Supplementary material to "Tuning-free Heterogeneous Inference in Massive Networks"

Zhao Ren[1], Yongjian Kang[2], Yingying Fan[2] and Jinchi Lv[2]

University of Pittsburgh[1] *and University of Southern California*[2]

This Supplementary Material contains a scalable HGSL algorithm with provable convergence in Section A, the proofs of Theorems 2.1–3.1, Theorem A.1 and Propositions 2.1–2.3 in Section C, as well as the proofs of key lemmas and additional technical details in Sections D and E, respectively. Additional computational cost comparison with existing methods is provided in Section B.

A Scalable HGSL algorithm with provable convergence

The tuning-free property of HGSL established in Section 3 provides a crucial step toward the scalability of our THI framework when one needs to analyze a large number of networks with massive number of nodes jointly. To further boost the scalability, we now introduce a new computational algorithm to solve the convex program of HGSL problem in (30) in a simple yet efficient fashion, which will be referred to as the HGSL algorithm hereafter for simplicity. As is common in regularization problems, we rescale each column of $\mathbf{X}^0_{*,-1}$ to have ℓ_2 norm $(n^{(t)})^{1/2}$ and denote by $\bar{\mathbf{X}}^0_{*,-1} = \mathrm{diag}\{\bar{\mathbf{X}}^{(1)}_{*,-1}, \cdots, \bar{\mathbf{X}}^{(k)}_{*,-1}\}$ the resulting new design matrix; that is, $\bar{\mathbf{X}}^0_{*,-1} = \mathbf{X}^0_{*,-1}\bar{D}_1^{-1/2}$ with the scaling matrix $\bar{D}_1$ given in Section 3. Let us consider another HGSL optimization problem

$$\hat{\bar{C}}^0_1 = \arg\min_{\beta^0\in\mathbb{R}^{(p-1)k}} \left\{ \sum_{t=1}^{k} \bar{Q}_t^{1/2}(\beta^{(t)}) + \lambda \sum_{l=2}^{p} \|\beta^0_{(l)}\| \right\}, \tag{A.1}$$

where $\bar{Q}_t(\beta^{(t)}) = \|X^{(t)}_{*,1} - \bar{\mathbf{X}}^{(t)}_{*,-1}\beta^{(t)}\|^2/n^{(0)}$ for $1 \le t \le k$ and the rest of the notation is defined similarly as in (30). In fact, the new HGSL optimization problem in (A.1) is closely related to the original HGSL optimization problem in (30), through a simple equation $\hat{\bar{C}}^0_1 = \bar{D}_1^{1/2}\hat{C}^0_1$

linking the minimizers of these two problems. Thus the problem of solving (30) reduces to that of solving (A.1).

To ease the presentation, we slightly abuse the notation and rewrite the new HGSL optimization problem (A.1) in a general form

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^{pk}} \left\{ (n^{(0)})^{-1/2} \sum_{t=1}^{k} \|Y^{(t)} - \mathbf{X}^{(t)} \beta^{(t)}\| + \lambda \sum_{l=1}^{p} \|\beta_{(l)}\| \right\}, \tag{A.2}$$

where $Y^{(t)} \in \mathbb{R}^{n^{(t)}}$, $\mathbf{X}^{(t)} \in \mathbb{R}^{n^{(t)} \times p}$, and $\beta^{(t)} \in \mathbb{R}^p$ are the response vector, the design matrix, and the regression coefficient vector, respectively, corresponding to the tth network for $1 \le t \le k$ with the pk-dimensional vector $\beta = ((\beta^{(1)})', \cdots, (\beta^{(k)})')'$ and $\beta_{(l)}$ a k-dimensional subvector of β formed by each lth component of $\beta^{(t)}$ with $1 \le t \le k$. Similarly we define the p-dimensional subvectors $\hat{\beta}^{(t)}$ of $\hat{\beta}$ with $1 \le t \le k$, and its k-dimensional subvectors $\hat{\beta}_{(l)}$ with $1 \le l \le p$.

So far our original HGSL optimization problem in (30) has been reduced to the general HGSL optimization problem in (A.2) with the same tuning-free choice of the parameter λ as discussed in Section 3 and the relationship between the two minimizers elucidated above. To solve the convex optimization problem in ($A.2$), we suggest a new scaled iterative thresholding algorithm. Our HGSL algorithm is designed specifically for the HGSL problem with convergence guarantees, motivated by the algorithm for the group square-root Lasso with homogeneous noises in [4] and a more general algorithm developed in [**?**]. In practice, to reduce the bias of the estimator $\hat{\beta}$ incurred by the regularization in ($A.2$) one can obtain the final estimate by a refit on the support of the computed sparse $\hat{\beta}$ using the ordinary least-squares estimator.

Our HGSL algorithm consists of two main steps, with the first step for rescaling and the second one for iteration. In the first step, we rescale the response vector, the design matrix, and the regularization parameter as

$$Y^{(t)}/K_0 \to Y^{(t)}, \ \mathbf{X}^{(t)}/K_0 \to \mathbf{X}^{(t)}, \ \lambda/K_0 \to \lambda \ \text{ for } 1 \le t \le k, \tag{A.3}$$

where $K_0 > 0$ is some preselected sufficiently large scalar. Clearly the solution to the optimization problem (A.2) remains the same after the rescaling specified in (A.3). Such step, however, reduces the norm of the design matrix, which can guarantee the convergence of the iterative algorithm as shown in Theorem A.1 later. We again slightly abuse the notation and still use

$Y^{(t)}$, $\mathbf{X}^{(t)}$, and λ to denote the response vector, the design matrix, and the regularization parameter after rescaling hereafter. In particular, the choice of $K_0 = \max_{1\leq t\leq k} \|\mathbf{X}^{(t)}\|_{\ell_2}$ with $\|\cdot\|_{\ell_2}$ denoting the spectral norm of a matrix, which is suggested by inequality (A.42) in the proof of Theorem A.1 in Section C.6 of the Supplementary Material, works well in our simulation studies.

In the second step, we solve iteratively the general HGSL optimization problem in (A.2) with the rescaled data matrix from the first step, and let $\beta(m)$ be the solution returned by the mth iteration for each integer $m \geq 0$. For the initial value $\beta(0)$, we set it as the zero vector in our numerical studies, which works well. Denote by $\beta(m)^{(t)}$ and $\beta(m)_{(l)}$ the subvectors of $\beta(m)$ similarly as in (A.2). For the $(m+1)$th iteration with input $\beta(m)$, we define $R(m) = ((R(m)^{(1)})', \cdots, (R(m)^{(k)})')' \in \mathbb{R}^{pk}$ with

$$R(m)^{(t)} = (\mathbf{X}^{(t)})' \left(\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\right) / \left[(n^{(0)})^{1/2} \left\|\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\right\|\right]$$

for $1 \leq t \leq k$, denote by $R(m)_{(l)}$ a k-dimensional subvector of $R(m)$ corresponding to the lth group for $1 \leq l \leq p$, and introduce a scaling factor $A(m) = \sum_{t=1}^{k} \left[(n^{(0)})^{1/2} \left\|\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\right\|\right]^{-1}$. Then we compute $\beta(m+1)$ as

$$\beta(m+1)_{(l)} = \overrightarrow{\Theta}\left(\beta(m)_{(l)} - \frac{R(m)_{(l)}}{A(m)}; \frac{\lambda}{A(m)}\right) \quad \text{for } 1 \leq l \leq p, \tag{A.4}$$

where $\overrightarrow{\Theta}$ is the multivariate soft-thresholding operator defined as

$$\overrightarrow{\Theta}(0; \lambda) = 0 \quad \text{and} \quad \overrightarrow{\Theta}(a; \lambda) = a\Theta(\|a\|\,; \lambda)/\|a\| \quad \text{for } a \neq \mathbf{0} \tag{A.5}$$

with $\Theta(t; \lambda) = \text{sgn}(t)(|t| - \lambda)_+$ representing the soft-thresholding rule. In practice, we stop the iteration when the difference between the solutions from two consecutive iterates falls below a prespecified small threshold for convergence.

Theorem A.1. *Assume that $\lambda > 0$ and $\min_{1\leq t\leq k} \inf_{\xi\in A^t} \left\|\mathbf{X}^{(t)}\xi - Y^{(t)}\right\| > c_0$ with $A^t = \{v\beta(m)^{(t)} + (1-v)\beta(m+1)^{(t)} : v \in [0,1], m = 0, 1, \cdots\}$ and $c_0 > 0$ some constant. Then for large enough K_0, the sequence of computed solutions $\beta(m)$ converges to the global optimum of the HGSL problem* (A.1).

Theorem A.1 justifies formally that our suggested scalable HGSL algorithm indeed enjoys provable convergence to the global optimum of our convex HGSL optimization problem. The scalability of the HGSL algorithm is rooted in both the tuning-free property and the simple iterative thresholding nature. It is also worth mentioning that a similar regularity condition to the one assumed in Theorem A.1 was imposed in [4] to prove the convergence of their algorithm for the group square-root Lasso with homogeneous noises. As mentioned before, in the end one can further apply a refit using the support of the computed sparse solution to obtain a final estimate with possibly reduced bias.

B Computational cost comparison with existing methods

We provide a comparison on the computational cost in Table 7 for simulation examples in Section 4.1.2. Since the computational cost of THI-ϕ_1 is almost identical to that of THI-ϕ_2, only the results for the latter are reported.

Table 7: Average computational costs of different methods in seconds.

	Setting 1 ($\times 10^0$)				Setting 2 ($\times 10^1$)				Setting 3 ($\times 10^2$)			
	THI	MPE	GGL	FGL	THI	MPE	GGL	FGL	THI	MPE	GGL	FGL
Model I	7.2	57.7	9.2	64.8	2.1	8.7	2.6	13.5	3.9	36.7	3.7	18.2
Model II	18.1	69.8	18.2	44.4	3.0	10.0	3.5	28.7	6.8	38.6	5.9	23.1

C Proofs of main results

C.1 Proofs of Theorem 2.1 and Proposition 2.1

The proofs of Theorems 2.1–2.2 and Propositions 2.1–2.2 rely on two key sets of results in Lemmas D.1 and D.2 in Sections D.1 and D.2, respectively, where we use the compact notation $[\ell]$ to denote the set $\{1, \cdots, \ell\}$ for any positive integer ℓ whenever there is no confusion. Our results are important consequences of Lemmas D.1 and D.2. Indeed, it holds that

$$\sum_{t=1}^{k}\left|\sqrt{n^{(t)}\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)}}\left(T_{n,k,1,2}^{(t)}-J_{n,k,1,2}^{(t)}\right)-V_{n,k,1,2}^{*(t)}\right|\leq T_1+T_2,$$

where

$$\begin{aligned}
T_1 &= \sum_{t=1}^{k}\sqrt{n^{(t)}\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)}}\left|T_{n,k,1,2}^{(t)}-J_{n,k,1,2}^{(t)}-\frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(E_{i,1}^{(t)}E_{i,2}^{(t)}-\mathbb{E}E_{i,1}^{(t)}E_{i,2}^{(t)}\right)\right|,\\
T_2 &= \sum_{t=1}^{k}\left|1-\sqrt{\frac{\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)}}{\omega_{2,2}^{(t)}\tilde{\omega}_{1,1}^{(t)}}}\right|\left|\sqrt{\frac{\omega_{2,2}^{(t)}\tilde{\omega}_{1,1}^{(t)}}{n^{(t)}}}\sum_{i=1}^{n^{(t)}}\left(E_{i,1}^{(t)}E_{i,2}^{(t)}-\mathbb{E}E_{i,1}^{(t)}E_{i,2}^{(t)}\right)\right|.
\end{aligned}$$

According to Lemma D.1, we have $|\hat{\omega}_{j,j}^{(t)}-\omega_{j,j}^{(t)}|\leq C'(\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}+s\frac{(k+\log p)}{n^{(0)}})=o(1)$ with probability at least $1-6p^{1-\delta}-2\delta_1$ uniformly for all $t\in[k]$ and $j=1,2$. Therefore, Condition 2.1 implies that all $\hat{\omega}_{j,j}^{(t)}$ are bounded from both below and above, which together with Lemma D.2 and $s\left(k+\log p\right)/n^{(0)}=o(1)$ leads to

$$T_1\leq C\left(s\frac{k+(\log p)}{\sqrt{n^{(0)}}}\right)$$

with probability at least $1-12p^{1-\delta}-2\delta_1$, where positive constant C depends on constants M, M_0, δ, C_1, C_2, and C_3.

It remains to upper bound term T_2. Note that Lemma D.1 together with Condition 2.1 implies that $\tilde{\omega}_{1,1}^{(t)}$ is bounded. In addition, Condition 2.1 also implies that $E_{i,1}^{(t)}E_{i,2}^{(t)}$, $i\in[n^{(t)}]$ are i.i.d. sub-exponential with bounded constant parameter. Consequently, Bernstein's inequality (see, e.g., Proposition 5.16, [35]) entails immediately that $\max_k|V_{n,k,1,2}^{*(t)}|<\sqrt{C'\log(k/\delta_1)}$ with probability at least $1-2\delta_1$, where positive constant C' depends on M only. Therefore, this fact and Lemma D.1 along with the union bound further yield with probability at least $1-6p^{1-\delta}-4\delta_1$ that

$$\begin{aligned}
T_2 &\leq \sqrt{C'\log(k/\delta_1)}\sum_{t=1}^{k}\left|1-\sqrt{\frac{\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)}}{\omega_{2,2}^{(t)}\tilde{\omega}_{1,1}^{(t)}}}\right|\\
&\leq C\sqrt{\log(k/\delta_1)}\left(\sum_{t=1}^{k}\left|\tilde{\omega}_{1,1}^{(t)}-\hat{\omega}_{1,1}^{(t)}\right|+\sum_{t=1}^{k}\left|\omega_{2,2}^{(t)}-\hat{\omega}_{2,2}^{(t)}\right|\right)\\
&\leq C\left(k\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}+s\frac{(k+(\log p))}{n^{(0)}}\right)\sqrt{\log(k/\delta_1)}\\
&\leq C(s\frac{k+(\log p)}{\sqrt{n^{(0)}}}),
\end{aligned}$$

where the second inequality follows from the fact that all $\hat{\omega}_{j,j}^{(t)}$, $\tilde{\omega}_{j,j}^{(t)}$, and $\omega_{j,j}^{(t)}$ are bounded from both below and above, the third inequality is due to Lemma D.1, and the last inequality follows from our sample size assumptions $\log(k/\delta_1) = O(s(1 + (\log p)/k))$ as well as $\log(k/\delta_1) = o(n^{(0)})$. The positive constant C above depends on constants M, δ, C_1, C_2, and C_3.

Combining the bounds of T_1 and T_2 above, we deduce that the following inequality holds with probability at least $1 - 12p^{1-\delta} - 4\delta_1$,

$$\sum_{t=1}^{k} \left| \sqrt{n^{(t)}\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)}} \left(T_{n,k,1,2}^{(t)} - J_{n,k,1,2}^{(t)}\right) - V_{n,k,1,2}^{*(t)} \right| \leq C\left(s\frac{k + \log p}{\sqrt{n^{(0)}}}\right), \tag{A.6}$$

where constant $C > 0$ depends only on M, M_0, δ, C_1, C_2, and C_3.

Aided with the key result in (A.6) above, the analysis of Theorem 2.1 is straightforward. Indeed we have

$$\begin{aligned}
&\left| \left(\sum_{t=1}^{k} n^{(t)}\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)} \left(T_{n,k,1,2}^{(t)} - J_{n,k,1,2}^{(t)}\right)^2 \right)^{1/2} - U_{n,k,1,2}^{*} \right| \\
\leq\ & \left[\sum_{t=1}^{k} \left(\sqrt{n^{(t)}\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)}} \left(T_{n,k,1,2}^{(t)} - J_{n,k,1,2}^{(t)}\right) - V_{n,k,1,2}^{*(t)} \right)^2 \right]^{1/2} \\
\leq\ & \sum_{t=1}^{k} \left| \sqrt{n^{(t)}\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)}} \left(T_{n,k,1,2}^{(t)} - J_{n,k,1,2}^{(t)}\right) - V_{n,k,1,2}^{*(t)} \right| \\
\leq\ & Cs\frac{k + (\log p)}{\sqrt{n^{(0)}}},
\end{aligned}$$

where the last inequality is due to (A.6). The remaining part of the proof for Theorem 2.1 follows easily.

Note that the chi distribution $U_{n,k,1,2}^{*}$ always has constant level standard deviation. Hence Proposition 2.1 follows from the fact that the error bound of $|U_{n,k,1,2} - U_{n,k,1,2}^{*}|$ is $o(1)$ with significant probability under the sample size assumption, which completes the proofs.

C.2 Proofs of Theorem 2.2 and Proposition 2.2

Theorem 2.2 is an immediate consequence of (A.6) established in Section C.1, since the left-hand side of (A.6) is an upper bound of the left-hand side of (20) regardless of what sign vector is picked.

Note that $V^*_{n,k,1,2}(\xi)$ follows distribution $N(0,k)$. The error bound of $|V_{n,k,1,2}(\xi)-V^*_{n,k,1,2}(\xi)|$ is negligible compared to the standard deviation of $V^*_{n,k,1,2}(\xi)$ with significant probability under the sample size assumption, that is, $s(k+(\log p))/\sqrt{n^{(0)}}=o(k^{1/2})$, which concludes the proofs of both Theorem 2.2 and Proposition 2.2.

C.3 Proof of Theorem 2.3

The first part of the analysis serves as a general tool for both the lower bound arguments in Theorem 2.3 and the proof of Theorem 2.4. It suffices to assume without loss of generality that the sample sizes of all k graphs are identical, that is, $n^{(1)}=\cdots=n^{(k)}=n^{(0)}$, noting that Condition 2.2 is valid under this setting. Consider a least favorable finite subset $\mathcal{G}=\{\Omega^0_1,\cdots,\Omega^0_m\}\subset\mathcal{A}$ in the alternative sets, where $\mathcal{A}=\mathcal{A}^{l2}(s,c'\sqrt{k^{1/2}/n^{(0)}})$ for Theorem 2.3 (1), $\mathcal{A}=\mathcal{A}^{l1}(s,c'\sqrt{k/n^{(0)}},\xi)$ for Theorem 2.3 (2), and $\mathcal{A}=\mathcal{A}^{l1}(s,c\sqrt{k/n^{(0)}},\xi)$ for Theorem 2.4. In addition, we consider one element in $\Omega^0_0\in\mathcal{N}(s)$. The choice of $\mathcal{G}$ and Ω^0_0 will be determined later.

