Supplement to "Iterative Algorithm for Discrete Structure Recovery"

Chao Gao[1] and Anderson Y. Zhang[2]

[1] *University of Chicago*
[2] *University of Pennsylvania*

The supplement includes a few more examples and all the technical proofs. We first analyze approximate ranking in Appendix A. $\mathbb{Z}/k\mathbb{Z}$ synchronization and permutation synchronization are studied in Appendix B and Appendix C, respectively. We then prove Theorem 3.1 in Appendix D. The rest of the proofs are organized from Appendix E to Appendix I.

A Approximate Ranking

In this section, we study the estimation of $z^* \in [p]^p$ using the pairwise interaction data generated according to $Y_{ij} \sim \mathcal{N}(\beta^*(z_i^* - z_j^*), 1)$ independently for all $1 \leq i \neq j \leq p$. This model can be viewed as a special case of the more general pairwise comparison model $Y_{ij} \sim \mathcal{N}(\theta^*_{z_i^*} - \theta^*_{z_j^*}, 1)$, where θ_i^* parametrizes the ability of the ith player, and the choice $\theta_i^* = \alpha^* + \beta^* i$ leads to $Y_{ij} \sim \mathcal{N}(\beta^*(z_i^* - z_j^*), 1)$ that will be studied in this section. Let Π_p be the set of all possible permutations of $[p]$. We assume the rank vector z^* belongs to the following class,

$$\mathcal{R} = \left\{ z \in [p]^p : \min_{\widetilde{z} \in \Pi_p} \|z - \widetilde{z}\|^2 \leq c_p \right\}, \tag{A.1}$$

for some sequence $1 \leq c_p = o(p)$. In other words, $\mathcal{R}$ is a set of approximate permutations. A rank vector $z^* \in \mathcal{R}$ is allowed to have ties and not necessarily to start from 1. To be more precise, a $z^* \in \mathcal{R}$ should be interpreted as discrete positions of the p players in the latent space of their abilities. This is in contrast to the exact ranking problem, also known as "noisy sorting" in the literature, where z^* is assumed to be a permutation [1, 11, 14].

For the loss function

$$L_2(z, z^*) = \frac{1}{p} \sum_{j=1}^{p} (z_j - z_j^*)^2, \tag{A.2}$$

the minimax rate of estimating z^* takes the following formula,

$$\inf_{\widehat{z}} \sup_{z^* \in \mathcal{R}} \mathbb{E} L_2(\widehat{z}, z^*) \asymp \begin{cases} \exp\left(-(1+o(1))\frac{p(\beta^*)^2}{4}\right), & p(\beta^*)^2 > 1, \\ \frac{1}{p(\beta^*)^2} \wedge p^2, & p(\beta^*)^2 \leq 1. \end{cases} \tag{A.3}$$

See Theorems 2.2 and 2.3 in [5][1]. Interestingly, the minimax rate either takes a polynomial form or an exponential form, depending on the signal strength parametrized by $p(\beta^*)^2$. In the paper [5], a combinatorial procedure is constructed to achieve the optimal rate (A.3), and whether (A.3) can be achieved by a polynomial-time algorithm is unknown. This is where our proposed iterative algorithm comes. We will particularly focus on the regime of $p(\beta^*)^2 \to \infty$, where the minimax rate takes an exponential form.

Specializing Algorithm 1 to the approximate ranking problem, we can write the iterative feature matching algorithm as

$$z_j^{(t)} = \underset{a\in[p]}{\operatorname{argmin}} \left| \sum_{i\in[p]\backslash\{j\}} (Y_{ji} - Y_{ij}) - 2p\widehat{\beta}(z^{(t-1)}) \left(a - \frac{p+1}{2}\right) \right|^2, \quad j \in [p],$$

where for each $z \in [p]^p$, we use the notation

$$\widehat{\beta}(z) = \frac{\sum_{1\leq i\neq j\leq p}(z_i - z_j)Y_{ij}}{\sum_{1\leq i\neq j\leq p}(z_i - z_j)^2}. \tag{A.4}$$

