using Bernstein's inequality for U-statistics, and using the fact that $|g(X,X')| \leqslant C_0 k_n$ always, with probability at least $1 - \frac{1}{72p_n}$,

$$\left|\frac{1}{\binom{n}{2}} \sum_{i<i'} g(X_i, X_{i'}) - \mathbb{E}[g(X,X')]\right| \leqslant \sqrt{\frac{2C_0^2 k_n^2 \log(2 \cdot 72 p_n)}{(n/2)}} + \frac{2C_0 k_n \log(2 \cdot 72 p_n)}{3(n/2)} \leqslant \sqrt{\frac{k_n^2 \log(p_n)}{n}} \cdot C'' ,$$

where

$$C'' = \sqrt{\frac{2C_0^2 \log_2(144)}{(1/2)}} + \frac{2C_0 \log_2(144)}{3/2} ,$$

and we use the assumption $n \geqslant k_n^2 \log(p_n)$ and $p_n \geqslant 2$. Therefore,

$$\left| \left(\frac{1}{\binom{n}{2}} \sum_{i<i'} g(X_i, X_{i'}) \right)^2 - \mathbb{E}[g(X,X')]^2 \right| \leqslant \left| \frac{1}{\binom{n}{2}} \sum_{i<i'} g(X_i, X_{i'}) - \mathbb{E}[g(X,X')] \right|^2 + \left| \frac{1}{\binom{n}{2}} \sum_{i<i'} g(X_i, X_{i'}) - \mathbb{E}[g(X,X')] \right| \cdot 2|\mathbb{E}[g(X,X')]| \leqslant C''' \sqrt{\frac{k_n^2 \log(p_n)}{n}} ,$$

where we set

$$C''' = C''^2 + 2C'' \cdot C_{\mathsf{moment}}^{1/3}$$

and again use $n \geqslant k_n^2 \log(p_n)$, and apply Lemma B.1 to bound $|\mathbb{E}\left[g(X,X')\right]|$. Combining everything, this proves that, with probability at least $1 - \frac{1}{36p_n}$,

$$\left| x^\top \widehat{\Sigma}_{h_1} x - x^\top \Sigma_{h_1} x \right| \leqslant \sqrt{\frac{k_n^2 \log(p_n)}{n}} \cdot C' + \frac{C_0^2 k_n^2}{n-1} + C''' \sqrt{\frac{k_n^2 \log(p_n)}{n}} .$$

Setting

$$C_{\mathsf{studentized}} = C' + C''' + 2C_0^2$$

and using the fact that $n \geqslant 2$ and $n \geqslant k_n^2 \log(p_n)$, we have

$$\left| x^\top \widehat{\Sigma}_{h_1} x - x^\top \Sigma_{h_1} x \right| \leqslant C_{\mathsf{studentized}} \sqrt{\frac{k_n^2 \log(p_n)}{n}} .$$

□

LEMMA D.8. *With probability at least* $1 - \frac{1}{9p_n}$,

$$||\widehat{\Sigma}_{h_1} - \Sigma_{h_1}||_\infty \leqslant 100 \sqrt{\frac{\log(p_n)}{n}} .$$

PROOF OF LEMMA D.8. From our definitions, we see that

$$\Sigma_{h_1} = \mathsf{Var}(h_1(X)) = \mathbb{E}\left[h_1(X) h_1(X)^\top\right] - \mathbb{E}[h_1(X)] \mathbb{E}[h_1(X)]^\top ,$$

and

$$\widehat{\Sigma}_{h_1} = \frac{1}{n} \sum_i \widehat{h}_1(X_i) \widehat{h}_1(X_i)^\top - \left(\frac{1}{n} \sum_i \widehat{h}_1(X_i) \right) \left(\frac{1}{n} \sum_i \widehat{h}_1(X_i) \right)^\top .$$

First, we bound $||\frac{1}{n}\sum_i \widehat{h}_1(X_i)\widehat{h}_1(X_i)^\top - \mathbb{E}\left[h_1(X)h_1(X)^\top\right]||_\infty$. We have

$$\text{(D.20)} \quad \left\|\frac{1}{n}\sum_i \widehat{h}_1(X_i)\widehat{h}_1(X_i)^\top - \mathbb{E}\left[h_1(X)h_1(X)^\top\right]\right\|_\infty \leqslant \\ \left\|\frac{1}{n}\sum_i \widehat{h}_1(X_i)\widehat{h}_1(X_i)^\top - \frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top\right\|_\infty + \\ \left\|\frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top - \mathbb{E}\left[h_1(X)h_1(X)^\top\right]\right\|_\infty .$$