Recall that each index denotes each of the k graphs, that is, $\Omega^0_h=\{\Omega^{(t)}_h\}_{t=1}^k$ for $h=0,\cdots,m$. Let $\mathbb{P}_h\equiv\mathbb{P}_{\Omega^0_h}$ denote the joint distribution of the observations when the true parameter is Ω^0_h. In other words, $\mathbb{P}_h$ is the joint distribution of $n^{(0)}$ copies of k graphs $\prod_{t=1}^k g_h^{(t)}(x_t)$, where $g_h^{(t)}(\cdot)$ is the density of $N(0,(\Omega^{(t)}_h)^{-1})$ for $t\in[k]$. We use $\mathbb{E}_v$ and f_h to denote the expectation under $\mathbb{P}_v$ and the density function under $\mathbb{P}_h$, respectively. Moreover, let $\bar{\mathbb{P}}=\frac{1}{m}\sum_{h=1}^m\mathbb{P}_h$ be the average measure of these joint distributions indexed by elements in $\mathcal{G}$. For any test ψ_0, we have

$$
\begin{aligned}
\sup_{v\in\mathcal{G}}\left(\mathbb{E}_0\psi_0+\mathbb{E}_v(1-\psi_0)\right) &\geq \inf_{\psi}\left(\sup_{v\in\mathcal{G}}\mathbb{E}_0\psi+\mathbb{E}_v(1-\psi)\right)\\
&\geq \inf_{\psi}\left(\mathbb{E}_0\psi+\bar{\mathbb{E}}(1-\psi)\right)\\
&= \left\|\mathbb{P}_0\wedge\bar{\mathbb{P}}\right\|,
\end{aligned}
$$

where $\|\mathbb{P}_0\wedge\bar{\mathbb{P}}\|$ is the total variation affinity between two measures. Therefore, if ψ_0 has significance level α it holds that

$$
\inf_{v\in\mathcal{A}}\mathbb{P}_v(\psi_0 \text{ rejects } H_{0,12})\leq\inf_{v\in\mathcal{G}}\mathbb{E}_v(\psi_0)\leq 1+\alpha-\left\|\mathbb{P}_0\wedge\bar{\mathbb{P}}\right\|. \tag{A.7}
$$

To show that for any given $\beta > \alpha$ and some constant $c > 0$, no test of significance level α satisfies (26), it is sufficient to prove that $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - (\beta - \alpha)/2$, which together with (A.7) implies that

$$\inf_{v \in \mathcal{A}} \mathbb{P}_v(\psi_0 \text{ rejects } H_{0,12}) \leq \beta - (\beta - \alpha)/2.$$

We will use this fact in the lower bound arguments in Theorem 2.3 and the proof of Theorem 2.4 with different constructions of $\mathcal{G}$ and Ω_0^0, and constant $c > 0$.

C.3.1 Proof of Theorem 2.3 (1)

To show that $\epsilon_n = \sqrt{k^{1/2}/n^{(0)}}$ is the separating rate, we first establish the lower bound (27) and then prove that our test ϕ_2 satisfies (26) with $\mathcal{A} = \mathcal{A}^{l2}(s, c\sqrt{k^{1/2}/n^{(0)}})$. With the aid of (A.7), it suffices to show that for fixed $\beta > \alpha$, there exists some constant $c' > 0$ such that $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - (\beta - \alpha)/2$ with appropriate choices of $\mathcal{G} \subset \mathcal{A} = \mathcal{A}^{l2}(s, c'\sqrt{k^{1/2}/n^{(0)}})$ and $\Omega_0^0 \in \mathcal{N}(s)$.

We define

$$\Omega_0^0 = \{\Omega_0^{(t)}\}_{t=1}^k \text{ such that } \Omega_0^{(1)} = \cdots = \Omega_0^{(k)} = I. \tag{A.8}$$

For simplicity, assume that $\tau\sqrt{k}$ is an integer with some small constant $\tau > 0$ to be determined later. Otherwise, $\tau\sqrt{k}$ can be replaced by its floor function $\lfloor \tau\sqrt{k} \rfloor$ in the analysis below. Then we construct a subset

$$\begin{aligned} \mathcal{G} \;=\; & \Big\{\Omega^0 = \{\Omega^{(t)}\}_{t=1}^k : \text{ there exists some } T \subset [k] \text{ with } |T| = \tau\sqrt{k} \text{ such that} \\ & \Omega^{(t)} = I \text{ for } t \notin T \text{ and } (\Omega_0^{(k)})^{-1} = I + (n^{(0)})^{-1/2} e_{12} \text{ for } t \in T\Big\}, \end{aligned} \tag{A.9}$$

where e_{12} is the matrix with the $(1, 2)$th and $(2, 1)$th entries being one and all other entries being zero. Therefore, there are $\binom{k}{\tau\sqrt{k}}$ distinct elements in $\mathcal{G}$ and thus $m = \binom{k}{\tau\sqrt{k}}$. It is easy to check that $\Omega_0^0 \in \mathcal{N}(s)$ and $\mathcal{G} \subset \mathcal{A}^{l2}(s, c'\sqrt{k^{1/2}/n^{(0)}})$ with $c' \equiv 2\sqrt{\tau}$, by noting that for each element in $\mathcal{G}$, $\|\omega_{h,12}^0\| = \frac{1}{1-1/n^{(0)}}\sqrt{\tau k^{1/2}/n^{(0)}}$. Hence we omit the details here. Lemma D.3 in Section D.3 helps us finish the proof of the lower bound, that is, (27).

It remains to show that the proposed chi-based test ϕ_2 satisfies (26), that is, with a sufficiently large $c > 0$, $\mathcal{A}(c) = \mathcal{A}^{l2}(s, c\sqrt{k^{1/2}/n^{(0)}})$, and $n^{(0)}$, we have

$$\inf_{v\in\mathcal{A}(c)} \mathbb{P}_v \left(U_{n,k,1,2} > z_k^{l2}(1-\alpha)\right) \geq \beta. \tag{A.10}$$

We show this fact in three steps. During the first two steps, we reduce the goal in (A.10) to a relatively simple one so that during the third step we are able to apply Chebyshev's inequality to finish our proof. Hereafter we use $C > 0$ to denote a generic constant. Before proceeding, note that under the assumptions of Proposition 2.1, including $\delta > 1$ and $\delta_1 = o(1)$, the last inequality of Lemma D.1 and Condition 2.1 entail that with probability $1 - o(1)$,

$$\max_{t\in[k],j=1,2} \left\{\left|\omega_{j,j}^{(t)}\left(\hat{\omega}_{j,j}^{(t)}\right)^{-1} - 1\right|\right\} \leq C\left(s\frac{(k+\log p)}{n^{(0)}} + \sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}\right), \tag{A.11}$$

$$J_{n,k,1,2}^{(t)}/\left(\omega_{1,2}^{(t)}/\left(\omega_{1,1}^{(t)}\omega_{2,2}^{(t)}\right)\right) \in (-1.1, -0.9), \tag{A.12}$$

where the second expression (A.12) follows from (A.11) and the definition of $J_{n,k,1,2}^{(t)}$ in (10).

Define $\bar{U}^2_{n,k,1,2} \equiv \sum_{t=1}^k n^{(t)}\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}(T_{n,k,1,2}^{(t)})^2$. Comparing $\bar{U}^2_{n,k,1,2}$ with the definition of $U^2_{n,k,1,2}$ in (11), we obtain that with probability $1 - o(1)$,

$$\frac{\bar{U}^2_{n,k,1,2}}{U^2_{n,k,1,2}} \leq \max_{t\in[k]} \frac{\omega_{1,1}^{(t)}}{\hat{\omega}_{1,1}^{(t)}}\frac{\omega_{2,2}^{(t)}}{\hat{\omega}_{2,2}^{(t)}} \leq 1 + C\left(s\frac{(k+\log p)}{n^{(0)}} + \sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}\right) \equiv \left(1+\eta_1^{l2}\right)^2,$$

where the second inequality follows from (A.11). Note that according to our assumptions, it holds that $\eta_1^{l2} \leq C(s\frac{(k+\log p)}{n^{(0)}} + \sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}) = o(1)$. Therefore, due to the union bound argument, to prove (A.10) it is sufficient to show

$$\inf_{v\in\mathcal{A}(c)} \mathbb{P}_v \left(\bar{U}_{n,k,1,2} > \left(1+\eta_1^{l2}\right)\cdot z_k^{l2}(1-\alpha)\right) > \beta. \tag{A.13}$$

We further reduce (A.13) in the second step. Denote by $\bar{V}^{*(t)}_{n,k,1,2} = \sqrt{\frac{\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}}{n^{(t)}}}\sum_{i=1}^{n^{(t)}}(E_{i,1}^{(t)}E_{i,2}^{(t)} - \mathbb{E}E_{i,1}^{(t)}E_{i,2}^{(t)})$ with $\mathbb{E}\bar{V}^{*(t)}_{n,k,1,2} = 0$. Lemma D.2 implies that with probability $1 - o(1)$,

$$\begin{aligned}
&\left|\bar{U}_{n,k,1,2} - \left(\sum_{t=1}^k\left[\sqrt{n^{(t)}\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}}J_{n,k,1,2}^{(t)} + \bar{V}^{*(t)}_{n,k,1,2}\right]^2\right)^{1/2}\right| \\
\leq\ & \sum_{t=1}^k \sqrt{n^{(t)}\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}}\left|T_{n,k,1,2}^{(t)} - J_{n,k,1,2}^{(t)} - \frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(E_{i,1}^{(t)}E_{i,2}^{(t)} - \mathbb{E}E_{i,1}^{(t)}E_{i,2}^{(t)}\right)\right| \\
\leq\ & C\left(s\frac{k+(\log p)}{n^{(0)}}\right) \equiv \eta_2^{l2}.
\end{aligned}$$

Therefore, by the union bound argument again, to show (A.13) it is sufficient to prove that

$$\inf_{v\in\mathcal{A}(c)} \mathbb{P}_v \left(\sum_{t=1}^{k} \left[\sqrt{n^{(t)}\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}} J_{n,k,1,2}^{(t)} + \bar{V}_{n,k,1,2}^{*(t)} \right]^2 > \left[\left(1+\eta_1^{l2}\right) \cdot z_k^{l2}(1-\alpha) + \eta_2^{l2} \right]^2 \right) > \beta.$$

We denote $\Xi_t \equiv (\sqrt{n^{(t)}\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}} J_{n,k,1,2}^{(t)} + \bar{V}_{n,k,1,2}^{*(t)})^2$, $t \in [k]$ to simplify our notation. Then it suffices to show

$$\inf_{v\in\mathcal{A}(c)} \mathbb{P}_v \left(\sum_{t=1}^{k} (\Xi_t - \mathbb{E}\Xi_t) > \left[\left(1+\eta_1^{l2}\right) \cdot z_k^{l2}(1-\alpha) + \eta_2^{l2} \right]^2 - \sum_{t=1}^{k} \mathbb{E}\Xi_t \right) > \beta. \tag{A.14}$$

In the third step, we need a careful analysis of both sides of (A.14). We first calculate the right-hand side term. According to the third result in Lemma E.1 in Section E with $z = \sqrt{2\log(1/\alpha)/k}$, it holds that $z_k^{l2}(1-\alpha) \le \sqrt{k}(1+\sqrt{2\log(1/\alpha)/k})$. By our sample size assumption $s^2\left(k+\log p\right)^2 = o(n^{(0)})$ and the definitions of η_1^{l2} and η_2^{l2}, we deduce that $s\frac{(k+\log p)}{n^{(0)}} \le C\left(n^{(0)}\right)^{-1/2}$, which further yields

$$\begin{aligned}
& \left[\left(1+\eta_1^{l2}\right) \cdot z_k^{l2}(1-\alpha) + \eta_2^{l2} \right]^2 \\
\le\ & \left(\sqrt{k}(1+\sqrt{2\log(1/\alpha)/k}) \left(1 + C\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}\right) + C\left(n^{(0)}\right)^{-1/2} \right)^2 \\
\le\ & \left(\sqrt{k}(1+\sqrt{2\log(1/\alpha)/k}) \left(1 + C\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}\right) \right)^2 + C\sqrt{\frac{k}{n^{(0)}}} \\
\le\ & \left(k + 3\sqrt{2k\log(1/\alpha)}\right) \left(1 + C\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}\right) \\
\le\ & k + 4\sqrt{2k\log(1/\alpha)}.
\end{aligned} \tag{A.15}$$

Next we calculate a lower bound of $\sum_{t=1}^{k} \mathbb{E}\Xi_t$. By the definition of $\bar{V}_{n,k,1,2}^{*(t)}$ and the joint Gaussianity of $E_{i,1}^{(t)}$ and $E_{i,2}^{(t)}$, we have $\mathbb{E}(\bar{V}_{n,k,1,2}^{*(t)})^2 = 1 + (\omega_{1,2}^{(t)})^2/(\omega_{2,2}^{(t)}\omega_{1,1}^{(t)})$. This fact together with (A.12) results in

$$\begin{aligned}
\sum_{t=1}^{k} \mathbb{E}\Xi_t &= \sum_{t=1}^{k} \mathbb{E} \left[\sqrt{n^{(t)}\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}} J_{n,k,1,2}^{(t)} + \bar{V}_{n,k,1,2}^{*(t)} \right]^2 \\
&\ge \sum_{t=1}^{k} \mathbb{E}\left(\bar{V}_{n,k,1,2}^{*(t)}\right)^2 + Cn^{(0)} \sum_{t=1}^{k} \frac{\left(\omega_{1,2}^{(t)}\right)^2}{\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}} \\
&\ge k + Cn^{(0)} \left\|\omega_{1,2}^{0}\right\|^2.
\end{aligned} \tag{A.16}$$

We can further upper bound the variance of $\sum_{t=1}^{k}(\Xi_t - \mathbb{E}\Xi_t)$ by the joint Gaussianity of $E_{i,1}^{(t)}$ and $E_{i,2}^{(t)}$,

$$\operatorname{var}\left(\sum_{t=1}^{k}(\Xi_t - \mathbb{E}\Xi_t)\right) \leq C\left(k + n^{(0)}\left\|\omega_{1,2}^{0}\right\|^2\right). \tag{A.17}$$

Expressions (A.15) and (A.16) imply that under alternative $\mathcal{A}(c) = \mathcal{A}^{l2}(s, c\sqrt{k^{1/2}/n^{(0)}})$ with a sufficiently large $c > 0$, the right-hand side of (A.14) is negative, that is,

$$\begin{aligned} & \left[\left(1+\eta_1^{l2}\right)\cdot z_k^{l2}(1-\alpha) + \eta_2^{l2}\right]^2 - \sum_{t=1}^{k}\mathbb{E}\Xi_t \\ < \ & -Cn^{(0)}\left\|\omega_{1,2}^{0}\right\|^2 + 4\sqrt{2k\log(1/\alpha)} \\ \leq \ & -cC\sqrt{k} + 4\sqrt{2k\log(1/\alpha)} < 0. \end{aligned} \tag{A.18}$$

Therefore, by Chebyshev's inequality we obtain that for any $v \in \mathcal{A}(c)$,

$$\begin{aligned} & \mathbb{P}_v\left(\sum_{t=1}^{k}(\Xi_t - \mathbb{E}\Xi_t) \leq \left[\left(1+\eta_1^{l2}\right)\cdot z_k^{l2}(1-\alpha) + \eta_2^{l2}\right]^2 - \sum_{t=1}^{k}\mathbb{E}\Xi_t\right) \\ \leq \ & \operatorname{var}\left(\sum_{t=1}^{k}(\Xi_t - \mathbb{E}\Xi_t)\right) \Big/ \left(Cn^{(0)}\left\|\omega_{1,2}^{0}\right\|^2\right)^2 < 1-\beta, \end{aligned}$$

where the first inequality follows from (A.18) and the last inequality follows from (A.17) and a large constant $c > 0$. Thus (A.14) is an immediate consequence, which completes the proof for the first part of Theorem 2.3.

C.3.2 Proof of Theorem 2.3 (2)

To prove that $\epsilon_n = \sqrt{k/n^{(0)}}$ is the separating rate, we first show the lower bound (27) and then establish that the proposed linear functional-based test ϕ_1 satisfies (26). Without loss of generality, assume that the sign vector $\xi = (1, \cdots, 1)'$ and denote by $\mathcal{A}^{l1}(s, c'\sqrt{k/n^{(0)}}) \equiv \mathcal{A}^{l1}(s, c'\sqrt{k/n^{(0)}}, \xi)$ for short. Facilitated with (A.7), it suffices to show that for fixed $\beta > \alpha$, there exists some constant $c' > 0$ such that $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - (\beta - \alpha)/2$ with appropriate choices of $\mathcal{G} \subset \mathcal{A} = \mathcal{A}^{l1}(s, c'\sqrt{k/n^{(0)}})$ and $\Omega_0^0 \in \mathcal{N}(s)$.

The constructions of $\mathcal{G}$ and Ω_0^0 are straightforward. There is only one element in $\mathcal{G}$, that is, $m = 1$ and $\bar{\mathbb{P}} = \mathbb{P}_1$. We define $\Omega_0^0 = \{\Omega_0^{(t)}\}_{t=1}^{k}$ such that $\Omega_0^{(1)} = \cdots = \Omega_0^{(k)} = I$ and set $\Omega_1^0 = \{\Omega_1^{(t)}\}_{t=1}^{k}$ such that $(\Omega_0^{(1)})^{-1} = \cdots = (\Omega_0^{(k)})^{-1} = I + (\tau/\sqrt{n^{(0)}k})e_{12}$, where $\tau > 0$

is some small constant to be determined later and e_{12} is the matrix with all but two entries being zero and the $(1,2)$th and $(2,1)$th entries being one. It is easy to see that $\Omega_0^0 \in \mathcal{N}(s)$. In addition, it is easy to check that all eigenvalues of Ω_1^0 are in $[M^{-1}, M]$, and thus $\Omega_1^0 \in \mathcal{F}(s)$ since $\tau/\sqrt{n^{(0)}k} = o(1)$. Note that $\left\|\omega_{1,12}^0\right\|_1 = \frac{\tau}{1-\tau^2/(n^{(0)}k)}\sqrt{k/n^{(0)}}$. Therefore, we have shown that $\Omega_1^0 \in \mathcal{A}^{l1}(s, c'\sqrt{k/n^{(0)}})$ with $c' \equiv 2\tau$, where we have used $\tau^2/(n^{(0)}k) < 1/2$.