A.1 Conditions

From (11), we have

$$\Delta_j(a,b)^2 = \frac{2p^2(\beta^*)^2}{p-1}\left[(a-b)^2 - 2(a-b)\left(\frac{1}{p}\sum_{j=1}^p z_j^* - \frac{p+1}{2}\right)\right], \tag{A.5}$$

and

$$\ell(z,z^*) = \frac{2p^2(\beta^*)^2}{p-1}\sum_{j=1}^p (z_j - z_j^*)^2 \tag{A.6}$$

in the current setting. It is easy to check that $\Delta_j(a,b)^2 > 0$ for all $a \neq b$ as long as $z^* \in \mathcal{R}$. From (10) and (11), we have $T_j = \mu_j(B^*, z_j^*) + \epsilon_j$ where $\epsilon_j \sim \mathcal{N}(0,1)$ for all $j \in [p]$. One can also write down the formulas of $F_j(a,b;z)$, $G_j(a,b;z)$ and $H_j(a,b)$, which are included in Appendix F due to the limit of space.

Lemma A.1. *Assume $z^* \in \mathcal{R}$, $\tau = o(p^2(\beta^*)^2)$, and $p(\beta^*)^2 \geq 1$. Then, for any $C' > 0$, there exists a constant $C > 0$ only depending on C' such that*

$$\max_{\{z:\ell(z,z^*)\leq\tau\}} \sum_{j=1}^p \max_{b\in[p]\backslash\{z_j^*\}} \frac{F_j(z_j^*,b;z)^2\|\mu_j(B^*,b) - \mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)} \leq Cp^{-2}, \tag{A.7}$$

$$\begin{aligned}\max_{\{z:\ell(z,z^*)\leq\tau\}} \quad & \max_{T\subset[p]} \frac{\tau}{4\Delta_{\min}^2|T| + \tau} \sum_{j\in T} \max_{b\in[p]\backslash\{z_j^*\}} \frac{G_j(z_j^*,b;z)^2\|\mu_j(B^*,b) - \mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)} \\ & \leq C\left(\frac{\tau}{p^2|\beta^*|^2} + \frac{1}{p|\beta^*|^2}\right),\end{aligned} \tag{A.8}$$

[1]The paper [5] considers a parameter space that is slightly different from $\mathcal{R}$. However, the proof of [5] can be modified so that the same minimax rate also applies to $\mathcal{R}$

and

$$\max_{j\in[p]}\max_{b\in[p]\backslash\{z_j^*\}}\frac{|H_j(z_j^*,b)|}{\Delta_j(z_j^*,b)^2}\leq C\frac{1}{\sqrt{p}|\beta^*|},\tag{A.9}$$

with probability at least $1-(C'p)^{-1}$ *for a sufficiently large* p.

Lemma A.1 implies that Conditions A , B and C hold with some sequence $\delta=\delta_p=o(1)$ as long as $\tau=o(p^2(\beta^*)^2)$ and $p(\beta^*)^2\to\infty$.

Next, we need to control $\xi_{\text{ideal}}(\delta)$ in Condition D. This is given by the following lemma.

Lemma A.2. *Assume* $p(\beta^*)^2\to\infty$. *Then, for any sequence* $\delta_p=o(1)$, *we have*

$$\xi_{\text{ideal}}(\delta_p)\leq p\exp\left(-(1+o(1))\frac{p(\beta^*)^2}{4}\right),$$

with probability at least $1-\exp\left(-\sqrt{p(\beta^*)^2}\right)-p^{-1}$.

We note that the signal condition $p(\beta^*)^2\to\infty$ implies that Conditions A, B, C and D hold simultaneously.

A.2 Convergence

With the help of Lemma A.1 and Lemma A.2, we can specialize Theorem 3.1 into the following result.