We handle these two terms separately. First, we bound $||\frac{1}{n}\sum_i \widehat{h}_1(X_i)\widehat{h}_1(X_i)^\top - \frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top||_\infty$. For convenience we define $A := \frac{1}{n}\sum_i \widehat{h}_1(X_i)\widehat{h}_1(X_i)^\top$ and $B := \frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top$. Since A and B are both positive semidefinite matrices with ones on the diagonal, we have

$$\text{(D.21)} \qquad ||A-B||_\infty = \frac{1}{2}\max_{j,k\in[p_n^2]}\left|f_{jk}^\top(A-B)f_{jk}\right| ,$$

where $f_{jk} \in \mathbb{R}^{p_n^2}$ is the vector with $(f_{jk})_j = 1$, $(f_{jk})_k = -1$, and zeros elsewhere. Next we have

$$\left|f_{jk}^\top(A-B)f_{jk}\right| = \left|\sqrt{f_{jk}^\top A f_{jk}} - \sqrt{f_{jk}^\top B f_{jk}}\right| \cdot \left(\sqrt{f_{jk}^\top A f_{jk}} + \sqrt{f_{jk}^\top B f_{jk}}\right) \\ \leqslant 4\left|\sqrt{f_{jk}^\top A f_{jk}} - \sqrt{f_{jk}^\top B f_{jk}}\right| = \frac{4}{\sqrt{n}}\left|\sqrt{\sum_i(\widehat{h}_1(X_i)^\top f_{jk})^2} - \sqrt{\sum_i(h_1(X_i)^\top f_{jk})^2}\right| \\ \leqslant \frac{4}{\sqrt{n}}\sqrt{\sum_i\left((\widehat{h}_1(X_i) - h_1(X_i))^\top f_{jk}\right)^2} ,$$

where the first inequality follows from the fact that $||f_{jk}||_1 \leqslant 2$ while $||A||_\infty, ||B||_\infty \leqslant 1$, and the second inequality follows from the triangle inequality. Next, for each i and each j,k, observe that

$$\widehat{h}_1(X_i)^\top f_{jk} = \frac{1}{n-1}\sum_{i'\neq i}\left(\text{sign}(X_i - X_{i'}) \otimes \text{sign}(X_i - X_{i'})\right)^\top f_{jk} ,$$

which after conditioning on X_i, is a mean of $(n-1)$ i.i.d. variables, each taking values in $[-2,2]$ since $||f_{jk}||_1 \leqslant 2$. Furthermore, conditioning on X_i, we have $\mathbb{E}[\widehat{h}_1(X_i)] = h_1(X_i)$. Therefore, applying Hoeffding's lemma (see, for example, Lemma 2.6 in Massart, 2007), for each i,j,k, for any $t \in \mathbb{R}$,

$$\text{(D.22)} \qquad \mathbb{E}\left[\exp\left\{t\cdot(\widehat{h}_1(X_i) - h_1(X_i))^\top f_{jk}\right\}\right] \leqslant \exp\left\{\frac{2t^2}{n-1}\right\} .$$

Applying Lemma D.10 (stated below), then,

$$\mathbb{P}\left\{\frac{1}{n}\sum_i \left((\widehat{h}_1(X_i)-h_1(X_i))^\top f_{jk}\right)^2 > \frac{80}{n-1}\cdot(1+\log(27p_n^5))\right\} \leqslant \frac{1}{27p_n^5}\ .$$

Taking a union bound over all $j,k\in[p_n^2]$, and returning to (D.21), we then have

$$\mathbb{P}\left\{\left\|\frac{1}{n}\sum_i \widehat{h}_1(X_i)\widehat{h}_1(X_i)^\top - \frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top\right\|_\infty > 2\sqrt{\frac{80}{n-1}\cdot(1+\log(27p_n^5))}\right\} \leqslant \frac{1}{27p_n}\ .$$