To finish the lower bound (27), it remains to prove $\|\mathbb{P}_0 \wedge \mathbb{P}_1\| > 1 - (\beta - \alpha)/2$. A similar argument to that in the proof of Lemma D.4 in Section D.4 (see expression (A.68)) implies that it is sufficient to show that the χ^2 divergence between $\mathbb{P}_0$ and $\mathbb{P}_1$ is small enough, that is, $\Delta = \int f_1^2/f_0 - 1 < (\beta - \alpha)^2$. By the simple constructions of Ω_0^0 and Ω_1^0, together with the χ^2 divergence of two Gaussian distributions (see expression (A.69)), it can be easily checked that $\Delta = (1 - \tau^2/(n^{(0)}k))^{-n^{(0)}k} - 1$. Since $\tau^2/(n^{(0)}k) < 1/2$, we can further bound the χ^2 divergence as

$$\Delta \leq (1 + 2\tau^2/(n^{(0)}k))^{n^{(0)}k} - 1 \leq \exp(2\tau^2) - 1.$$

Therefore, by picking τ small enough we deduce that $\Delta < (\beta - \alpha)^2$ and thus $\|\mathbb{P}_0 \wedge \mathbb{P}_1\| > 1 - (\beta - \alpha)/2$, which finishes the proof of (27).

It remains to show that the proposed linear functional-based test ϕ_1 satisfies (26), that is, with a sufficiently large $c > 0$, $\mathcal{A}(c) = \mathcal{A}^{l1}(s, c\sqrt{k/n^{(0)}})$, and $n^{(0)}$, it holds that

$$\inf_{v \in \mathcal{A}(c)} \mathbb{P}_v \left(\frac{V_{n,k,1,2}(\xi)}{\sqrt{k}} < z(\alpha) \right) \geq \beta.$$

Observe that under the assumptions of Proposition 2.2, including $\delta > 1$ and $\delta_1 = o(1)$, the last three inequalities of Lemma D.1 and Condition 2.1 lead to the following two facts: (i) $\omega_{1,1}^{(t)}(\hat{\omega}_{1,1}^{(t)})^{-1} = 1 + o(1)$ and $\omega_{2,2}^{(t)}(\hat{\omega}_{2,2}^{(t)})^{-1} = 1 + o(1)$ uniformly over $t \in [k]$, and (ii) $\sum_{t=1}^k |(\omega_{1,1}^{(t)})^{1/2} - (\tilde{\omega}_{1,1}^{(t)})^{1/2}| = o(1)$ with probability $1 - o(1)$, which will be used later in our analysis.

With bound (20) in Theorem 2.2 and the definition of $V_{n,k,1,2}(\xi)$ in (19), along with a union bound argument, we see that it suffices to prove that as $n^{(0)} \to \infty$,

$$\inf_{v \in \mathcal{A}(c)} \mathbb{P}_v \left(\frac{V_{n,k,1,2}^*}{\sqrt{k}} < z(\alpha) - \eta_1^{l1} - \Psi \right) > \beta, \tag{A.19}$$

where $\Psi \equiv \sum_{t=1}^k \xi_t (n^{(t)}\hat{\omega}_{2,2}^{(t)}\hat{\omega}_{1,1}^{(t)})^{1/2} J_{n,k,1,2}^{(t)}/\sqrt{k}$ and $\eta_1^{l1} \equiv Cs\,(k + \log p)\,/\sqrt{n^{(0)}k}$. To deal with the bias issue of $V_{n,k,1,2}^*$, we define $\bar{V}_{n,k,1,2}^* = \sum_{t=1}^k \xi_t (\frac{\omega_{2,2}^{(t)}\omega_{1,1}^{(t)}}{n^{(t)}})^{1/2} \sum_{i=1}^{n^{(t)}} (E_{i,1}^{(t)}E_{i,2}^{(t)} - \mathbb{E}E_{i,1}^{(t)}E_{i,2}^{(t)})$

and reduce the problem of showing (A.19) to that of showing

$$\inf_{v\in\mathcal{A}(c)}\mathbb{P}_v\left(\frac{\bar{V}^*_{n,k,1,2}}{\sqrt{k}}<z(\alpha)-\eta_1^{l1}-\eta_2^{l1}-\Psi\right)>\beta, \tag{A.20}$$

where $\eta_2^{l1}\equiv(V^*_{n,k,1,2}-\bar{V}^*_{n,k,1,2})/\sqrt{k}$.

We claim that $\eta_1^{l1}+\eta_2^{l1}=o_P(1)$ and $z(\alpha)-\Psi<0$ under alternative $v\in\mathcal{A}(c)$ with a sufficiently large constant $c>0$. Note that by definition $\mathbb{E}\bar{V}^*_{n,k,1,2}=0$. Hence according to Chebyshev's inequality and the union bound argument, it suffices to prove that

$$\operatorname{var}(\bar{V}^*_{n,k,1,2}/\sqrt{k})/\left|z(\alpha)-\Psi\right|^2<(1-\beta)/2$$

under alternative $v\in\mathcal{A}(c)$. We finish the proof by showing $\eta_1^{l1}+\eta_2^{l1}=o_P(1)$, $\operatorname{var}(\bar{V}^*_{n,k,1,2}/\sqrt{k})\le 2$ and that $\Psi<0$ can be arbitrarily small under alternative $v\in\mathcal{A}(c)$ by picking a sufficiently large constant $c>0$, respectively. Indeed, assuming that the latter two facts hold, $\operatorname{var}(\bar{V}^*_{n,k,1,2}/\sqrt{k})/\left|z(\alpha)-\Psi\right|^2<(1-\beta)/2$ follows as an immediate consequence, which will finish our proof.

In particular, fact (i) above entails that $J^{(t)}_{n,k,1,2}=(-1+o(1))\omega^{(t)}_{1,2}/(\omega^{(t)}_{1,1}\omega^{(t)}_{2,2})$ uniformly over $t\in[k]$, following from the definition of $J^{(t)}_{n,k,1,2}$ in (10). Since the sign vector of $\omega^0_{1,2}$ is encoded in ξ, the boundedness of $\omega^{(t)}_{1,1}\omega^{(t)}_{2,2}$ and $(\hat{\omega}^{(t)}_{2,2}\hat{\omega}^{(t)}_{1,1})^{1/2}$ for $t\in[p]$ (due to Condition 2.1 and fact (i) above) further implies that with some constant $C>0$,

$$\Psi\le -C\sqrt{\frac{n^{(0)}}{k}}\left\|\omega^0_{1,2}\right\|_1\le -Cc,$$

under alternative $\mathcal{A}(c)=\mathcal{A}^{l1}(s,c\sqrt{k/n^{(0)}})$. Therefore, with a sufficiently large constant $c>0$, $\Psi<0$ is smaller than any pre-determined negative constant.

Note that by the independence and joint Gaussianity of $E^{(t)}_{1,1}$ and $E^{(t)}_{1,2}$, we have $\operatorname{var}(\bar{V}^*_{n,k,1,2}/\sqrt{k})=k^{-1}\sum_{t=1}^k\operatorname{var}(E^{(t)}_{1,1}E^{(t)}_{1,2})\omega^{(t)}_{2,2}\omega^{(t)}_{1,1}\le 2$. Thus it remains to show that $\eta_1^{l1}+\eta_2^{l1}=o_P(1)$. It is easy to see that $\eta_1^{l1}=Cs\left(k+\log p\right)/\sqrt{n^{(0)}k}=o(1)$ by our sample size assumption. In addition, we have with probability at least $1 - 2\delta_1^{-10}$,

$$\begin{aligned}
|\eta_2^{l1}| &= \left| \sum_{t=1}^{k} \frac{\xi_t}{\sqrt{k}} \cdot \sqrt{\frac{\omega_{2,2}^{(t)}}{n^{(t)}}} \sum_{i=1}^{n^{(t)}} \left(E_{i,1}^{(t)} E_{i,2}^{(t)} - \mathbb{E} E_{i,1}^{(t)} E_{i,2}^{(t)} \right) \left(\sqrt{\omega_{1,1}^{(t)}} - \sqrt{\tilde{\omega}_{1,1}^{(t)}} \right) \right| \\
&\leq \frac{1}{\sqrt{k}} \max_{t \in [k]} \left| \sqrt{\frac{\omega_{2,2}^{(t)}}{n^{(t)}}} \sum_{i=1}^{n^{(t)}} \left(E_{i,1}^{(t)} E_{i,2}^{(t)} - \mathbb{E} E_{i,1}^{(t)} E_{i,2}^{(t)} \right) \right| \cdot \sum_{t=1}^{k} \left| \sqrt{\omega_{1,1}^{(t)}} - \sqrt{\tilde{\omega}_{1,1}^{(t)}} \right| \\
&< C \sqrt{\frac{\log(k/\delta_1)}{k}} \cdot \sum_{t=1}^{k} \left| \sqrt{\omega_{1,1}^{(t)}} - \sqrt{\tilde{\omega}_{1,1}^{(t)}} \right|, \qquad \text{(A.21)}
\end{aligned}$$

where the first inequality is due to Hölder's inequality and the second one follows from Bernstein's inequality (see, e.g., Proposition 5.16, [35]). It follows from fact (ii) above and inequality (A.21) that $\eta_2^{l1} = o_P(1)$, in view of $\delta_1 = o(1)$. Therefore, we have shown (A.20), which further entails that ϕ_1 satisfies (26) with a sufficiently large constant $c > 0$. This concludes the proof for the second part of Theorem 2.3.

C.4 Proof of Theorem 2.4

The general tool established in (A.7) of Section C.3 plays a key role in our analysis. We need to show that for any fixed $\beta > \alpha$ and $c > 0$, there is no test of significance level α satisfying (26) with $\mathcal{A} = \mathcal{A}^{l1}(s, c\sqrt{k/n^{(0)}}, \xi)$. In light of (A.7), it is sufficient to show that as long as $s^2 k^{-1}(k + \log p) > Cn^{(0)}$ for some sufficiently large positive constant C depending on M_1, μ, and c, we have

$$\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - (\beta - \alpha)/2$$

with appropriate choices of $\mathcal{G} \subset \mathcal{A}^{l1}(s, c\sqrt{k/n^{(0)}}, \xi)$ and $\Omega_0^0 \in \mathcal{N}(s)$. Since the lower bound does not depend on the choice of the sign vector ξ, hereafter we assume $\xi = (1, \cdots, 1)'$ without loss of generality.

To construct $\mathcal{G}$ and Ω_0^0, it suffices to assume that the k precision matrices are identical for each Ω_h^0, $h = 0, \cdots, m$, that is, $\Omega_h^{(1)} = \cdots = \Omega_h^{(k)}$. Therefore, we only need to construct $\Omega_h^{(1)}$ for each h. The element in null is defined as $\Omega_0^{(1)} = I$ which gives

$$\Omega_0^0 = \{\Omega_0^{(t)}\}_{t=1}^{k} \text{ with } \Omega_0^{(1)} = \cdots = \Omega_0^{(k)} = I. \qquad \text{(A.22)}$$

Besides, we construct a subset

$$\mathcal{G} = \left\{\Omega^0 = \{\Omega^{(t)}\}_{t=1}^k : \Omega^{(1)} = \cdots = \Omega^{(k)} = (I + aH)^{-1}\text{for some } H \in \mathcal{H}\right\} \tag{A.23}$$

with $a = \sqrt{\tau \frac{1+(\log p)/k}{n^{(0)}}}$ and $\tau > 0$ some small constant to be determined later. Here $\mathcal{H}$ is the set containing the collection of all $p \times p$ symmetric matrices with exactly $s - 1$ elements equal to 1 between the third and the last elements of the first and second rows (and hence columns by symmetry) and the rest all zeros. We also assume that for each $H \in \mathcal{H}$, the supports of the first row and the second row are identical. Clearly, there are $\binom{p-2}{s-1}$ distinct elements in $\mathcal{G}$ and thus $m = \binom{p-2}{s-1}$. To finish the proof, we need to show two claims: (i) $\mathcal{G} \subset \mathcal{A}^{l1}(s, c\sqrt{k/n^{(0)}}, \xi)$ and $\Omega_0^0 \in \mathcal{N}(s)$ and (ii) $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - (\beta - \alpha)/2$.

The desired result in claim (ii) is established in Lemma D.4 in Section D.4. Thus it remains to prove the desired result in claim (i). It is easy to see that $\Omega_0^0 \in \mathcal{N}(s)$ since all k precision matrices are identity matrices and particularly $\omega_{0,12}^0 = \mathbf{0}$. For each $\Omega_h^0 \in \mathcal{G}$, we can check that Ω_h^0 satisfies the sparsity assumption $\max_a \sum_{b \neq a} 1\{\omega_{h,ab}^0 \neq \mathbf{0}\} \leq s$. Moreover, the largest and smallest eigenvalues of $\Omega_h^{(1)}$ are

$$\lambda_{\max}(\Omega_h^{(1)}) = \frac{1 + \sqrt{2(s-1)a^2}}{1 - 2(s-1)a^2}, \lambda_{\min}(\Omega_h^{(1)}) = \frac{1 - \sqrt{2(s-1)a^2}}{1 - 2(s-1)a^2},$$

respectively, with all remaining eigenvalues being ones. Under the assumption that $s(1 + (\log p)/k)/n^{(0)} = o(1)$, we see that $2(s-1)a^2$ is sufficiently small and hence all eigenvalues are bounded between $1/M$ and M, which satisfies Condition 2.1. Therefore, we have shown that $\mathcal{G} \subset \mathcal{F}(s)$.

Finally, some elementary algebra implies that for each $\Omega_h^0 \in \mathcal{G}$, we always have $\omega_{h,12}^{(1)} = \frac{(s-1)a^2}{1-2(s-1)a^2}$. As a result, it holds that

$$\left\|\omega_{h,12}^0\right\|_1 = \frac{k(s-1)a^2}{1 - 2(s-1)a^2} \geq 2k(s-1)\tau \left(\frac{1 + (\log p)/k}{n^{(0)}}\right) > c\sqrt{\frac{k}{n^{(0)}}},$$

where the first inequality follows from $2(s-1)a^2 < 1/2$ and the last inequality is due to the main assumption of Theorem 2.4, that is, $s^2 k^{-1}(k + \log p)^2 > Cn^{(0)}$ with $C \equiv (c/\tau)^2$. Therefore, we have shown $\mathcal{G} \subset \mathcal{A}^{l1}(s, c\sqrt{k/n^{(0)}}, \xi)$, which completes the proof.

C.5 Proof of Theorem 3.1

Without loss of generality, we only prove the results for the case of $j = 1$. This is because by symmetry, the results remain valid for any $j \in [p]$. Hereafter, we follow the same notation for any vector $u \in \mathbb{R}^{(p-1)k}$ as defined for C_1^0, that is, $u^{(t)}$ denotes its subvector corresponding to the tth class and $u_{(l)}$ represents its subvector corresponding to the lth group. The purpose of normalization diagonal matrices $\bar{D}_1^{(t)}$ for our method HGSL defined in (30) is to obtain a tight universal regularization parameter λ by normalizing each column of $\mathbf{X}_{*,-1}^{(t)}$ such that its ℓ_2 norm is $\sqrt{n^{(t)}}$, that is, $\bar{\mathbf{X}}_{*,-1}^{(t)} = \mathbf{X}_{*,-1}^{(t)}(\bar{D}_1^{(t)})^{-1/2}$.

Define $\bar{C}_1^{(t)} = (\bar{D}_1^{(t)})^{1/2} C_1^{(t)}$ and $\hat{\bar{C}}_1^{(t)} = (\bar{D}_1^{(t)})^{1/2} \hat{C}_1^{(t)}$, and correspondingly $\bar{C}_1^0$ and $\hat{\bar{C}}_1^0$. Then the right-hand side of (29) becomes $\bar{\mathbf{X}}_{*,-1}^0 \bar{C}_1^0 + E_{*,1}^0$ and the method HGSL in (30) becomes

$$\hat{\bar{C}}_1^0 = \arg\min_{\beta^0 \in \mathbb{R}^{k(p-1)}} \left\{ \sum_{t=1}^{k} \bar{Q}_t^{1/2}(\beta^{(t)}) + \lambda \sum_{l=2}^{p} \|\beta_{(l)}^0\| \right\}$$

with $\bar{Q}_t(\beta^{(t)}) = \frac{1}{n^{(0)}} \|X_{*,1}^{(t)} - \bar{\mathbf{X}}_{*,-1}^{(t)} \beta^{(t)}\|^2$. Our main results involve the difference $\Delta = \hat{C}_1^0 - C_1^0$. In what follows, we establish all results in terms of $\bar{\Delta} = \hat{\bar{C}}_1^0 - \bar{C}_1^0 = \left(\bar{D}_1^0\right)^{1/2} \Delta$. It is worth mentioning that this does not affect our results much. Indeed, our Condition 2.1 and the fact of $\mathbf{X}_{*,l}^{(t)\prime} \mathbf{X}_{*,l}^{(t)} / \sigma_{ll}^{(t)} \sim \chi^2(n^{(t)})$, together with an application of Lemma E.1 and the union bound, entail that with probability at least $1 - 2pk \exp(-n^{(0)}/32)$, all diagonal entries of $\bar{D}_1^0$ are bounded from below by $M/2$ and from above by $3M/2$ simultaneously. Therefore, Δ and $\bar{\Delta}$ are of the same order componentwise and globally. To make it rigorous, define an event

$$\mathcal{E}_{scale} = \left\{ \mathbf{X}_{*,l}^{(t)\prime} \mathbf{X}_{*,l}^{(t)} / n^{(t)} \in [1/(2M), 3M/2] \text{ for all } t \in [k], l \in [p] \right\}$$

and it holds that $\mathbb{P}\{\mathcal{E}_{scale}\} \geq 1 - 2pk \exp(-n^{(0)}/32)$.