Theorem A.1. *Assume* $p(\beta^*)^2\to\infty$ *and* $z^*\in\mathcal{R}$. *Suppose* $z^{(0)}$ *satisfies* $\ell(z^{(0)},z^*)=o(p^2(\beta^*)^2)$ *with probability at least* $1-\eta$. *Then, we have*

$$\ell(z^{(t)},z^*)\leq p\exp\left(-(1+o(1))\frac{p(\beta^*)^2}{4}\right)+\frac{1}{2}\ell(z^{(t-1)},z^*)\quad\text{for all } t\geq 1,$$

with probability at least $1-\eta-\exp\left(-\sqrt{p(\beta^*)^2}\right)-2p^{-1}$.

Using the relation from (A.2) and (A.6) that

$$L_2(z,z^*)=\frac{p-1}{2p^3(\beta^*)^2}\ell(z,z^*),\tag{A.10}$$

we immediately obtain the following result on the loss $L_2(z,z^*)$.

Corollary A.1. *Assume* $p(\beta^*)^2\to\infty$ *and* $z^*\in\mathcal{R}$. *Suppose* $z^{(0)}$ *satisfies* $\ell(z^{(0)},z^*)=o(p^2(\beta^*)^2)$ *with probability at least* $1-\eta$. *Then, we have*

$$L_2(z^{(t)},z^*)\leq\exp\left(-(1+o(1))\frac{p(\beta^*)^2}{4}\right)+2^{-t}\quad\text{for all } t\geq 1,\tag{A.11}$$

with probability at least $1-\eta-\exp\left(-\sqrt{p(\beta^*)^2}\right)-2p^{-1}$.

We observe that $L_2(z, z^*)$ takes value in the set $\{j/p : j \in \mathbb{N} \cup \{0\}\}$, the term 2^{-t} in (A.11) is negligible as long as $2^{-t} = o(p^{-1})$. We therefore can claim

$$L_2(z^{(t)}, z^*) \leq \exp\left(-(1+o(1))\frac{p(\beta^*)^2}{4}\right) \quad \text{for all } t \geq 3\log p.$$

Hence, by (A.3) the iterative feature matching algorithm achieves the minimax rate of approximate ranking in the regime of $p(\beta^*)^2 \to \infty$ after at most $\lceil 3\log p \rceil$ iterations.

A.3 Initialization

To initialize the iterative feature matching algorithm, we consider a simple ranking procedure based on the statistics $\{T_j\}_{j\in[p]}$. That is, letting $T_{(1)} \leq \cdots \leq T_{(p)}$ be the order statistics of $\{T_j\}_{j\in[p]}$, we define $z^{(0)}$ to be a permutation vector that satisfies $T_{z_j^{(0)}} = T_{(j)}$ for all $j \in [p]$.

Proposition A.1. *Assume $z^* \in \mathcal{R}$ and $\beta^* > 0$. Then, we have*

$$L_2(z^{(0)}, z^*) \lesssim \begin{cases} o(1), & p(\beta^*)^2 \to \infty, \\ \frac{1}{p(\beta^*)^2} \wedge p^2, & p(\beta^*)^2 = O(1), \end{cases}$$

with probability at least $1 - p^{-1}$.

Note that the additional condition $\beta^* > 0$ guarantees that $z^{(0)}$ estimates z^* instead of its reverse order. In the regime of $p(\beta^*)^2 \to \infty$, the initialization procedure achieves $L_2(z^{(0)}, z^*) = o(1)$ with high probability. Given the relation (A.10), this implies that $\ell(z^{(0)}, z^*) = o(p^2(\beta^*)^2)$, and thus the initialization condition of Theorem A.1 is satisfied. In the regime of $p(\beta^*)^2 = O(1)$, the initialization procedure achieves the rate $\frac{1}{p(\beta^*)^2} \wedge p^2$, which is already minimax optimal according to (A.3), and there is no need for the improvement via the iterative algorithm.