Next we turn to the second term in (D.20). Since $||h_1(X)||_\infty \leqslant 1$ always, we see that for each $j,k\in[p_n]$,

$$\left(\frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top\right)_{jk}$$

is a mean of n i.i.d. terms, each taking values in $[-1,1]$. Applying Hoeffding's inequality, for each j,k,

$$\mathbb{P}\left\{\left|\left(\frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top - \mathbb{E}[h(X)h(X)^\top]\right)_{jk}\right| \geqslant t\right\} \leqslant 2e^{-nt^2/2}$$

for any $t\geqslant 0$. Setting $t=\sqrt{\frac{2\log(54p_n^3)}{n}}$, and taking a union bound, we see that

$$\mathbb{P}\left\{\left\|\frac{1}{n}\sum_i h_1(X_i)h_1(X_i)^\top - \mathbb{E}[h(X)h(X)^\top]\right\|_\infty \geqslant \sqrt{\frac{2\log(54p_n^3)}{n}}\right\}$$
$$\leqslant 2p_n^2\cdot e^{-n\left(\sqrt{\frac{54\log(p_n^3)}{n}}\right)^2/2} = \frac{1}{27p_n}\ .$$

Returning to (D.20), then, with probability at least $1-\frac{2}{27p_n}$,

(D.23)
$$||\frac{1}{n}\sum_i \widehat{h}_1(X_i)\widehat{h}_1(X_i)^\top - \mathbb{E}\left[h_1(X)h_1(X)^\top\right]||_\infty \leqslant 2\sqrt{\frac{80}{n-1}\cdot(1+\log(27p_n^3))} + \sqrt{\frac{2\log(54p_n^3)}{n}}\ .$$

This proves a bound on (D.20).

Next, to complete the proof, we bound

$$\left\|\left(\frac{1}{n}\sum_i \widehat{h}_1(X_i)\right)\left(\frac{1}{n}\sum_i \widehat{h}_1(X_i)\right)^\top - \mathbb{E}\left[h_1(X)\right]\mathbb{E}\left[h_1(X)\right]^\top\right\|_\infty\ .$$

We have

$$\left(\frac{1}{n}\sum_i \widehat{h}_1(X_i)\right)\left(\frac{1}{n}\sum_i \widehat{h}_1(X_i)\right)^\top - \mathbb{E}\left[h_1(X)\right]\mathbb{E}\left[h_1(X)\right]^\top$$
$$= \left(\frac{1}{n}\sum_i \widehat{h}_1(X_i)\right)\left(\frac{1}{n}\sum_i \widehat{h}_1(X_i) - \mathbb{E}\left[h_1(X)\right]\right)^\top - \left(\frac{1}{n}\sum_i \widehat{h}_1(X_i) - \mathbb{E}\left[h_1(X)\right]\right)\mathbb{E}\left[h_1(X)\right]^\top$$

and, since $||\frac{1}{n}\sum_i \widehat{h}_1(X_i)||_\infty, ||\mathbb{E}\left[h_1(X)\right]||_\infty \leqslant 1$, we therefore have

$$\left\|\left(\frac{1}{n}\sum_i \widehat{h}_1(X_i)\right)\left(\frac{1}{n}\sum_i \widehat{h}_1(X_i)\right)^\top - \mathbb{E}\left[h_1(X)\right]\mathbb{E}\left[h_1(X)\right]^\top\right\|_\infty \leqslant 2|| \left| || \frac{1}{n}\sum_i \widehat{h}_1(X_i) - \mathbb{E}\left[h_1(X)\right]\right\|_\infty .$$

For each sign $s \in \{\pm 1\}$, for each $j \in [p_n^2]$, writing $\mathbf{e}_j$ to denote the jth basis vector in $\mathbb{R}^{p_n^2}$, we have

$$\mathbb{E}\left[\exp\left\{t \cdot s \cdot \mathbf{e}_j^\top \left(\frac{1}{n}\sum_i \widehat{h}_1(X_i) - \mathbb{E}\left[h_1(X)\right]\right)\right\}\right]$$
$$\leqslant \frac{1}{n}\sum_i \mathbb{E}\left[\exp\left\{t \cdot s \cdot \mathbf{e}_j^\top \left(\widehat{h}_1(X_i) - \mathbb{E}\left[h_1(X)\right]\right)\right\}\right] \leqslant \exp\left\{\frac{t^2}{2(n-1)}\right\} ,$$

where the first inequality follows from the convexity of $x \mapsto e^x$, while the second applies Hoeffding's lemma, as in (D.22) above. Then,

$$\mathbb{P}\left\{s \cdot \mathbf{e}_j^\top \left(\widehat{h}_1(X_i) - \mathbb{E}\left[h_1(X)\right]\right) > \sqrt{\frac{2\log(54p_n^3)}{n-1}}\right\} \leqslant \frac{1}{54p^3} ,$$

and therefore taking a union bound over each $s \in \{\pm 1\}$ and each $j \in [p_n^2]$,

$$\mathbb{P}\left\{\left\|\frac{1}{n}\sum_i \widehat{h}_1(X_i) - \mathbb{E}\left[h_1(X)\right]\right\|_\infty > \sqrt{\frac{2\log(54p_n^3)}{n-1}}\right\} \leqslant \frac{1}{27p_n} .$$