We begin with introducing the group-wise restricted eigenvalue (gRE) condition proposed by [?] and [22], which is needed to establish our main results. Recall that the true coefficient vector C_1^0 is a group sparse vector. Denote by $T = \{l : \bar{C}_{1(l)}^0 \neq \mathbf{0}\}$. By the definition of the maximum node degree given in (14) and the relationship between $\bar{C}_1^{(t)}$ and $\Omega^{(t)}$, we deduce that $|T| \leq s$, where $|\cdot|$ stands for the cardinality of a set.

Definition C.1. *The group-wise restricted eigenvalue (gRE) condition holds on the design matrix* $\bar{\mathbf{X}}^0_{*,-1}$ *if*

$$gRE(\xi, T) \equiv \inf_{u \neq 0} \left\{ \frac{\left\|\bar{\mathbf{X}}^0_{*,-1} u\right\|}{\sqrt{n^{(0)}}\,\|u\|} : u \in \Psi(\xi, T) \right\} > 0,$$

where $\Psi(L, T) = \{u \in \mathbb{R}^{(p-1)k} : \sum_{j \in T^c} \|u_{(j)}\| \leq L \sum_{j \in T} \|u_{(j)}\|\}$ *is a cone.*

The above gRE condition is an extension of the restricted eigenvalue (RE) condition for the regular Lasso proposed in [?], in which the ℓ_1 norm is replaced by the group-wise ℓ_1 norm. It was also assumed in [22] to tackle the usual group Lasso as a direct condition. [?] derived the gRE condition based on some incoherence condition. However, to the best of our knowledge, there is no existing result for the random design matrix satisfying the gRE condition in the literature. In this paper, we first establish that the gRE condition is satisfied with large probability as a consequence of our assumptions in Lemma D.5 presented in Section D.5.

We would like to mention that other commonly used conditions on the design matrix $\bar{\mathbf{X}}^0_{*,-1}$, including the group-wise compatibility condition [4] and the group-wise cone invertibility factor condition [27], can also be applied here. In fact, the group-wise compatibility condition $\kappa(\xi, T) > 0$ is a natural consequence of the gRE condition thanks to the Cauchy-Schwarz inequality, since

$$\begin{aligned}
\kappa(\xi, T) &\equiv \inf_{u \neq \mathbf{0}} \left\{ \frac{\sqrt{|T|}\left\|\bar{\mathbf{X}}^0_{*,-1} u\right\|}{\sqrt{n^{(0)}} \sum_{l \in T} \left\|u_{(l)}\right\|} : u \in \Psi(\xi, T) \right\} \\
&\geq \inf_{u \neq \mathbf{0}} \left\{ \frac{\left\|\bar{\mathbf{X}}^0_{*,-1} u\right\|}{\sqrt{n^{(0)}} \left(\sum_{l \in T} \left\|u_{(l)}\right\|^2\right)^{1/2}} : u \in \Psi(\xi, T) \right\} \\
&\geq \inf_{u \neq 0} \left\{ \frac{\left\|\bar{\mathbf{X}}^0_{*,-1} u\right\|}{\sqrt{n^{(0)}}\,\|u\|} : u \in \Psi(\xi, T) \right\} = gRE(\xi, T). \qquad \text{(A.24)}
\end{aligned}$$

In particular, on the event $\mathcal{E}_{1,gRE}$ defined in Lemma D.5 it holds that

$$\kappa(\xi, T) > \min_{l,t} \{(n^{(t)}/\mathbf{X}^{(t)\prime}_{*,l}\mathbf{X}^{(t)}_{*,l})^{1/2}\}/(2M)^{1/2}.$$

As discussed in Section 3, the analysis of Theorem 3.1 relies critically on the event $\mathcal{B}_1$ defined in (31), which guides us to pick a sharp parameter λ. Lemma D.6 in Section D.6 implies that our explicit choice of λ is indeed feasible. Thus with the aid of Lemmas D.5 and D.6, we are now ready to establish our main results in the following two steps.

Step 1. It follows from the definition that

$$\sum_{t=1}^{k}\left(\bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)})-\bar{Q}_t^{1/2}(\bar{C}_1^{(t)})\right) \le \lambda\sum_{l=2}^{p}\left(\left\|\bar{C}_{1(l)}^{0}\right\|-\left\|\hat{\bar{C}}_{1(l)}^{0}\right\|\right)$$
$$\le \lambda\left(\sum_{l\in T}\left\|\bar{\Delta}_{(l)}\right\|-\sum_{l\in T^c}\left\|\bar{\Delta}_{(l)}\right\|\right). \tag{A.25}$$

Observe that $\frac{\partial \bar{Q}_t^{1/2}(\bar{C}_1^{(t)})}{\partial\beta^{(t)}}=\frac{-1}{\sqrt{n^{(0)}}}\frac{\bar{\mathbf{X}}_{*,-1}^{(t)\prime}E_{*,1}^{(t)}}{\|E_{*,1}^{(t)}\|}$. By the convexity of $\bar{Q}_t^{1/2}(\cdot)$, we have

$$\sum_{t=1}^{k}\left(\bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)})-\bar{Q}_t^{1/2}(\bar{C}_1^{(t)})\right) \ge -\frac{1}{\sqrt{n^{(0)}}}\sum_{t=1}^{k}\frac{\bar{\Delta}^{(t)\prime}\bar{\mathbf{X}}_{*,-1}^{(t)\prime}E_{*,1}^{(t)}}{\left\|E_{*,1}^{(t)}\right\|}$$
$$\ge -\left(\sum_{l=2}^{p}\left\|\bar{\Delta}_{(l)}\right\|\right)\cdot\max_{2\le l\le p}\frac{\left\|\bar{D}_{E1}^{-1/2}\bar{\mathbf{X}}_{*,(l)}^{0\prime}E_{*,1}^{0}\right\|}{\sqrt{n^{(0)}}}$$
$$\ge -\lambda\frac{\xi-1}{\xi+1}\sum_{l=2}^{p}\left\|\bar{\Delta}_{(l)}\right\|, \tag{A.26}$$

where the last inequality follows from Lemma D.6. Combining inequalities (A.25) and (A.26), we obtain

$$-\lambda\frac{\xi-1}{\xi+1}\sum_{l=2}^{p}\left\|\bar{\Delta}_{(l)}\right\|\le\lambda\left(\sum_{l\in T}\left\|\bar{\Delta}_{(l)}\right\|-\sum_{l\in T^c}\left\|\bar{\Delta}_{(l)}\right\|\right),$$

which entails that

$$\sum_{l\in T^c}\left\|\bar{\Delta}_{(l)}\right\|\le\xi\sum_{l\in T}\left\|\bar{\Delta}_{(l)}\right\|.$$

Hence, we have shown that $\bar{\Delta}\in\Psi(\xi,T)$.

Step 2. We will make use of the following facts with $\zeta_t=\bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)})+\bar{Q}_t^{1/2}(\bar{C}_1^{(t)})$

$$\bar{Q}_t(\hat{\bar{C}}_1^{(t)})-\bar{Q}_t(\bar{C}_1^{(t)}) = \frac{\left\|\bar{\mathbf{X}}_{*,-1}^{(t)}\bar{\Delta}^{(t)}\right\|^2}{n^{(0)}}-\frac{2\bar{\Delta}^{(t)\prime}\bar{\mathbf{X}}_{*,-1}^{(t)\prime}E_{*,1}^{(t)}}{n^{(0)}}, \tag{A.27}$$
$$\bar{Q}_t(\hat{\bar{C}}_1^{(t)})-\bar{Q}_t(\bar{C}_1^{(t)}) = \left(\bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)})-\bar{Q}_t^{1/2}(\bar{C}_1^{(t)})\right)\cdot\zeta_t, \tag{A.28}$$
$$\sum_{l\in T}\left\|\bar{\Delta}_{(l)}\right\| \le \frac{\sqrt{s}\left\|\bar{\mathbf{X}}_{*,-1}^{0}\bar{\Delta}\right\|}{\sqrt{n^{(0)}}\kappa(\xi,T)}, \tag{A.29}$$
$$\sum_{t=1}^{k}\frac{\bar{\Delta}^{(t)\prime}\bar{\mathbf{X}}_{*,-1}^{(t)\prime}E_{*,1}^{(t)}}{n^{(0)}\zeta_t} \le \left(\sum_{l=2}^{p}\left\|\bar{\Delta}_{(l)}\right\|\right)\max_{2\le l\le p}\frac{\left\|\bar{D}_{E1}^{-1/2}\bar{\mathbf{X}}_{*,(l)}^{0\prime}E_{*,1}^{0}\right\|}{\sqrt{n^{(0)}}}\cdot\max_{t\in[k]}\frac{\left\|E_{*,1}^{(t)}\right\|}{\zeta_t\sqrt{n^{(0)}}}, \tag{A.30}$$

where the first two facts are due to some elementary algebra and the third one follows from the definition of $\kappa(\xi,T)$ in (A.24) and the fact of $\bar{\Delta}\in\Psi(\xi,T)$ proved in Step 1. It follows from

(A.27) and (A.28) that

$$\sum_{t=1}^{k}(\bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)}) - \bar{Q}_t^{1/2}(\bar{C}_1^{(t)})) = \sum_{t=1}^{k}\left(\frac{\left\|\bar{\mathbf{X}}_{*,-1}^{(t)}\bar{\Delta}^{(t)}\right\|^2}{n^{(0)}\zeta_t} - \frac{2\bar{\Delta}^{(t)\prime}\bar{\mathbf{X}}_{*,-1}^{(t)\prime}E_{*,1}^{(t)}}{n^{(0)}\zeta_t}\right).$$

Therefore, by (A.30), Lemma D.6, and the fact of $\max_{t\in[k]}(\frac{\|E_{*,1}^{(t)}\|}{\zeta_t\sqrt{n^{(0)}}}) \leq 1$, we further deduce that

$$\begin{aligned}
\sum_{t=1}^{k}\frac{\left\|\bar{\mathbf{X}}_{*,-1}^{(t)}\bar{\Delta}^{(t)}\right\|^2}{n^{(0)}\zeta_t} &\leq \sum_{t=1}^{k}\left(\bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)}) - \bar{Q}_t^{1/2}(\bar{C}_1^{(t)})\right) + 2\lambda\frac{\xi-1}{\xi+1}\left(\sum_{l=2}^{p}\|\bar{\Delta}_{(l)}\|\right) \\
&\leq \lambda\left(\sum_{l\in T}\|\bar{\Delta}_{(l)}\| - \sum_{l\in T^c}\|\bar{\Delta}_{(l)}\|\right) + 2\lambda\frac{\xi-1}{\xi+1}\left(\sum_{l=2}^{p}\|\bar{\Delta}_{(l)}\|\right) \\
&= \lambda\left(\frac{3\xi-1}{\xi+1}\sum_{l\in T}\|\bar{\Delta}_{(l)}\| + \frac{\xi-3}{\xi+1}\sum_{l\in T^c}\|\bar{\Delta}_{(l)}\|\right) \\
&\leq \lambda\left(\frac{3\xi-1}{\xi+1} + \xi\frac{(\xi-3)_+}{\xi+1}\right)\sum_{l\in T}\|\bar{\Delta}_{(l)}\| \\
&\leq \frac{\sqrt{s}\left\|\bar{\mathbf{X}}_{*,-1}^{0}\bar{\Delta}\right\|}{\sqrt{n^{(0)}}\kappa(\xi,T)}\lambda\left(\frac{3\xi-1}{\xi+1} + \xi\frac{(\xi-3)_+}{\xi+1}\right), \qquad \text{(A.31)}
\end{aligned}$$

where the second inequality is due to (A.25) and the last one follows from the definition of $\kappa(\xi, T)$ in (A.24).

Lemma D.7 presented in Section D.7 provides a natural constant level upper bound for the fitted prediction error. Then we can lower bound the left-hand side of (A.31) according to Lemma D.7 on the event $\mathcal{E}_{1,up}$ as

$$\sum_{t=1}^{k}\frac{\left\|\bar{\mathbf{X}}_{*,-1}^{(t)}\bar{\Delta}^{(t)}\right\|^2}{n^{(0)}\zeta_t} \geq \frac{1}{\sqrt{6MM_0}}\sum_{t=1}^{k}\frac{\left\|\bar{\mathbf{X}}_{*,-1}^{(t)}\bar{\Delta}^{(t)}\right\|^2}{n^{(0)}}.$$

Thus combining (A.31) with the above inequality leads to

$$\frac{\left\|\bar{\mathbf{X}}_{*,-1}^{0}\bar{\Delta}\right\|}{\sqrt{n^{(0)}}} \leq \frac{\sqrt{s}}{\kappa(\xi,T)}\lambda\left(\frac{3\xi-1}{\xi+1} + \xi\frac{(\xi-3)_+}{\xi+1}\right)\sqrt{6MM_0}.$$

In summary, by (A.24) and with our well specified λ, on the event $\mathcal{E}_{scale}\cap\mathcal{E}_{1,up}\cap\mathcal{B}_1\cap\mathcal{E}_{1,gRE}$ there exists some constant $C > 0$ such that

$$\sum_{t=1}^{k}\frac{\left\|\mathbf{X}_{*,-1}^{(t)}\left(\hat{C}_1^{(t)} - C_1^{(t)}\right)\right\|^2}{n^{(0)}} = \sum_{t=1}^{k}\frac{\left\|\bar{\mathbf{X}}_{*,-1}^{0}\left(\hat{\bar{C}}_1^{(t)} - \bar{C}_1^{(t)}\right)\right\|^2}{n^{(0)}} \leq Cs\frac{k+\log p}{n^{(0)}}.$$

Moreover, since $\hat{\bar{C}}_1^0 - \bar{C}_1^0 = \bar{\Delta} \in \Psi(\xi, T)$, by the definitions of $\kappa(\xi, T)$ in (A.24) and the gRE condition in Definition C.1 we can derive the following two inequalities from the expression above

$$\begin{aligned}
\sum_{l=2}^{p} \left\| \hat{C}_{1(l)}^0 - C_{1(l)}^0 \right\| &\leq \sqrt{2M} \sum_{l=2}^{p} \left\| \hat{\bar{C}}_{1(l)}^0 - \bar{C}_{1(l)}^0 \right\| \leq Cs \left(\frac{k + \log p}{n^{(0)}} \right)^{1/2}, \\
\left\| \hat{C}_1^0 - C_1^0 \right\| &\leq \sqrt{2M} \left\| \hat{\bar{C}}_1^0 - \bar{C}_1^0 \right\| \leq C \left(s \frac{k + \log p}{n^{(0)}} \right)^{1/2},
\end{aligned}$$

noting that conditional on the event $\mathcal{E}_{scale}$, Δ is less than or equal to$\sqrt{2M}\bar{\Delta}$ componentwise. Finally we conclude the proof by an application of the union bound argument using Lemmas D.5–D.7.

C.6 Proof of Theorem A.1

The main idea of the proof consists of two parts. First we prove that our suggested algorithm in Section A has a unique guaranteed point of convergence β^*. Then we show that such a point is the global optimum of the HGSL optimization problem (A.1).

Step 1: Convergence of $\beta(m)$. Let us denote by

$$F(\beta) = (n^{(0)})^{-1/2} \sum_{t=1}^{k} \|Y^{(t)} - \mathbf{X}^{(t)} \beta^{(t)}\| + \lambda \sum_{l=1}^{p} \|\beta_{(l)}\| \tag{A.32}$$

the objective function in (A.2) which is a reformulation of (A.1) in simplified notation. To prove the desired result, we first construct a surrogate function and show that the updating rule optimizes the surrogate function. Then we characterize the relationship between the objective function and the surrogate function, which entails that the limit of $\beta(m)$ from the mth iteration of the algorithm is in fact optimal for our objective function.

We begin with introducing a surrogate function

$$\begin{aligned}
G(\beta, \gamma) = &\sum_{t=1}^{k} \frac{\|Y^{(t)} - \mathbf{X}^{(t)} \beta^{(t)}\|}{\sqrt{n^{(0)}}} + \frac{1}{2} \sum_{t=1}^{k} \frac{1}{\sqrt{n^{(0)}} \|Y^{(t)} - \mathbf{X}^{(t)} \beta^{(t)}\|} \|\gamma - \beta\|^2 + \lambda \sum_{l=1}^{p} \|\gamma_{(l)}\| \\
&+ \sum_{t=1}^{k} \frac{1}{\sqrt{n^{(0)}} \|Y^{(t)} - \mathbf{X}^{(t)} \beta^{(t)}\|} (\gamma^{(t)} - \beta^{(t)})' (\mathbf{X}^{(t)})' (\mathbf{X}^{(t)} \beta^{(t)} - Y^{(t)}),
\end{aligned} \tag{A.33}$$

where $\gamma^{(t)}$ and $\gamma_{(l)}$ are the subvectors of γ defined similarly as $\beta^{(t)}$ and $\beta_{(l)}$, respectively. It is easy to see that

$$F(\beta) = G(\beta, \beta). \tag{A.34}$$

Denote by $R^{(t)} = (n^{(0)})^{-1/2}(\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)})/\|Y^{(t)} - \mathbf{X}^{(t)}\beta^{(t)}\|$ and $R = ((R^{(1)})', \cdots, (R^{(k)})')'$. Then we can rewrite the last term in (A.33) as

$$\sum_{t=1}^{k} \frac{1}{\sqrt{n^{(0)}}\,\|Y^{(t)} - \mathbf{X}^{(t)}\beta^{(t)}\|}(\gamma^{(t)} - \beta^{(t)})'(\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)}) = (\gamma - \beta)'R.$$

Thus given a fixed β, minimizing the above surrogate function G over γ is equivalent to minimizing the following objective function formed by the last three terms of G in (A.33) with respect to γ

$$\frac{1}{2}A\,\|\gamma - \beta\|^2 + \lambda \sum_{l=1}^{p} \|\gamma_{(l)}\| + (\gamma - \beta)'R,$$

where we denote by $A = \sum_{t=1}^{k}(n^{(0)})^{-1/2}\|Y^{(t)} - \mathbf{X}^{(t)}\beta^{(t)}\|^{-1}$. The optimization problem above is further equivalent to minimizing the following objective function with respect to γ

$$\frac{1}{2}\left\|\gamma - \beta + \frac{R}{A}\right\|^2 + \frac{\lambda}{A}\sum_{l=1}^{p}\|\gamma_{(l)}\|. \tag{A.35}$$

Combining the above results yields that for any given β, the minimizer of the objective function $G(\beta, \gamma)$ defined in (A.33) with respect to γ is the same as that of the objective function given in (A.35).