B $\mathbb{Z}/k\mathbb{Z}$ Synchronization

$\mathbb{Z}/k\mathbb{Z}$ synchronization is known as the famous problem of joint alignment from pairwise differences [3]. It has some interesting applications in the haplotype assembly problem [15] and and Dixon imaging [18]. The group $\mathbb{Z}/k\mathbb{Z}$ consists of elements $\{0, 1, 2, \cdots, k-1\}$. For any $g, h \in \mathbb{Z}/k\mathbb{Z}$, the group operation is defined by $g \circ h = (g + h) \bmod k$. Since it implies the inverse of an element g is $g^{-1} = (k - g)$ mode k, which is equal to $k - g$ if $g \neq 0$ and 0 if $g = 0$. Then we have $g \circ h^{-1} = (g - h) \bmod k$.

In $\mathbb{Z}/k\mathbb{Z}$ synchronization, we have independent observations $Y_{ij} \sim \mathcal{N}(\lambda^*(z_i^* \circ {z_j^*}^{-1}), 1)$ for all $1 \leq i \neq j \leq p$ with $\lambda^* \in \mathbb{R}$ and $z_1^*, \cdots, z_p^* \in \mathbb{Z}/k\mathbb{Z}$, and our goal is to recover $z_1^*, \cdots, z_p^*$. When $k = 2$, the problem is very similar but not identical to $\mathbb{Z}_2$ synchronization. Though $\mathbb{Z}/2\mathbb{Z} = \{0, 1\}$ is isomorphic to $\mathbb{Z}_2 = \{-1, 1\}$, the current model assumption $Y_{ij} \sim \mathcal{N}(\lambda^*(z_i^* \circ {z_j^*}^{-1}), 1)$ does not lead to the symmetric property $\mathbb{E}Y_{ij} = \mathbb{E}Y_{ji}$ in $\mathbb{Z}_2$ synchronization.

This is because $(z_i^*-z_j^*)$ mod $k \neq (z_j^*-z_i^*)$ mod k. As a consequence, it is required that both Y_{ij} and Y_{ji} are observed in the more general setting of $\mathbb{Z}/k\mathbb{Z}$ synchronization.

Minimax rate for estimating $z_1^*,\cdots,z_p^*$ is unknown in the literature. We first present a minimax lower bound for the problem.

Theorem B.1. *If $p\lambda^{*2}\to\infty$, we have*

$$\inf_{\widehat{z}}\sup_{z^*}\mathbb{E}\min_{a\in\mathbb{Z}/k\mathbb{Z}}\left(\frac{1}{p}\sum_{j=1}^{p}\mathbf{1}_{\left\{\widehat{z}_j\neq z_j^*\circ a^{-1}\right\}}\right)\geq\exp\left(-\frac{(1+o(1))p\lambda^{*2}}{4}\right).$$

Otherwise if $p\lambda^{*2}=O(1)$, we then have

$$\inf_{\widehat{z}}\sup_{z^*}\mathbb{E}\min_{a\in\mathbb{Z}/k\mathbb{Z}}\left(\frac{1}{p}\sum_{j=1}^{p}\mathbf{1}_{\left\{\widehat{z}_j\neq z_j^*\circ a^{-1}\right\}}\right)\geq c,$$

for some constant $c>0$.

Next, we will show the above minimax lower bound can be achieved via an computationally efficient algorithm under the signal-to-noise ratio condition $\frac{p\lambda^{*2}}{k^7\log k}\to\infty$. We first explain how to view the $\mathbb{Z}/k\mathbb{Z}$ synchronization problem from the perspective of our general framework. Though it is not obvious because of the nonlinear group operation in the model, we can still adopt a similar idea used in $\mathbb{Z}_2$ synchronization and view one of the z_j^* in $\lambda^*(z_i^*\circ z_j^{*^{-1}})$ and λ^* jointly as the continuous model parameter. Write $Y=\mathbb{E}Y+W\in\mathbb{R}^{p\times p}$ with $\mathbb{E}Y_{ij}=\lambda^*(z_i^*\circ z_j^{*-1})$ and $W_{ij}\sim\mathcal{N}(0,1)$ for all $i\neq j$ and $\mathbb{E}Y_{ii}=W_{ii}=0$ for all i. Define $\beta^*=(\lambda^*,z^*)$. It is clear that $\beta^*\in\mathcal{B}_{z^*}=\{(\lambda,z^*):\lambda\in\mathbb{R}\}=\mathbb{R}\times\{z^*\}$. We can then write $\mathbb{E}Y=\mathscr{X}_{z^*}(\beta^*)$, where the operator $\mathscr{X}_{z^*}(\cdot)$ is determined by $[\mathscr{X}_{z^*}(\beta^*)]_{ij}=[\mathscr{X}_{z^*}((\lambda^*,z^*))]_{ij}=\lambda^*(z_i^*\circ z_j^{*-1})$.