Therefore, combining this with (D.23), with probability at least $1 - \frac{1}{9p_n}$,

$$||\widehat{\Sigma}_{h_1} - \Sigma_{h_1}||_\infty \leqslant 2\sqrt{\frac{80}{n-1} \cdot (1 + \log(27p_n^3))} + \sqrt{\frac{2\log(54p_n^3)}{n}} + 2\sqrt{\frac{2\log(54p_n^3)}{n-1}} \leqslant 100\sqrt{\frac{\log(p_n)}{n}} ,$$

where the last step uses the fact that $n, p_n \geqslant 2$. □

LEMMA D.9. *Let Σ_{h_1} be defined as in Assumption 3.4. For every $z \in \mathbb{R}^{p_n^2}$,*

$$z^\top \Sigma_{h_1} z \leqslant \lambda_{\max}(\Sigma) \cdot ||\mathsf{mat}(z)||^2_{\ell_1/\ell_2} ,$$

where $||\mathsf{mat}(z)||_{\ell_1/\ell_2}$ is defined as in the statement of Lemma D.6.

PROOF OF LEMMA D.9. Since the statement is deterministic, we can treat $M = \mathsf{mat}(z) \in \mathbb{R}^{p_n \times p_n}$ as fixed. Then $z = \mathsf{vec}(M)$ and

$$\begin{aligned}
&\mathsf{vec}(M)^\top \Sigma_{h_1} \mathsf{vec}(M) = \mathsf{Var}\left(\mathsf{vec}(M)^\top h_1(X)\right) \\
&= \mathsf{Var}\left(\mathsf{vec}(M)^\top \mathbb{E}[h(X,X') \mid X]\right) \\
&= \mathsf{Var}\left(\mathbb{E}[\mathsf{vec}(M)^\top h(X,X') \mid X]\right) \\
&\leqslant \mathsf{Var}\left(\mathsf{vec}(M)^\top h(X,X')\right) \quad \text{(by the law of total variance)} \\
&\leqslant \mathbb{E}\left[\left(\mathsf{vec}(M)^\top h(X,X')\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathsf{vec}(M)^\top \left(\mathrm{sign}(X-X') \otimes \mathrm{sign}(X-X')\right)\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathrm{sign}(X-X')^\top M\, \mathrm{sign}(X-X')\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_j \mathrm{sign}(X-X')^\top M_j \cdot \mathrm{sign}(X_j - X'_j)\right)^2\right] \quad \text{(where } M_j \text{ is the } j\text{th column of } M\text{)} \\
&\leqslant \mathbb{E}\left[\left(\sum_j \left|\mathrm{sign}(X-X')^\top M_j\right|\right)^2\right] \\
&= \sum_{jk} \mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top M_j\right| \cdot \left|\mathrm{sign}(X-X')^\top M_k\right|\right] \\
&\leqslant \sum_{jk} \sqrt{\mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top M_j\right|^2\right]} \cdot \sqrt{\mathbb{E}\left[\left|\mathrm{sign}(X-X')^\top M_k\right|^2\right]} \\
&= \sum_{jk} \sqrt{M_j^\top \mathbb{E}\left[\mathrm{sign}(X-X')\, \mathrm{sign}(X-X')^\top\right] M_j} \cdot \sqrt{M_k^\top \mathbb{E}\left[\mathrm{sign}(X-X')\, \mathrm{sign}(X-X')^\top\right] M_k} \\
&= \sum_{jk} \sqrt{M_j^\top T M_j} \cdot \sqrt{M_k^\top T M_k} \\
&\leqslant \sum_{jk} \sqrt{||M_j||_2^2 \cdot \lambda_{\max}(T)} \cdot \sqrt{||M_k||_2^2 \cdot \lambda_{\max}(T)} \\
&= \lambda_{\max}(T) \cdot \left(\sum_j ||M_j||_2\right)^2 .
\end{aligned}$$