We now set $\beta = \beta(m)$ and correspondingly define the vector $R(m)$ and the scalar $A(m)$ similarly as R and A, respectively, with $\beta(m)$ in place of β. We update $\beta(m+1)$ as the minimizer of the objective function (A.35) with respect to γ given $\beta = \beta(m)$. Thus $\beta(m+1)$ is also the minimizer of $G(\beta(m), \gamma)$ with respect to γ. Since the optimization problem in (A.35) is separable, it can be rewritten in the following form

$$\sum_{l=1}^{p}\left\{\frac{1}{2}\left\|\beta_{(l)} - \frac{R_{(l)}}{A} - \gamma_{(l)}\right\|^2 + \frac{\lambda}{A}\|\gamma_{(l)}\|\right\}. \tag{A.36}$$

In view of (A.36), the optimization problem in (A.35) can be solved componentwise by minimizing each of the p summands above. In particular, the resulting solution admits an explicit form and we obtain by Lemmas 1 and 2 in [**?**] that $\beta(m+1)$ is given by

$$\beta(m+1)_{(l)} = \overrightarrow{\Theta}\left(\beta(m)_{(l)} - \frac{R(m)_{(l)}}{A(m)}; \frac{\lambda}{A(m)}\right), \qquad l \in [p], \tag{A.37}$$

where $R(m)_{(l)}$ is a subvector of $R(m)$ defined in a similar way to $\beta_{(l)}$ as a subvector of β and $\overrightarrow{\Theta}(\cdot;\cdot)$ is the multivariate soft-thresholding operator introduced in (A.5). Thus, it follows from (A.34) that

$$G(\beta(m), \beta(m+1)) \leq G(\beta(m), \beta(m)) = F(\beta(m)). \tag{A.38}$$

Let us consider the function $(n^{(0)})^{-1/2} \left\|Y^{(t)} - \mathbf{X}^{(t)}\gamma^{(t)}\right\|$ with respect to $\gamma^{(t)}$. Some routine calculations show that its gradient is given by

$$(n^{(0)})^{-1/2} \left\|Y^{(t)} - \mathbf{X}^{(t)}\gamma^{(t)}\right\|^{-1} (\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\gamma^{(t)} - Y^{(t)}) \tag{A.39}$$

and its Hessian matrix is

$$\begin{aligned}(n^{(0)})^{-1/2} \left\|Y^{(t)} - \mathbf{X}^{(t)}\gamma^{(t)}\right\|^{-1} (\mathbf{X}^{(t)})'\mathbf{X}^{(t)} - (n^{(0)})^{-1/2} \left\|Y^{(t)} - \mathbf{X}^{(t)}\gamma^{(t)}\right\|^{-3} \\ \cdot (\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\gamma^{(t)} - Y^{(t)})(\mathbf{X}^{(t)}\gamma^{(t)} - Y^{(t)})'\mathbf{X}^{(t)} \\ \leq (n^{(0)})^{-1/2} \left\|Y^{(t)} - \mathbf{X}^{(t)}\gamma^{(t)}\right\|^{-1} (\mathbf{X}^{(t)})'\mathbf{X}^{(t)},\end{aligned} \tag{A.40}$$

where $\leq$ means that the difference between the matrices on the right-hand side and the left-hand side of the inequality is positive semidefinite. Thus for any given β and γ, an application of the Taylor expansion of the function $(n^{(0)})^{-1/2} \left\|Y^{(t)} - \mathbf{X}^{(t)}\gamma^{(t)}\right\|$ at the point $\beta^{(t)}$ to the first order with the Lagrange remainder, together with (A.39)–(A.40), results in

$$\begin{aligned}\sum_{t=1}^{k} \frac{\left\|Y^{(t)} - \mathbf{X}^{(t)}\beta^{(t)}\right\|}{\sqrt{n^{(0)}}} + \sum_{t=1}^{k} \frac{1}{\sqrt{n^{(0)}} \left\|\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)}\right\|} (\gamma^{(t)} - \beta^{(t)})'(\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)}) \\ - \sum_{t=1}^{k} \frac{\left\|Y^{(t)} - \mathbf{X}^{(t)}\gamma^{(t)}\right\|}{\sqrt{n^{(0)}}} \\ \geq \sum_{t=1}^{k} -\frac{(\gamma^{(t)} - \beta^{(t)})'(\mathbf{X}^{(t)})'\mathbf{X}^{(t)}(\gamma^{(t)} - \beta^{(t)})}{2\sqrt{n^{(0)}} \left\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\right\|},\end{aligned} \tag{A.41}$$

where $\xi^{(t)}$ lies on the line segment connecting $\beta^{(t)}$ and $\gamma^{(t)}$ for each $t \in [k]$.

For now set $\beta = \beta(m)$ and $\gamma = \beta(m+1)$. Then it follows from (A.33) and (A.41) that

$$
\begin{aligned}
F(\beta(m)) - F(\beta(m+1)) &\geq G(\beta(m), \beta(m+1)) - F(\beta(m+1)) \\
&\geq \sum_{t=1}^{k} - \frac{(\beta(m+1)^{(t)} - \beta(m)^{(t)})'(\mathbf{X}^{(t)})'\mathbf{X}^{(t)}(\beta(m+1)^{(t)} - \beta(m)^{(t)})}{2\sqrt{n^{(0)}}\,\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\|} \\
&\quad + \frac{1}{2}A(m)\,\|\beta(m+1) - \beta(m)\|^2 \\
&= \sum_{t=1}^{k}(\beta(m+1)^{(t)} - \beta(m)^{(t)})' \left(\frac{A(m)}{2} I - \frac{(\mathbf{X}^{(t)})'\mathbf{X}^{(t)}}{2\sqrt{n^{(0)}}\,\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\|} \right) \\
&\quad \cdot (\beta(m+1)^{(t)} - \beta(m)^{(t)}) \\
&\geq \sum_{t=1}^{k} \frac{1}{2\sqrt{n^{(0)}}} \left(\frac{1}{\|\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\|} - \frac{\left\|\mathbf{X}^{(t)}\right\|_{\ell_2}^2}{2\,\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\|} \right) \\
&\quad \cdot \left\|\beta(m+1)^{(t)} - \beta(m)^{(t)}\right\|^2, \qquad\qquad \text{(A.42)}
\end{aligned}
$$

where I stands for the identity matrix and $\|\mathbf{X}\|_{\ell_2}$ denotes the spectral norm of matrix $\mathbf{X}$.

To show the descent property of our algorithm and thus the convergence of the sequence $\beta(m)$ due to the nonnegativity of the objective function $F(\beta)$ in (A.32), we need to prove that the right-hand side of (A.42) is positive. At the initial step $m = 0$, it is easy to see that this can be achieved by picking a large enough scalar $K_0 > 0$ in the scaling step (A.3) as long as $\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\| \neq 0$. This fact and the regularity condition assumed in Theorem A.1 can guarantee that $F(\beta(m))$ is monotonically decreasing. To see this, set $B_0 = (n^{(0)})^{1/2}F(\beta(0))$ and recall that $\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\| > c_0$ by assumption. It suffices to show that $\|\mathbf{X}^{(t)}\|_{\ell_2}^2 < c_0/B_0$. From the definition of B_0, this claim is equivalent to

$$\|\mathbf{X}^{(t)}\|_{\ell_2}^2 F(\beta(0)) < (n^{(0)})^{-1/2}c_0. \qquad \text{(A.43)}$$

In light of the rescaling step for $Y^{(t)}$, $\mathbf{X}^{(t)}$, and λ in (A.3), we see that the term on the left-hand side of (A.43) scales down with a factor of K_0^{-3}. This entails that as long as $K_0 > 0$ is chosen large enough, inequality (A.43) can be easily satisfied and thus the above claim $\|\mathbf{X}^{(t)}\|_{\ell_2}^2 < c_0/B_0$ holds.

Moreover, we can use the induction later to prove

$$\left\|\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\right\| \leq B_0 \quad \text{and} \quad F(\beta(m)) \leq F(\beta(0)) \qquad \text{(A.44)}$$

for all t and m. Combining the above inequalities (A.44), $\|\mathbf{X}^{(t)}\|_{\ell_2}^2 < c_0/B_0$, and $\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\| > c_0$ results in

$$\frac{1}{\|\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\|} - \frac{\|\mathbf{X}^{(t)}\|_{\ell_2}^2}{2\,\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\|} \geq \frac{1}{2B_0},$$

which along with (A.42) entails that

$$F(\beta(m)) - F(\beta(m+1)) \geq \frac{1}{4}(n^{(0)})^{-1/2}B_0^{-1}\sum_{t=1}^{k}\left\|\beta(m+1)^{(t)} - \beta(m)^{(t)}\right\|^2. \tag{A.45}$$

This shows that $F(\beta(m)) \geq F(\beta(m+1))$. Since $F(\beta(m))$ is always bounded from below by zero, it follows that $\lim_{m\to\infty} F(\beta(m))$ exists and $\lim_{m\to\infty} |F(\beta(m+1)) - F(\beta(m))| = 0$. Thus in view of (A.45), we have

$$\lim_{m\to\infty} \|\beta(m+1) - \beta(m)\| = 0. \tag{A.46}$$

Observe that for each $m \geq 0$,

$$\|\beta(m)\| \leq \sum_{l=1}^{p} \|\beta(m)_{(l)}\| \leq \frac{F(\beta(m))}{\lambda} \leq \frac{F(\beta(0))}{\lambda},$$

which means that all $\beta(m)$ lie in a compact subset of $\mathbb{R}^{kp}$. This fact entails that the sequence $\beta(m)$ has at least one point of convergence. Furthermore, (A.46) ensures that $\beta(m)$ has a unique limit point β^*, which is a fixed point of the soft-thresholding rule given in (A.37).

It now remains to establish the results in (A.44) using induction. When $m = 0$, it is easy to verify that $\|\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\| \leq B_0$ and $F(\beta(m)) \leq F(\beta(0))$. Let us assume that the inequalities $\|\mathbf{X}^{(t)}\beta(m)^{(t)} - Y^{(t)}\| \leq B_0$ and $F(\beta(m)) \leq F(\beta(0))$ in (A.44) hold for all $m \leq T$. Then it follows that

$$\frac{1}{\|\mathbf{X}^{(t)}\beta(T)^{(t)} - Y^{(t)}\|} - \frac{\|\mathbf{X}^{(t)}\|_{\ell_2}^2}{2\,\|\mathbf{X}^{(t)}\xi^{(t)} - Y^{(t)}\|} \geq \frac{1}{2B_0},$$

which together with (A.42) leads to

$$F(\beta(T+1)) \leq F(\beta(T)) \leq F(\beta(0)).$$

We can also obtain $\|\mathbf{X}^{(t)}\beta(T+1)^{(t)} - Y^{(t)}\| \leq (n^{(0)})^{1/2}F(\beta(T+1)) \leq (n^{(0)})^{1/2}F(\beta(0)) = B_0$. Thus (A.44) also holds for $m = T+1$. This completes the proof of (A.44) for all m and t and also concludes the proof of the first step.

Step 2: Global optimality. To conclude the proof, we need to show that the unique point of convergence β^* of our algorithm established in Step 1 is the global optimum of the HGSL optimization problem (A.1). Since $F(\beta)$ defined in (A.32) is the sum of two convex functions of β, it follows that $F(\beta)$ is also a convex function. Thus a vector β is a global minimizer of the objective function $F(\cdot)$ if and only if it satisfies the Karush-Kuhn-Tucker (KKT) conditions

$$\frac{((\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)}))_l}{\sqrt{n^{(0)}}\,\|\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)}\|} = -\lambda\frac{\beta_l^{(t)}}{\|\,\beta_{(l)}\|} \qquad \text{for } \beta_{(l)} \neq \mathbf{0}, \tag{A.47}$$

$$\frac{\left|((\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)}))_l\right|}{\sqrt{n^{(0)}}\,\|\mathbf{X}^{(t)}\beta^{(t)} - Y^{(t)}\|} \leq \lambda \qquad \text{for } \beta_{(l)} = \mathbf{0}, \tag{A.48}$$

where the subscript l in both expressions represents the lth component of a vector.

Recall that we have shown in Step 1 that β^* is the fixed point of the soft-thresholding rule in (A.37), that is,

$$\beta^*_{(l)} = \overrightarrow{\Theta}\left(\beta^*_{(l)} - \frac{R^*_{(l)}}{A^*}; \frac{\lambda}{A^*}\right), \qquad l \in [p],$$

where $R^*_{(l)}$ and A^* are defined similarly as $R(m)_{(l)}$ and $A(m)$ in (A.37) with $\beta(m)$ replaced by β^*. Let us first consider the case when $\beta^*_{(l)} = \mathbf{0}$. Then by the definition of the soft-thresholding rule, we have $\|R^*_{(l)}/A^*\| \leq \lambda/A^*$, which entails that $\|R^*_{(l)}\| \leq \lambda$. Thus it holds that

$$\frac{\left|((\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\beta^{*(t)} - Y^{(t)}))_l\right|}{\sqrt{n^{(0)}}\,\|\mathbf{X}^{(t)}\beta^{*(t)} - Y^{(t)}\|} = |R_l^{*(t)}| \leq \|R^*_{(l)}\| \leq \lambda \tag{A.49}$$

for $\beta^*_{(l)} = \mathbf{0}$, which verifies the second KKT condition (A.48) for the fixed point β^*.

We next consider the case when $\beta^*_{(l)} \neq \mathbf{0}$. It follows from the soft-thresholding rule that

$$\beta^*_{(l)} = \frac{\left\|\beta^*_{(l)} - \frac{R^*_{(l)}}{A^*}\right\| - \frac{\lambda}{A^*}}{\left\|\beta^*_{(l)} - \frac{R^*_{(l)}}{A^*}\right\|}\left(\beta^*_{(l)} - \frac{R^*_{(l)}}{A^*}\right). \tag{A.50}$$

Taking the ℓ_2 norm on both sides of the above equation leads to $\|\beta^*_{(l)}\| = \|\beta^*_{(l)} - R^*_{(l)}/A^*\| - \lambda/A^*$. Moreover, equation (A.50) can be rewritten as

$$-\frac{\lambda}{A^*}\left(\beta^*_{(l)} - \frac{R^*_{(l)}}{A^*}\right) = \frac{R^*_{(l)}}{A^*}\left\|\beta^*_{(l)} - \frac{R^*_{(l)}}{A^*}\right\|,$$

which along with the above fact results in

$$\lambda\beta^*_{(l)} = R^*_{(l)}\left(\left\|\beta^*_{(l)} - \frac{R^*_{(l)}}{A^*}\right\| - \frac{\lambda}{A^*}\right) = R^*_{(l)}\left\|\beta^*_{(l)}\right\|. \tag{A.51}$$

The representation in (A.51) further entails that

$$R_l^{*(t)} = \frac{\left((\mathbf{X}^{(t)})'(\mathbf{X}^{(t)}\beta^{*(t)} - Y^{(t)})\right)_l}{\sqrt{n^{(0)}}\,\|\mathbf{X}^{(t)}\beta^{*(t)} - Y^{(t)}\|} = -\lambda\frac{\beta_l^{(t)}}{\|\beta_{(l)}\|} \tag{A.52}$$

for $\beta^*_{(l)} \neq \mathbf{0}$, which establishes the first KKT condition (A.47) for the fixed point β^*. Combining (A.49) and (A.52), we conclude that $\beta^{(*)}$ is indeed a global minimizer of the HGSL optimization problem (A.1), which completes the proof of Theorem A.1.

C.7 Proof of Proposition 2.3

The support recovery property of our THI estimator $\hat{\mathcal{E}}$ given in (28) follows from the proofs of Theorems 2.1 and 2.3 (1) in Sections C.1 and C.3.1, in view of the conditions of Proposition 2.1 and the assumption that the minimum signal strength $\min_{(a,b)\in\mathcal{E}} \|\omega^0_{a,b}\|$ is above the value of $C\sqrt{[(k\log p)^{1/2} + \log p]/n^{(0)}}$. Specifically, we need a refined technical analysis in the proof of Theorem 2.3 (1) in Section C.3.1 through replacing Chebyshev's inequality used in the third step by an accurate coupling inequality such as Proposition KMT in [**?**], which was also used in Theorem 2 (iii) of [30] for support recovery in the setting of a single Gaussian graphical model. We omit the details here for simplicity.

D Key lemmas and their proofs

D.1 Lemma D.1 and its proof

Lemma D.1. *Assume that Conditions 2.1–2.2 hold and* $\max\{\log p, \log k\} = o(n^{(0)})$. *Let* $\hat{C}^0_j = (\hat{C}^{(1)\prime}_j, \cdots, \hat{C}^{(k)\prime}_j)'$ *be any estimator satisfying working assumptions (15)–(17) for a fixed* $j \in [p]$. *Then there exists some positive constant* C *depending on constants* M, δ, C_1, *and* C_3 *such that*

$$\mathbb{P}\left(\frac{1}{k}\sum_{t=1}^{k}\left|\left(\hat{\omega}^{(t)}_{j,j}\right)^{-1} - \frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(E^{(t)}_{i,j}\right)^2\right| \geq Cs\frac{1+(\log p)/k}{n^{(0)}}\right) \leq 3p^{1-\delta},$$

$$\mathbb{P}\left(\frac{1}{k}\sum_{t=1}^{k}\left|\left(\hat{\omega}^{(t)}_{j,j}\right)^{-1} - \left(\omega^{(t)}_{j,j}\right)^{-1}\right| \geq C\left(\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}} + s\frac{1+(\log p)/k}{n^{(0)}}\right)\right) \leq 3p^{1-\delta} + \delta_1$$

as long as $\log(\delta_1^{-1}) = o(n^{(0)})$. *Moreover, whenever* $\max\{\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}, s\frac{(k+\log p)}{n^{(0)}}\} = o(1)$*, there exists some positive constant* C' *depending on* M,δ*,* C_1*, and* C_3 *such that*

$$\mathbb{P}\left(\frac{1}{k}\sum_{t=1}^{k}\left|\hat{\omega}_{j,j}^{(t)} - \left(\frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(E_{i,j}^{(t)}\right)^2\right)^{-1}\right| \geq C's\frac{(1+(\log p)/k)}{n^{(0)}}\right) \leq 3p^{1-\delta},$$

$$\mathbb{P}\left(\frac{1}{k}\sum_{t=1}^{k}\left|\hat{\omega}_{j,j}^{(t)} - \omega_{j,j}^{(t)}\right| \geq C'\left(\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}} + s\frac{(1+(\log p)/k)}{n^{(0)}}\right)\right) \leq 3p^{1-\delta} + \delta_1,$$

$$\mathbb{P}\left(\max_{t\in[k]}\left|\hat{\omega}_{j,j}^{(t)} - \omega_{j,j}^{(t)}\right| \geq C'\left(\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}} + s\frac{(k+\log p)}{n^{(0)}}\right)\right) \leq 3p^{1-\delta} + \delta_1.$$

Proof. Observe that $\frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}(\hat{E}_{i,j}^{(t)})^2 = (\hat{\omega}_{j,j}^{(t)})^{-1}$. For each $j \in [p]$, in view of $\hat{E}_{i,j}^{(t)} = E_{i,j}^{(t)} + X_{i,-j}^{(t)\prime}(C_j^{(t)} - \hat{C}_j^{(t)})$ we deduce that

$$\frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(\hat{E}_{i,j}^{(t)}\right)^2 = \frac{1}{n^{(t)}}\Bigg\{\sum_{i=1}^{n^{(t)}}\left(E_{i,j}^{(t)}\right)^2 + 2E_{*,j}^{(t)\prime}\mathbf{X}_{*,-j}^{(t)}(C_j^{(t)} - \hat{C}_j^{(t)}) + (C_j^{(t)} - \hat{C}_j^{(t)})'\mathbf{X}_{*,-j}^{(t)\prime}\mathbf{X}_{*,-j}^{(t)}(C_j^{(t)} - \hat{C}_j^{(t)})\Bigg\}. \tag{A.53}$$

Thus we have

$$\begin{aligned} &\frac{1}{k}\sum_{t=1}^{k}\left|\left(\hat{\omega}_{j,j}^{(t)}\right)^{-1} - \sum_{i=1}^{n^{(t)}}\left(E_{i,j}^{(t)}\right)^2/n^{(t)}\right| \\ \leq\ & \frac{1}{k}\sum_{t=1}^{k}\frac{1}{n^{(t)}}\left(2\left|E_{*,j}^{(t)\prime}\mathbf{X}_{*,-j}^{(t)}(C_j^{(t)} - \hat{C}_j^{(t)})\right| + \left\|\mathbf{X}_{*,-j}^{(t)}(C_j^{(t)} - \hat{C}_j^{(t)})\right\|^2\right) \\ \equiv\ & T_1 + T_2. \end{aligned} \tag{A.54}$$

We will consider the above two terms T_1 and T_2 separately.

For the second term T_2, we can bound it by our working assumption (17) as

$$T_2 = \frac{1}{k}\sum_{t=1}^{k}\frac{1}{n^{(t)}}\left\|\mathbf{X}_{*,-j}^{(t)}(C_j^{(t)} - \hat{C}_j^{(t)})\right\|^2 \leq C_3 s\frac{1+(\log p)/k}{n^{(0)}}. \tag{A.55}$$

The first term T_1 can be bounded with probability at least $1-3p^{1-\delta}$ as

$$
\begin{aligned}
T_1 &\le \frac{2}{k}\sum_{l\neq j}\sum_{t=1}^{k}\left|\frac{E_{*,j}^{(t)\prime}X_{*,l}^{(t)}}{n^{(t)}}\right|\cdot\left|C_{j,l}^{(t)}-\hat{C}_{j,l}^{(t)}\right| \\
&\le \sum_{l\neq j}\left(\frac{1}{k}\sum_{t=1}^{k}(\frac{E_{*,j}^{(t)\prime}X_{*,l}^{(t)}}{n^{(t)}})^2\right)^{1/2}\left(\frac{1}{k}\sum_{t=1}^{k}(C_{j,l}^{(t)}-\hat{C}_{j,l}^{(t)})^2\right)^{1/2} \\
&\le \max_{l\neq j}\left(\frac{1}{k}\sum_{t=1}^{k}(\frac{E_{*,j}^{(t)\prime}X_{*,l}^{(t)}}{n^{(t)}})^2\right)^{1/2}\sum_{l\neq j}\frac{1}{\sqrt{k}}\|\Delta_{j(l)}\| \\
&\le c_\delta\left(\frac{1+(\log p)/k}{n^{(0)}}\right)^{1/2}s\left(\frac{1+(\log p)/k}{n^{(0)}}\right)^{1/2}, \qquad \text{(A.56)}
\end{aligned}
$$

where the last inequality is due to working assumption (16) and Lemma E.2 in Section E with c_δ some positive constant depending only on δ, M, and C_1. Thus we have shown the first desired result.

Let us further bound the difference between the oracle estimator $\sum_{i=1}^{n^{(t)}}(E_{i,j}^{(t)})^2/n^{(t)}$ and its mean $(\omega_{j,j}^{(t)})^{-1}$. Indeed, it holds that $\sum_{i=1}^{n^{(t)}}(E_{i,j}^{(t)})^2(\omega_{j,j}^{(t)})\sim\chi^2(n^{(t)})$. This representation entails that as long as $\log(\delta_1^{-1})=o(n^{(0)})$, by Lemma E.1 and $n^{(0)}\le n^{(t)}$ we have

$$
\left|\frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(E_{i,j}^{(t)}\right)^2-1/\omega_{j,j}^{(t)}\right|=\frac{1}{n^{(t)}}\left|\sum_{i=1}^{n^{(t)}}\left(\left(E_{i,j}^{(t)}\right)^2-\mathbb{E}\left(E_{i,j}^{(t)}\right)^2\right)\right|\le c_M\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}} \qquad \text{(A.57)}
$$

with probability at least $1-\delta_1/k$, where c_M is some positive constant depending only on M. Combining inequalities (A.54)–(A.57) with the union bound argument, we obtain the second desired result that with probability at least $1-3p^{1-\delta}-\delta_1$,

$$
\frac{1}{k}\sum_{t=1}^{k}\left|\left(\hat{\omega}_{j,j}^{(t)}\right)^{-1}-\left(\omega_{j,j}^{(t)}\right)^{-1}\right|\le C\left(\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}+s\frac{1+(\log p)/k}{n^{(0)}}\right),
$$

where C is some positive constant that depends on M, δ, C_1, and C_3.

Note that whenever $\max\{\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}},s\frac{(k+\log p)}{n^{(0)}}\}=o(1)$, it follows from inequalities (A.54)–(A.57) and the union bound argument that with probability at least $1-3p^{1-\delta}-\delta_1$,

$$
\max_t\left|1/\hat{\omega}_{j,j}^{(t)}-1/\omega_{j,j}^{(t)}\right|\le C\left(\sqrt{\frac{\log(k/\delta_1)}{n^{(0)}}}+s\frac{k+\log p}{n^{(0)}}\right), \qquad \text{(A.58)}
$$

which is sufficiently small for large $n^{(0)}$. Consequently, we see that $\hat{\omega}_{j,j}^{(t)}$ is uniformly bounded from above by some positive constant for all $t\in[k]$, since $\omega_{j,j}^{(t)}$ is bounded from above by M

by Condition 2.1. Therefore, in light of $|\hat{\omega}_{j,j}^{(t)} - \omega_{j,j}^{(t)}| = |1/\hat{\omega}_{j,j}^{(t)} - 1/\omega_{j,j}^{(t)}|\omega_{j,j}^{(t)}\hat{\omega}_{j,j}^{(t)}$ the last three desired inequalities follow from the first two established above and inequality (A.58), which concludes the proof.

D.2 Lemma D.2 and its proof

Lemma D.2. *Assume that Conditions 2.1–2.2 hold, working assumptions (15)–(17) are valid for $j = 1, 2$, and $\max\{\log p, \log k\} = o(n^{(0)})$. Then there exists some positive constant C depending only on constants M, δ, C_1, C_2, and C_3 such that*

$$\frac{1}{k}\sum_{t=1}^{k}\left|T_{n,k,1,2}^{(t)} - J_{n,k,1,2}^{(t)} - \frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(E_{i,1}^{(t)}E_{i,2}^{(t)} - \mathbb{E}E_{i,1}^{(t)}E_{i,2}^{(t)}\right)\right|$$
$$\leq C''\left(s\frac{1+(\log p)/k}{n^{(0)}}(1+\sqrt{ks\frac{1+(\log p)/k}{n^{(0)}}})\right) \tag{A.59}$$

holds with probability at least $1 - 6p^{1-\delta}$.

Proof. At a high level, the first term $\frac{1}{k}\sum_{t=1}^{k}|\sum_{i=1}^{n^{(t)}}\hat{E}_{i,1}^{(t)}\hat{E}_{i,2}^{(t)}/n^{(t)}|$ in $T_{n,k,1,2}$ is constructed to approximate $\frac{1}{k}\sum_{t=1}^{k}|\sum_{i=1}^{n^{(t)}}E_{i,1}^{(t)}E_{i,2}^{(t)}/n^{(t)}|$, but some bias appears in the approximation. The remaining two terms $\sum_{i=1}^{n^{(t)}}(\hat{E}_{i,1}^{(t)})^2\hat{C}_{2,1}/n^{(t)}$ and $\sum_{i=1}^{n^{(t)}}(\hat{E}_{i,2}^{(t)})^2\hat{C}_{1,2}/n^{(t)}$ in each $T_{n,k,1,2}^{(t)}$ serve as the remedy to correct the bias when the null $\omega_{1,2}^0 = \mathbf{0}$ is true. In view of $\hat{E}_{i,j}^{(t)} = E_{i,j}^{(t)} + X_{i,-j}^{(t)\prime}(C_j^{(t)} - \hat{C}_j^{(t)})$, we can deduce

$$\begin{aligned}
&\frac{1}{k}\sum_{t=1}^{k}\left|\frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\hat{E}_{i,1}^{(t)}\hat{E}_{i,2}^{(t)}\right| \\
=\ &\frac{1}{k}\sum_{t=1}^{k}\left|\frac{1}{n^{(t)}}E_{*,1}^{(t)\prime}E_{*,2}^{(t)} + \frac{1}{n^{(t)}}E_{*,1}^{(t)\prime}\mathbf{X}_{*,-2}^{(t)}(C_2^{(t)} - \hat{C}_2^{(t)})\right. \\
&+\frac{1}{n^{(t)}}E_{*,2}^{(t)\prime}\mathbf{X}_{*,-1}^{(t)}(C_1^{(t)} - \hat{C}_1^{(t)}) \\
&\left.+\frac{1}{n^{(t)}}(C_1^{(t)} - \hat{C}_1^{(t)})^T\mathbf{X}_{*,-1}^{(t)\prime}\mathbf{X}_{*,-2}^{(t)}(C_2^{(t)} - \hat{C}_2^{(t)})\right| \\
=\ &\frac{1}{k}\sum_{t=1}^{k}\left|H_1^{(t)} + H_2^{(t)} + H_3^{(t)} + H_4^{(t)}\right|.
\end{aligned} \tag{A.60}$$

The main term $H_1^{(t)}$ above enjoys the following property

$$\begin{aligned} H_1^{(t)} &= \frac{1}{n^{(t)}} \sum_{i=1}^{n^{(t)}} E_{i,1}^{(t)} E_{i,2}^{(t)} = \mathbb{E} E_{1,1}^{(t)} E_{1,2}^{(t)} + \frac{1}{n^{(t)}} \sum_{i=1}^{n^{(t)}} \left(E_{i,1}^{(t)} E_{i,2}^{(t)} - \mathbb{E} E_{i,1}^{(t)} E_{i,2}^{(t)} \right) \\ &= \frac{\omega_{1,2}^{(t)}}{\omega_{1,1}^{(t)} \omega_{2,2}^{(t)}} + \frac{1}{n^{(t)}} \sum_{i=1}^{n^{(t)}} \left(E_{i,1}^{(t)} E_{i,2}^{(t)} - \mathbb{E} E_{i,1}^{(t)} E_{i,2}^{(t)} \right). \end{aligned} \tag{A.61}$$

We can bound the last term $\sum_{t=1}^{k} |H_4^{(t)}|/k$ in (A.60) as

$$\begin{aligned} \frac{1}{k} \sum_{t=1}^{k} \left| H_4^{(t)} \right| &\leq \frac{1}{k} \sum_{t=1}^{k} \frac{1}{n^{(t)}} \left\| \mathbf{X}_{*,-2}^{(t)} (C_2^{(t)} - \hat{C}_2^{(t)}) \right\| \left\| \mathbf{X}_{*,-1}^{(t)} (C_1^{(t)} - \hat{C}_1^{(t)}) \right\| \\ &\leq \frac{1}{2k} \sum_{t=1}^{k} \frac{1}{n^{(t)}} \left(\left\| \mathbf{X}_{*,-2}^{(t)} (C_2^{(t)} - \hat{C}_2^{(t)}) \right\|^2 + \left\| \mathbf{X}_{*,-1}^{(t)} (C_1^{(t)} - \hat{C}_1^{(t)}) \right\|^2 \right) \\ &\leq C_3 s \frac{1 + (\log p)/k}{n^{(0)}}, \end{aligned}$$

where the last inequality follows from our working assumption (17).

The second term $H_2^{(t)}$ in (A.60) can be further decomposed as

$$\begin{aligned} H_2^{(t)} &= \frac{1}{n^{(t)}} \left(E_{*,1}^{(t)\prime} X_{*,1}^{(t)} (C_{2,1}^{(t)} - \hat{C}_{2,1}^{(t)}) + E_{*,1}^{(t)\prime} \mathbf{X}_{*,\{1,2\}^c}^{(t)} (C_{2,-1}^{(t)} - \hat{C}_{2,-1}^{(t)}) \right) \\ &\equiv H_{2,0}^{(t)} + H_{2,1}^{(t)}. \end{aligned} \tag{A.62}$$

We can bound $\sum_{t=1}^{k} |H_{2,1}^{(t)}|/k$ such that with probability at least $1 - 3p^{1-\delta}$,

$$\begin{aligned} \frac{1}{k} \sum_{t=1}^{k} \left| H_{2,1}^{(t)} \right| &\leq \frac{1}{k} \sum_{j=3}^{p} \sum_{t=1}^{k} \left| \frac{E_{*,1}^{(t)\prime} X_{*,j}^{(t)}}{n^{(t)}} \right| \cdot \left| C_{2,j}^{(t)} - \hat{C}_{2,j}^{(t)} \right| \\ &\leq \sum_{j=3}^{p} \left(\frac{1}{k} \sum_{t=1}^{k} \left(\frac{E_{*,1}^{(t)\prime} X_{*,j}^{(t)}}{n^{(t)}} \right)^2 \right)^{1/2} \left(\frac{1}{k} \sum_{t=1}^{k} (C_{2,j}^{(t)} - \hat{C}_{2,j}^{(t)})^2 \right)^{1/2} \\ &\leq \max_j \left(\frac{1}{k} \sum_{t=1}^{k} \left(\frac{E_{*,1}^{(t)\prime} X_{*,j}^{(t)}}{n^{(t)}} \right)^2 \right)^{1/2} \sum_{j=3}^{p} \frac{1}{\sqrt{k}} \left\| \Delta_{2(j)} \right\| \\ &\leq C \left(\frac{1 + (\log p)/k}{n^{(0)}} \right)^{1/2} s \left(\frac{1 + (\log p)/k}{n^{(0)}} \right)^{1/2}, \end{aligned}$$

where the last inequality is due to working assumption (16) and Lemma E.2. Observe that similar decomposition, notation, and analysis apply to term $H_3^{(t)}$ as well. Hence, it holds that with probability at least $1 - 3p^{1-\delta}$,

$$\frac{1}{k} \sum_{t=1}^{k} \left| \frac{1}{n^{(t)}} \sum_{i=1}^{n^{(t)}} \hat{E}_{i,1}^{(t)} \hat{E}_{i,2}^{(t)} - \left(H_1^{(t)} + H_{2,0}^{(t)} + H_{3,0}^{(t)} \right) \right| \leq C \left(\frac{s}{n^{(0)}} (1 + (\log p)/k) \right). \tag{A.63}$$

Let us decompose term $H_{2,0}^{(t)}$ in (A.62) as

$$
\begin{aligned}
H_{2,0}^{(t)} &= \frac{1}{n^{(t)}}\Big\{\hat{E}_{*,1}^{(t)\prime}\hat{E}_{*,1}^{(t)}(C_{2,1}^{(t)}-\hat{C}_{2,1}^{(t)})+\sum_{i=1}^{n^{(t)}}E_{i,1}^{(t)}X_{i,-1}^{(t)\prime}C_1^{(t)}(C_{2,1}^{(t)}-\hat{C}_{2,1}^{(t)}) \\
&\quad +(E_{*,1}^{(t)\prime}E_{*,1}^{(t)}-\hat{E}_{*,1}^{(t)\prime}\hat{E}_{*,1}^{(t)})(C_{2,1}^{(t)}-\hat{C}_{2,1}^{(t)})\Big\} \\
&\equiv H_{2,0,0}^{(t)}+H_{2,0,1}^{(t)}+H_{2,0,2}^{(t)}. \qquad (A.64)
\end{aligned}
$$

Now we control the two terms $\sum_{t=1}^k|H_{2,0,1}^{(t)}|/k$ and $\sum_{t=1}^k|H_{2,0,2}^{(t)}|/k$ separately, and leave $H_{2,0,0}^{(t)}$ as the main term. By Lemma E.2 and working assumption (16), we obtain that with probability at least $1-3p^{-\delta}$,

$$
\begin{aligned}
\frac{1}{k}\sum_{t=1}^k\left|H_{2,0,1}^{(t)}\right| &\le \left(\frac{1}{k}\sum_{t=1}^k\Big(\frac{E_{*,1}^{(t)\prime}\mathbf{X}_{*,-1}^{(t)}C_1^{(t)}}{n^{(t)}}\Big)^2\right)^{1/2}\left(\frac{1}{k}\sum_{t=1}^k(C_{2,1}^{(t)}-\hat{C}_{2,1}^{(t)})^2\right)^{1/2} \\
&\le C\left(\frac{1+(\log p)/k}{n^{(0)}}\right)^{1/2}\left(s\frac{1+(\log p)/k}{n^{(0)}}\right)^{1/2}. \qquad (A.65)
\end{aligned}
$$