Finally, by Wegkamp and Zhao (2013, Theorem 2.3), $\lambda_{\max}(T) \leqslant \lambda_{\max}(\Sigma)$. □

LEMMA D.10. *Let $v \in \mathbb{R}^p$ be a fixed vector and let $Z_1, \ldots, Z_n \in [-1,1]^p$ be random vectors, not necessarily independent, such that $v^\top(Z_i - \mathbb{E}[Z_i])$ is C-subgaussian for each i, that is,*

$$\mathbb{E}[\exp\{tv^\top(Z_i - \mathbb{E}[Z_i])\}] \leqslant \exp(Ct^2/2).$$

Then for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\frac{1}{n}\sum_i \left(v^\top(Z_i - \mathbb{E}[Z_i])\right)^2 \leqslant 20C(1+\log(1/\delta)) \; .$$

PROOF OF LEMMA D.10. For each i, by assumption,

$$\mathbb{E}\left[\exp\left\{t \cdot \frac{1}{\sqrt{C}} v^\top(Z_i - \mathbb{E}[Z_i])\right\}\right] \leqslant \exp\left\{\frac{t^2}{2}\right\} \; .$$

By Vershynin (2012, Lemma 5.5) (and tracking constants carefully in this Lemma), for each i,

$$\mathbb{E}\left[\exp\left\{\frac{1}{20C} \cdot \left(v^\top(Z_i - \mathbb{E}[Z_i])\right)^2\right\}\right] \leqslant e \; .$$

By the convexity of $x \mapsto e^x$, then,

$$\mathbb{E}\left[\exp\left\{\frac{1}{20C} \cdot \frac{1}{n}\sum_i \left(v^\top(Z_i - \mathbb{E}[Z_i])\right)^2\right\}\right] \leqslant \frac{1}{n}\sum_i \mathbb{E}\left[\exp\left\{\frac{1}{20C} \cdot \left(v^\top(Z_i - \mathbb{E}[Z_i])\right)^2\right\}\right] \leqslant e \; .$$

Therefore, we have

$$\begin{aligned}\mathbb{P}\left\{\frac{1}{n}\sum_i \left(v^\top(Z_i - \mathbb{E}[Z_i])\right)^2 > t\right\} &\leqslant \mathbb{E}\left[\exp\left\{\frac{1}{20C}\frac{1}{n}\sum_i \left(v^\top(Z_i - \mathbb{E}[Z_i])\right)^2 - \frac{1}{20C}t\right\}\right] \\ &\leqslant \exp\left\{1 - \frac{1}{20C}t\right\} \; .\end{aligned}$$

Setting $t = 20C(1+\log(1/\delta))$, then, we have proved the desired result. □

REFERENCES

Baraniuk, R. G., Davenport, M. A., and Wakin, M. B. (2008). A Simple Proof of the Restricted Isometry Property for Random Matrices. *Constructive Approximation* **28** 253–263.

Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *Ann. Stat.* **6** 417–421. MR0464359 (57 ##4290)

de la Pena, V. and Giné, E. (1999). *Decoupling: from dependence to independence*. Springer.

Lindskog, F., McNeil, A. and Schmock, U. (2003). Kendall's Tau for Elliptical Distributions. *Credit Risk* 149–156.

Massart, P. (2007). *Concentration inequalities and model selection. Lecture Notes in Mathematics* **1896**. Springer, Berlin Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR2319879 (2010a:62008)

Peel, T., Anthoine, S. and Ralaivola, L. (2010). Empirical Bernstein Inequalities for U-Statistics. In *Adv. Neural Inf. Process. Syst. 23* (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta, eds.) 1903–1911. Curran Associates, Inc.

Srebro, N. and Shraibman, A. (2005). Rank, trace-norm and max-norm. In *Learning theory. Lecture Notes in Comput. Sci.* **3559** 545–560. Springer, Berlin. MR2203286 (2006i:15002)

Sun, T. and Zhang, C.-H. (2012). Sparse Matrix Inversion with Scaled Lasso.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications* (Y. C. Eldar and G. Kutyniok, eds.). Cambridge University Press.

Wegkamp, M. and Zhao, Y. (2013). Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *ArXiv e-prints, arXiv:1305.6526.*

Address of the First and Second authors,
E-mail: rina@uchicago.edu; mkolar@chicagobooth.edu