As for the term $H_{2,0,2}^{(t)}$ in (A.64), we can show that with probability at least $1-3p^{1-\delta}$,

$$
\begin{aligned}
\frac{1}{k}\sum_{t=1}^k\left|H_{2,0,2}^{(t)}\right| &= \frac{1}{k}\sum_{t=1}^k\left|(E_{*,1}^{(t)\prime}E_{*,1}^{(t)}-\hat{E}_{*,1}^{(t)\prime}\hat{E}_{*,1}^{(t)})(C_{2,1}^{(t)}-\hat{C}_{2,1}^{(t)})\right| \\
&\le \frac{1}{k}\sum_{t=1}^k\left|\frac{1}{n^{(t)}}\Big(\sum_{i=1}^{n^{(t)}}\left(\hat{E}_{i,1}^{(t)}\right)^2-\sum_{i=1}^{n}\left(E_{i,1}^{(t)}\right)^2\Big)\right|\max_t\left|C_{2,1}^{(t)}-\hat{C}_{2,1}^{(t)}\right| \\
&\le Cs\frac{1+(\log p)/k}{n^{(0)}}\cdot\max_t\left\|\Delta_{1(t)}\right\| \\
&\le Cs\frac{1+(\log p)/k}{n^{(0)}}\cdot\left(ks\frac{1+(\log p)/k}{n^{(0)}}\right)^{1/2}, \qquad (A.66)
\end{aligned}
$$

where the second inequality follows from expressions (A.53)–(A.56) in the earlier proof of Lemma D.1 in Section D.1 and the last inequality follows from our working assumption (15). Note that similar decomposition, notation, and analysis also apply to term $H_{3,0}^{(t)}$. Thus combining the above expressions (A.63)–(A.66) yields that with probability at least $1-3p^{-\delta}-3p^{1-\delta}$,

$$
\begin{aligned}
&\frac{1}{k}\sum_{t=1}^k\left|\frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\hat{E}_{i,1}^{(t)}\hat{E}_{i,2}^{(t)}-\left(H_1^{(t)}+H_{2,0,0}^{(t)}+H_{3,0,0}^{(t)}\right)\right| \\
\le\ & C\left(\frac{s}{n^{(0)}}(1+(\log p)/k)\right)\left(1+(ks\frac{1+(\log p)/k}{n^{(0)}})^{1/2}\right). \qquad (A.67)
\end{aligned}
$$

We finally correct the bias in $H_{2,0,0}^{(t)}$ and $H_{3,0,0}^{(t)}$ induced from $\hat{C}_{2,1}$. To this end, we take the sum of $\hat{E}_{*,1}^{(t)\prime}\hat{E}_{*,2}^{(t)}/n^{(t)}$ and two terms $\hat{E}_{*,1}^{(t)\prime}\hat{E}_{*,1}^{(t)}\hat{C}_{2,1}/n^{(t)}$, $\hat{E}_{*,1}^{(t)\prime}\hat{E}_{*,1}^{(t)}\hat{C}_{1,2}/n^{(t)}$ out of $H_{2,0,0}^{(t)}$ and $H_{3,0,0}^{(t)}$ as the statistic $T_{n,k,1,2}^{(t)}$. The remaining terms in $H_{2,0,0}^{(t)}$ and $H_{3,0,0}^{(t)}$ together with the first term of decomposition of $H_1^{(t)}$ in (A.61) form $J_{n,k,1,2}^{(t)}$ defined in (10), in light of $C_{2,1}^{(t)} = -\omega_{1,2}^{(t)}/\omega_{2,2}^{(t)}$ and $C_{1,2}^{(t)} = -\omega_{1,2}^{(t)}/\omega_{1,1}^{(t)}$. Therefore, the desired result follows from (A.67), that is, with probability at least $1 - 3p^{-\delta} - 3p^{1-\delta}$,

$$\begin{aligned}
&\frac{1}{k}\sum_{t=1}^{k}\left|T_{n,k,1,2}^{(t)} - J_{n,k,1,2}^{(t)} - \frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\left(E_{i,1}^{(t)}E_{i,2}^{(t)} - \mathbb{E}E_{i,1}^{(t)}E_{i,2}^{(t)}\right)\right| \\
\leq\ & C''\left(\frac{s}{n^{(0)}}(1+(\log p)/k)\right)\left(1+(ks\frac{1+(\log p)/k}{n^{(0)}})^{1/2}\right)
\end{aligned}$$

with C'' some positive constant. Keeping track of all relevant constants, we see that the positive constant C'' depends only on M, δ, C_1, C_2, and C_3, which completes the proof.

D.3 Lemma D.3 and its proof

Lemma D.3. *With $\mathcal{G}$ and Ω_0^0 chosen as in (A.9) and (A.8), we have $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - \frac{1}{2}(\beta - \alpha)$ with some sufficiently small constant $\tau > 0$ depending only on $\beta - \alpha$.*

Proof. A similar argument to that used in the later proof of Lemma D.4 in Section D.4 (see inequality (A.68)) entails that it is sufficient to show that the χ^2 divergence between $\mathbb{P}_0$ and $\bar{\mathbb{P}}$ is small enough, that is,

$$\Delta = \int\left(\frac{1}{m}\sum_{h=1}^{m} f_h\right)^2/f_0 - 1 = \sum_{h_1,h_2=1}^{m}(\int(\frac{f_{h_1}f_{h_2}}{f_0}) - 1)/(m)^2 < (\beta-\alpha)^2.$$

Recall that $g_h^{(t)}$ denotes the density of $N(0, (\Omega_h^{(t)})^{-1})$ for $h = 0, \cdots, m$. By our construction of Ω_0^0 and Ω_1^0, together with the χ^2 divergence of two Gaussian distributions in (A.69), we can deduce that for any $h_1, h_2 \in [m]$,

$$\begin{aligned}
\int\frac{f_{h_1}f_{h_2}}{f_0} &= \left(\int\prod_{t=1}^{h} g_{h_1}^{(t)}g_{h_2}^{(t)}/g_0^{(t)}\right)^{n^{(0)}} = \left(1 - 1/n^{(0)}\right)^{-\dot{J}(h_1,h_2)n^{(0)}} \\
&\leq \left(1 + 2/n^{(0)}\right)^{\dot{J}(h_1,h_2)n^{(0)}} \leq \exp(2\dot{J}(h_1,h_2)),
\end{aligned}$$

where we have used $1/n^{(0)} < 1/2$ in the second to last inequality and $\dot{J} = \dot{J}(h_1, h_2)$ is the cardinality of $T_{h1} \cap T_{h_2}$ with the index sets $T_{h_i} \subset [k]$ denoting those graphs with non-identity precision matrices in (A.9) for $i = 1, 2$. In other words, $\dot{J}(h_1, h_2)$ is the number of overlapping non-identity precision matrices between two sets of k precision matrices indexed by $\Omega^0_{h_1}$ and $\Omega^0_{h_2}$. It is easy to see that integer $\dot{J} = \dot{J}(h_1, h_2) \in [0, \cdots, \tau\sqrt{k}]$.

Recall that $m = \binom{k}{\tau\sqrt{k}}$. Thus we have

$$\begin{aligned}\Delta &= \frac{1}{(m)^2} \sum_{0 \le j \le \tau\sqrt{k}} \sum_{\dot{J}(h_1,h_2)=j} \Big(\exp(2\dot{J}(h_1, h_2)) - 1\Big) \\ &\le \frac{1}{(m)^2} \sum_{1 \le j \le \tau\sqrt{k}} \binom{k}{\tau\sqrt{k}} \binom{\tau\sqrt{k}}{j} \binom{k-j}{\tau\sqrt{k}-j} \exp(2j) \\ &= \sum_{1 \le j \le \tau\sqrt{k}} \binom{\tau\sqrt{k}}{j} \binom{k-j}{\tau\sqrt{k}-j} \Big/ \binom{k}{\tau\sqrt{k}} \cdot \exp(2j) \\ &\le \sum_{1 \le j \le \tau\sqrt{k}} \frac{1}{j!} \left(\frac{\tau^2 k \exp(2)}{k - \tau\sqrt{k}}\right)^j \\ &\le \exp(\lambda)\mathbb{P}(Z > 0) = \exp(\lambda) - 1,\end{aligned}$$

where in the last inequality we bounded the sum using a Poisson random variable Z with parameter $\lambda = \tau^2 k \exp(2)/(k - \tau\sqrt{k})$. Finally, we can conclude the proof by picking a small enough constant τ depending on $\beta - \alpha$ to obtain $\Delta \le (\beta - \alpha)^2$.

D.4 Lemma D.4 and its proof

Lemma D.4. *With $\mathcal{G}$ and Ω^0_0 specified in (A.23) and (A.22), it holds that $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - \frac{1}{2}(\beta - \alpha)$ with some sufficiently small constant $\tau > 0$ depending only on M_1 and μ.*

Proof. Recall that the densities of distributions $\mathbb{P}_h$ and $N(0, (\Omega^{(1)}_h)^{-1})$ are denoted as f_h and g_h, respectively, for each $0 \le h \le m$. By Jensen's inequality we have

$$\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| = \int (f_0 \wedge \bar{f}) \ge 1 - \frac{1}{2}\Big(\int \frac{\bar{f}^2}{f_0} - 1\Big)^{1/2} = 1 - \sqrt{\Delta}/2.$$

Thus it suffices to show that the χ^2 divergence is small enough

$$\Delta = \int \frac{\left(\frac{1}{m}\sum_{h=1}^m f_h\right)^2}{f_0} - 1 = \frac{1}{m^2} \sum_{h_1,h_2=1}^{m} \left(\int \Big(\frac{f_{h_1} f_{h_2}}{f_0}\Big) - 1\right) < (\beta - \alpha)^2, \tag{A.68}$$

which yields the desired bound $\|\mathbb{P}_0 \wedge \bar{\mathbb{P}}\| > 1 - \frac{1}{2}(\beta - \alpha)$.

The following representation of the χ^2 divergence of two Gaussian distributions

$$\int \frac{g_1 g_2}{g_0} = [\det(I - \Sigma_0^{-1}(\Sigma_1 - \Sigma_0)\Sigma_0^{-1}(\Sigma_2 - \Sigma_0))]^{-1/2}, \tag{A.69}$$

with g_i the density of $N(0, \Sigma_i)$ for $i = 0, 1, 2$, is helpful to our analysis. By our construction of $\mathbb{P}_h$ and (A.69), some algebra results in

$$\int \frac{f_{h_1} f_{h_2}}{f_0} = \left(\int \prod_{t=1}^{h} g_{h_1}^{(t)} g_{h_2}^{(t)} / g_0^{(t)}\right)^{n^{(0)}} = \left(1 - 2Ja^2\right)^{-n^{(0)}k},$$

where $J = J(h_1, h_2)$ is the number of overlapping a between the first rows of $(\Omega_{h_1}^{(1)})^{-1}$ and $(\Omega_{h_2}^{(1)})^{-1}$. Hence it follows that

$$\begin{aligned} \Delta &= \frac{1}{m^2} \sum_{0 \le j \le s-1} \sum_{J(h_1,h_2)=j} \left(\left(1 - 2ja^2\right)^{-n^{(0)}k} - 1\right) \\ &= \frac{1}{m^2} \sum_{1 \le j \le s-1} \binom{p-1}{s-1} \binom{s-1}{j} \binom{p-s}{s-1-j} \left(\left(1 - 2ja^2\right)^{-n^{(0)}k} - 1\right). \end{aligned}$$

Observe that since $2ja^2 \le 2(s-1)a^2 < 1/2$ and $k \le M_1 \log p$, we have

$$\begin{aligned} \left(1 - 2ja^2\right)^{-n^{(0)}k} &\le \left(1 + 4ja^2\right)^{n^{(0)}k} \le \exp(4ja^2 n^{(0)} k) = \exp(4j\tau(k + \log p)) \\ &\le (p)^{4(1+M_1)\tau j}. \end{aligned}$$

Moreover, it can be checked that with $m = \binom{p-1}{s-1}$,

$$\frac{1}{m^2} \binom{p-1}{s-1} \binom{s-1}{j} \binom{p-s}{s-1-j} \le \left(\frac{s^2}{p-s}\right)^j.$$

Therefore, combining the three expressions above we can complete the proof by noting that

$$\Delta \le \sum_{1 \le j \le s-1} \left(\frac{s^2 p^{4(1+M_1)\tau}}{p-s}\right)^j \to 0,$$

where we have used $p > s^\mu$ for some $\mu > 2$ and picked a small enough constant τ depending on μ and M_1.

D.5 Lemma D.5 and its proof

Lemma D.5. *For any fixed ξ, under Conditions 2.1–2.2 and the assumption of $s < C_\xi n^{(0)} / \log p$ with some sufficiently small constant $C_\xi > 0$ depending on ξ, M, and M_0, we have $\mathbb{P}\{\mathcal{E}_{1,gRE}\} >$*

$1 - 2k\exp(-cn^{(0)})$, *where* $\mathcal{E}_{1,gRE} = \{gRE(\xi, T) > \min_{l,t}\{(n^{(t)}/\mathbf{X}^{(t)\prime}_{*,l}\mathbf{X}^{(t)}_{*,l})^{1/2}\}/(2M)^{1/2}\}$ *and* $c > 0$ *is some constant depending on* ξ, M, *and* M_0.

Proof. The proof of the group-wise restricted eigenvalue (gRE) condition follows from a similar reduction principle to that developed in [**?**] and [**?**] for dealing with the regular restricted eigenvalue (RE) condition. First of all, due to the normalization constant, that is, $\bar{\mathbf{X}}^0_{*,-1} = \mathbf{X}^0_{*,-1}(\bar{D}_1)^{-1/2}$, it suffices to show that with probability at least $1 - 2k\exp(-cn^{(0)})$,

$$\inf_{u \neq 0} \left\{ \frac{\left\|\mathbf{X}^0_{*,-1}u\right\|}{\sqrt{n^{(0)}}\left\|u\right\|} : u \in \Psi(\xi, T) \right\} \geq (2M)^{-1/2}. \tag{A.70}$$

To further reduce the condition in (A.70), we note that

$$\frac{u'\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1}u}{n^{(0)}\left\|u\right\|^2} = \frac{u'\mathbb{E}\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1}\right)u}{n^{(0)}\left\|u\right\|^2} + \frac{u'\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1} - \mathbb{E}\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1}\right)\right)u}{n^{(0)}\left\|u\right\|^2}$$

and the first term above is lower bounded by M^{-1}, that is,

$$\frac{u'\mathbb{E}\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1}\right)u}{n^{(0)}\left\|u\right\|^2} = \sum_{t=1}^{k} \frac{u^{(t)\prime}\Sigma^{(t)}_{-1,-1}u^{(t)}}{\left\|u^{(t)}\right\|^2} \cdot \frac{n^{(t)}}{n^{(0)}} \geq \frac{1}{M},$$

where the last inequality follows from Conditions 2.1–2.2. Thus it remains to prove that with probability at least $1 - 2k\exp(-cn^{(0)})$,

$$\left|\frac{u'\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1} - \mathbb{E}\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1}\right)\right)u}{n^{(0)}\left\|u\right\|^2}\right| \leq \frac{1}{2M} \quad \text{for all } u \in \Psi(\xi, T). \tag{A.71}$$

Before proceeding, let us introduce some notation. Let

$$\mathbb{K}(m) = \{u \in \mathbb{R}^{k(p-1)} : \sum_{l=2}^{p} 1\{u_{(l)} \neq 0\} \leq m\}$$

be the group-wise m-sparse set. The proof of (A.71) is comprised of two steps. In the first step we prove that the following inequality holds with probability at least $1 - 2k\exp(-cn^{(0)})$ for all $u \in \mathbb{K}(2s)$,

$$\begin{aligned} &\left|\frac{u'\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1} - \mathbb{E}\left(\mathbf{X}^{0\prime}_{*,-1}\mathbf{X}^0_{*,-1}\right)\right)u}{n^{(0)}\left\|u\right\|^2}\right| \\ =\; &\left|\sum_{t=1}^{k} \frac{u^{(t)\prime}(\mathbf{X}^{(t)\prime}_{*,-1}\mathbf{X}^{(t)}_{*,-1}/n^{(t)} - \Sigma^{(t)}_{-1,-1})u^{(t)}}{\left\|u^{(t)}\right\|^2} \cdot \frac{n^{(t)}}{n^{(0)}}\right| \\ \leq\; &\frac{1}{6(2+\xi)^2 M}, \end{aligned} \tag{A.72}$$

while the second step shows that (A.72) entails (A.71) deterministically.

The inequality (A.72) can be established by the standard δ-net argument for each of the design matrices $\mathbf{X}_{*,-1}^{(t)}$ and a union bound argument. Denote by

$$\mathbb{K}^{(t)}(m) = \left\{ u^{(t)} \in \mathbb{R}^{(p-1)} : \sum_{l=2}^{p} 1\{u_l^{(t)} \neq 0\} \leq m \right\}.$$

Then an application of Lemma 15 in [?] implies that there exists some absolute constant $c_0 > 0$ such that

$$\begin{aligned} &\mathbb{P}\left(\sup_{u^{(t)} \in \mathbb{K}^{(t)}(2s)} \left| \frac{u^{(t)\prime} \left(\mathbf{X}_{*,-1}^{(t)\prime} \mathbf{X}_{*,-1}^{(t)} / n^{(t)} - \Sigma_{-1,-1}^{(t)} \right) u^{(t)}}{\|u^{(t)}\|^2} \right| > x \right) \\ \leq \;& 2\exp(-c_0 n^{(t)} \min\{x^2/M^2, x/M\} + 4s \log p). \end{aligned}$$

Note that $n^{(t)}/n^{(0)} \leq M_0$ from Condition 2.2. Therefore, the union bound of the above inequality for all $t \in [k]$, together with the choice $x = (6(2+\xi)^2 M M_0)^{-1}$ and our assumption $s < C_\xi n^{(0)}/\log p$ with some sufficiently small constant $C_\xi > 0$ depending on ξ, M, and M_0, yields that (A.72) holds with probability at least $1 - 2k\exp(-cn^{(0)})$ for some positive constant c depending on ξ, M, and M_0.

It remains to show that (A.72) in fact implies the desired result in (A.71). From now on, denote by

$$\mathbf{\Gamma} = (\mathbf{X}_{*,-1}^{0\prime} \mathbf{X}_{*,-1}^{0} - \mathbb{E}(\mathbf{X}_{*,-1}^{0\prime} \mathbf{X}_{*,-1}^{0}))/n^{(0)}.$$

In order to show (A.71), by the scaling property it suffices to establish

$$|u'\mathbf{\Gamma} u| \leq \frac{1}{2M} \quad \text{for all } u \in \Psi(\xi, T) \cap B_2(1), \tag{A.73}$$

where $B_2(1)$ is the unit ℓ_2 ball in $\mathbb{R}^{k(p-1)}$. To finish our proof, given (A.72) we show that $|u'\mathbf{\Gamma} u| \leq \frac{1}{2M}$ for any $u \in \text{cl}(\text{conv}\{\mathbb{K}(s) \cap B_2(2+\xi)\})$, the closure of the convex hull covering $\mathbb{K}(2s) \cap B_2(2+\xi)$, followed by an argument showing that $\Psi(\xi, T) \cap B_2(1) \subset \text{cl}(\text{conv}\{\mathbb{K}(s) \cap B_2(2+\xi)\})$.

For any $u \in \text{cl}(\text{conv}\{\mathbb{K}(s) \cap B_2(2+\xi)\})$, we can write $u = \sum_i \alpha_i u_i$, where $u_i \in \mathbb{K}(s)$, $\|u_i\| \leq 2+\xi$, $\alpha_i > 0$, and $\sum_i \alpha_i = 1$. Thus it follows from (A.72) and the fact of $u_i + u_j \in$

$\mathbb{K}(2s)$ for any i and j that

$$
\begin{aligned}
|u'\mathbf{\Gamma}u| &= \left|(\sum_i \alpha_i u_i)'\mathbf{\Gamma}(\sum_i \alpha_i u_i)\right| \le \sum_{i,j} \alpha_i\alpha_j\, |u_i{}'\mathbf{\Gamma}u_j| \\
&= \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j\, |(u_i+u_j)'\mathbf{\Gamma}(u_i+u_j) - u_i{}'\mathbf{\Gamma}u_i - u_j{}'\mathbf{\Gamma}u_j| \\
&\le \frac{1}{2}\frac{1}{6(2+\xi)^2 M}\sum_{i,j} \alpha_i\alpha_j \left(4(2+\xi)^2 + (2+\xi)^2 + (2+\xi)^2\right) \\
&\le \frac{1}{2M}\sum_{i,j} \alpha_i\alpha_j = \frac{1}{2M},
\end{aligned}
$$

where (A.72) has been applied in the second inequality. It remains to show that

$$\Psi(\xi, T) \cap B_2(1) \subset \mathrm{cl}(\mathrm{conv}\{\mathbb{K}(s) \cap B_2(2+\xi)\}).$$

We exploit a similar analysis to that designed for the regular sparse set (see Lemma 1,1 of [**?**]). To show that a set A belongs to a convex set B, it suffices to prove

$$\phi_A(z) \le \phi_B(z) \quad \text{for all } z \in \mathbb{R}^{k(p-1)},$$

where $\phi_A(z) = \sup_{u\in A}\langle u, z\rangle$; see, e.g., Theorem 2.3.1 of [**?**].

Hereafter we denote by $A = \Psi(\xi, T) \cap B_2(1)$ and $B = \mathrm{cl}(\mathrm{conv}\{\mathbb{K}(s) \cap B_2(2+\xi\})$. For any $z \in \mathbb{R}^{k(p-1)}$, let the index set S consist of the top s groups of z in terms of the ℓ_2 norm. Consequently, for any $l \in S^c$ we have $\|z_{(l)}\| \le (\sum_{l\in S}\|z_{(l)}\|^2)^{1/2}/\sqrt{s}$. Now we upper bound $\phi_A(z)$ by considering index sets S and S^c separately,

$$
\begin{aligned}
\phi_A(z) &\le \sup_{u\in A}\sum_{l\in S}\left\langle u_{(l)}, z_{(l)}\right\rangle + \sup_{u\in A}\sum_{l\in S^c}\left\langle u_{(l)}, z_{(l)}\right\rangle \\
&\le (\sum_{l\in S}\left\|z_{(l)}\right\|^2)^{1/2} + \max_{l\in S^c}\left\|z_{(l)}\right\| \cdot \sum_{l\in S^c}\left\|u_{(l)}\right\| \\
&\le (\sum_{l\in S}\left\|z_{(l)}\right\|^2)^{1/2}(1 + (1+\xi)\sqrt{s}/\sqrt{s}) = (2+\xi)(\sum_{l\in S}\left\|z_{(l)}\right\|^2)^{1/2},
\end{aligned}
$$

where we have used the fact that u is a unit vector and the Cauchy–Schwarz inequality in the second inequality, and the third inequality follows from the fact that

$$\sum_{l\in S^c}\|u_{(l)}\| \le \sum_{l=2}^{p}\|u_{(l)}\| \le (1+\xi)\sum_{l\in T}\|u_{(l)}\| \le (1+\xi)\sqrt{s}\|u\|$$

in light of $u \in \Psi(\xi, T)$. On the other hand, since B is a convex set we have

$$\phi_B(z) = \sup_{u \in B} \langle u, z \rangle = (2+\xi) \max_{L:|L|=s} \sup_{u \in B_2(1)} \sum_{l \in L} \langle u_{(l)}, z_{(l)} \rangle = (2+\xi)(\sum_{l \in S} \|z_{(l)}\|^2)^{1/2},$$

where we have used the definition of the index set S. Clearly, it holds that $\phi_A(z) \leq \phi_B(z)$ for all $z \in \mathbb{R}^{k(p-1)}$, which concludes the proof.

D.6 Lemma D.6 and its proof

Lemma D.6. *With the choice of regularization parameter λ specified in Theorem 3.1, the event $\mathcal{B}_1$ defined in (31) holds with probability at least $1 - 3p^{-\delta+1}$.*

Proof. Throughout this proof we condition on $\mathbf{X}^0_{*,-1}$. For any fixed $l \in [k]$, we have

$$\bar{D}_{1(l)}^{-1/2} \mathbf{X}^{0\prime}_{*,(l)} E^0_{*,1} \overset{d}{\sim} \left(N(0, n^{(1)}/\omega^{(1)}_{1,1}), \cdots, N(0, n^{(k)}/\omega^{(k)}_{1,1}) \right)',$$

where $\overset{d}{\sim}$ denotes equivalence in distribution and the k components on the right-hand side are independent of each other. By the definition of $\bar{D}_{E1}$, we can further write

$$\bar{D}_{E1}^{-1/2} \bar{D}_{1(l)}^{-1/2} \mathbf{X}^{0\prime}_{*,(l)} E^0_{*,1} \overset{d}{\sim} (T^{(1)} Z^{(1)}, \cdots, T^{(k)} Z^{(k)})',$$

where $Z^{(t)}$, $t \in [k]$, are i.i.d. standard Gaussian and $(T^{(t)})^{-2} \overset{d}{\sim} \chi^2(n^{(t)})/n^{(t)}$. Consequently, we obtain

$$\mathbb{P}\left(\left\| \bar{D}_{E1}^{-1/2} \bar{D}_{1(l)}^{-1/2} \mathbf{X}^{0\prime}_{*,(l)} E^0_{*,1} \right\|^2 > z \right) \leq \mathbb{P}\left(\max_{t \in [k]} \left(T^{(t)}\right)^2 \chi^2(k) > z \right). \tag{A.74}$$

To control the term $T^{(t)}$, we apply Lemma E.1 with $x = \tau = (8(\delta \log p + \log k)/n^{(0)})^{1/2} = o(1)$ to deduce that

$$\mathbb{P}\left(\left(T^{(t)}\right)^2 > \frac{1}{1-\tau} \right) \leq 2k^{-1} p^{-\delta}, \tag{A.75}$$

where we have used the fact of $n^{(0)} \leq n^{(t)}$. Similarly, to control the term $\chi^2(k)$ an application of Lemma E.1 with $y = \delta \log p$ leads to

$$\mathbb{P}\left(\chi^2(k) > k + 2\delta \log p + 2\sqrt{\delta k \log p} \right) \leq p^{-\delta}. \tag{A.76}$$

Thus the union bound argument applied to inequalities (A.75) over $t \in [k]$ and (A.76) yields

$$\mathbb{P}\left(\max_{t\in[k]} \left(T^{(t)}\right)^2 \chi^2(k) > \frac{k + 2\delta \log p + 2\sqrt{\delta k \log p}}{1-\tau}\right) \le 3p^{-\delta}.$$

Finally, we can apply another union bound argument over all $2 \le l \le p$ and (A.74) to obtain

$$\mathbb{P}\left(\max_{2\le l\le p} \left\|\bar{D}_{E1}^{-1/2}\bar{D}_{1(l)}^{-1/2}\mathbf{X}_{*,(l)}^{0\prime}E_{*,1}^{0}\right\|^2 > \frac{k + 2\delta \log p + 2\sqrt{\delta k \log p}}{1-\tau}\right) \le 3p^{-\delta+1},$$

which completes the proof by noting that the above conditional probability is free of $\mathbf{X}_{*,-1}^0$.

D.7 Lemma D.7 and its proof

Lemma D.7. *Under Conditions 2.1–2.2, for the event $\mathcal{E}_{1,up} = \{\zeta_t \le \sqrt{6MM_0}$ simultaneously for all $t \in [k]\}$ it holds that $\mathbb{P}\{\mathcal{E}_{1,up}\} \ge 1 - 4k\exp(-n^{(0)}/32)$.*

Proof. Be definition, we have $\zeta_t = \bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)}) + \bar{Q}_t^{1/2}(\bar{C}_1^{(t)})$. Since $\hat{\bar{C}}_1^0$ is the solution to the HGSL optimization problem (A.1), for the vector $\breve{\beta} = (\mathbf{0}, \hat{\bar{C}}_1^{(2)\prime}, \cdots, \hat{\bar{C}}_1^{(k)\prime})'$ with $\breve{\beta}_{(l)} = (0, \hat{\bar{C}}_{1,l}^{(2)}, \cdots, \hat{\bar{C}}_{1,l}^{(k)})'$ it holds that

$$\sum_{t=1}^{k} \bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)}) + \lambda \sum_{l=2}^{p} \left\|\hat{\bar{C}}_{1(l)}^{0}\right\| \le \bar{Q}_1^{1/2}(\mathbf{0}) + \sum_{t\neq t_0} \bar{Q}_t^{1/2}(\hat{\bar{C}}_1^{(t)}) + \lambda \sum_{l=2}^{p} \left\|\breve{\beta}_{(l)}\right\|.$$

Note that $\|\hat{\bar{C}}_{1(l)}^0\| \ge \|\breve{\beta}_{(l)}\|$ by our choice of $\breve{\beta}_{(l)}$. Thus we deduce that

$$\bar{Q}_1^{1/2}(\hat{\bar{C}}_1^{(1)}) \le \bar{Q}_1^{1/2}(\mathbf{0}) = \|X_{*,1}^{(1)}\|/(n^{(0)})^{1/2}.$$

By symmetry, for all $t \in [k]$ we have with probability at least $1 - 4k\exp(-n^{(0)}/32)$,

$$\begin{aligned} \zeta_t &\le \frac{\left\|X_{*,1}^{(t)}\right\| + \left\|E_{*,1}^{(t)}\right\|}{\sqrt{n^{(0)}}} \le \frac{\left\|X_{*,1}^{(t)}\right\| + \left\|E_{*,1}^{(t)}\right\|}{\sqrt{n^{(t)}}}\frac{\sqrt{n^{(t)}}}{\sqrt{n^{(0)}}} \\ &\le 2\sqrt{3M/2}\cdot\sqrt{M_0}, \end{aligned}$$

where the last inequality follows from Conditions 2.1–2.2 and the facts of $X_{*,1}^{(t)\prime}X_{*,1}^{(t)}/\sigma_{1,1}^{(t)} \sim \chi^2(n^{(t)})$ and $E_{*,1}^{(t)\prime}E_{*,1}^{(t)}(\omega_{1,1}^{(t)}) \sim \chi^2(n^{(t)})$. Specifically, the union bound for $t \in [k]$ with an application of Lemma E.1 using $x = 1/2$ yields

$$(\|X_{*,1}^{(t)}\| + \|E_{*,1}^{(t)}\|)/(n^{(t)})^{1/2} \le (3\sigma_{1,1}^{(t)}/2)^{1/2} + (3/2\omega_{1,1}^{(t)})^{1/2}$$

with probability at least $1 - 4k\exp(-n^{(0)}/32)$, which concludes the proof.

E Additional technical details

The following two technical lemmas are used throughout the paper from place to place.

Lemma E.1 ([18]). *The chi-square distribution with n degrees of freedom satisfies the following tail probability bounds*

$$\begin{aligned}
\mathbb{P}\left(\left|\chi^2(n)/n-1\right|>x\right) &\leq 2\exp(-nx(x\wedge 1)/8) \quad \textit{for any } x>0,\\
\mathbb{P}\left(\chi^2(n)/n-1>2y/n+2\sqrt{y/n}\right) &\leq \exp(-y) \quad \textit{for any } y>0,\\
\mathbb{P}\left(\sqrt{\chi^2(n)/n}-1>z\right) &\leq \exp(-nz^2/2) \quad \textit{for any } z>0.
\end{aligned}$$

Lemma E.2. *Assume that Conditions 2.1–2.2 hold and $\max\{\log p,\log k\}=o(n^{(0)})$. Then for any given constant $\delta>0$, there exists some positive constant C depending only on M and δ such that for any fixed j,*

$$\begin{aligned}
\mathbb{P}\left(\max_{l\neq j}\frac{1}{k}\sum_{t=1}^{k}(\frac{E_{*,j}^{(t)\prime}X_{*,l}^{(t)}}{n^{(t)}})^2\geq C\frac{1+(\log p)/k}{n^{(0)}}\right) &\leq 3p^{1-\delta},\\
\mathbb{P}\left(\frac{1}{k}\sum_{t=1}^{k}(\frac{E_{*,j}^{(t)\prime}\mathbf{X}_{*,-j}^{(t)}C_j^{(t)}}{n^{(t)}})^2\geq C\frac{1+(\log p)/k}{n^{(0)}}\right) &\leq 3p^{-\delta}.
\end{aligned}$$

Proof. Since $E_{*,j}^{(t)}\sim N(0,I\cdot(\omega_{j,j}^{(t)})^{-1})$ is independent of $\mathbf{X}_{*,-j}^{(t)}$ for each $t\in[k]$, it holds that for each $l\neq j$, $(E_{*,j}^{(t)\prime}X_{*,l}^{(t)})(\omega_{j,j}^{(t)})^{1/2}/\|X_{*,l}^{(t)}\|\sim N(0,1)$. In addition, these random variables are independent among different $t\in[k]$. By Lemma E.1, we have

$$\mathbb{P}\left(\frac{1}{k}\sum_{t=1}^{k}\omega_{j,j}^{(t)}\left(E_{*,j}^{(t)\prime}X_{*,l}^{(t)}/\left\|X_{*,l}^{(t)}\right\|\right)^2\geq 1+2\sqrt{\frac{\delta\log p}{k}}+\frac{2\delta\log p}{k}\right)\leq 2p^{-\delta}. \tag{A.77}$$

To control the term $\|X_{*,l}^{(t)}\|$, we apply Lemma E.1 with $X_{*,l}^{(t)}\sim N(0,I\cdot\sigma_{l,l}^{(t)})$ to deduce that

$$\mathbb{P}\left(\left\|X_{*,l}^{(t)}\right\|/\sqrt{\sigma_{l,l}^{(t)}n^{(t)}}\geq 1+\sqrt{\frac{2(\delta\log p+\log k)}{n^{(t)}}}\right)\leq p^{-\delta}k^{-1},$$

where $\sigma_{l,l}^{(t)}$ stands for the variance of $X_l^{(t)}$. The union bound, together with the assumption of $\max\{\log p,\log k\}=o(n^{(0)})$, entails that

$$\|X_{*,l}^{(t)}\|\leq 2(\sigma_{l,l}^{(t)}n^{(t)})^{1/2}\leq(4Mn^{(t)})^{1/2} \tag{A.78}$$

simultaneously for all $t \in [k]$ with probability at least $1 - p^{-\delta}$.

We now condition on the event given by (A.78). Due to Conditions 2.1–2.2, we have

$$\frac{1}{k}\sum_{t=1}^{k} \omega_{j,j}^{(t)} \left(\frac{E_{*,j}^{(t)\prime} X_{*,l}^{(t)}}{\left\| X_{*,l}^{(t)} \right\|} \right)^2 \geq \frac{n^{(0)}}{4M^2} \frac{1}{k} \sum_{t=1}^{k} \left(\frac{E_{*,j}^{(t)\prime} X_{*,l}^{(t)}}{n^{(t)}} \right)^2,$$

which along with (A.77) leads to

$$\mathbb{P}\left(\frac{1}{k}\sum_{t=1}^{k} \left(E_{*,j}^{(t)\prime} X_{*,l}^{(t)} / n^{(t)} \right)^2 \geq \frac{4M^2}{n^{(0)}} \left(1 + 2\sqrt{\frac{\delta \log p}{k}} + \frac{2\delta \log p}{k} \right) \right) \leq 3p^{-\delta}. \tag{A.79}$$

Thus we see that the first desired result follows immediately from (A.79) with a union bound for all $l \neq j$ and $C = 4M^2(2 + 3\delta)$, in view of $2((\delta \log p)/k)^{1/2} \leq 1 + (\delta \log p)/k$. Since $\mathbf{X}_{*,-1}^{(t)} C_1^{(t)}$ has i.i.d. Gaussian entries with bounded variance and is independent of $E_{*,j}^{(t)}$, the second desired result follows from a similar analysis as for (A.79), which completes the proof.