To derive an iterative algorithm, we define the local statistic $T_j=Y_j\in\mathbb{R}^p$, the jth column of the matrix Y. We then have $\mathbb{E}T_j=\mu_j(\beta^*,z_j^*)=\nu_j(\beta^*,z_j^*)\in\mathbb{R}^p$. For any $i\in[p]$ and any $a\in\mathbb{Z}/k\mathbb{Z}$, the ith entry of $\mu_j(\beta^*,a)=\nu_j(\beta^*,a)$ is

$$[\mu_j(\beta^*,a)]_i=[\nu_j(\beta^*,a)]_i=\lambda^*(z_i^*\circ a^{-1}).$$

Then, we can specialize Algorithm 1 into the following iterative procedure,

$$\beta^{(t)} = \operatorname*{argmin}_{\beta=(\lambda,z^{(t-1)}):\lambda\in\mathbb{R}}\|Y-\mathscr{X}_{z^{(t-1)}}(\beta)\|_{\mathrm{F}}^2, \tag{B.1}$$

$$z_j^{(t)} = \operatorname*{argmin}_{a\in\mathbb{Z}/k\mathbb{Z}}\|Y_j-\mu_j(\beta^{(t)},a)\|^2. \tag{B.2}$$

The second step (B.2) is straightforward, since one can easily evaluate $\|Y_j-\mu_j(\beta^{(t-1)},a)\|^2$ for all $a\in\mathbb{Z}/k\mathbb{Z}$. The first step (B.1) looks complicated, but thanks to the constraint $\beta=(\lambda,z^{(t-1)})$, it is a simple one-dimensional optimization problem. In fact, there is a closed form solution to (B.1), and we can use that to write down an equivalent form of the algorithm (B.1)-(B.2) as follows,

$$\lambda^{(t)} = \frac{\sum_{1\leq i\neq j\leq p} Y_{ij}\left(z_i^{(t-1)} \circ (z_j^{(t-1)})^{-1}\right)}{\sum_{1\leq i\neq j\leq p}\left(z_i^{(t-1)} \circ (z_j^{(t-1)})^{-1}\right)^2}, \tag{B.3}$$

$$z_j^{(t)} = \operatorname*{argmin}_{a\in\mathbb{Z}/k\mathbb{Z}} \sum_{i=1}^{p}\left(Y_{ij} - \lambda^{(t)}(z_i^{(t-1)} \circ a^{-1})\right)^2. \tag{B.4}$$

This example really demonstrates the flexibility of our general framework, especially the possible dependence of the space $\mathcal{B}_z$ on the discrete structure z. It allows us to separate z_i from z_j in $z_i \circ z_j^{-1}$, and derive the simple iterative algorithm (B.3)-(B.4).

B.1 Conditions

To analyze the algorithmic convergence of (B.3)-(B.4), we note that

$$\Delta_j(a,b)^2 = \lambda^{*2}\sum_{i=1}^{p}\left((z_i^* \circ a^{-1}) - (z_i^* \circ b^{-1})\right)^2,$$

under the current setting. Thus, the loss function is

$$\ell(z,z^*) = \lambda^{*2}\sum_{j=1}^{p}\sum_{i=1}^{p}\left((z_i^* \circ z_j^{-1}) - (z_i^* \circ z_j^{*-1})\right)^2. \tag{B.5}$$

One can also write down the formulas of $F_j(a,b;z)$, $G_j(a,b;z)$ and $H_j(a,b)$, which are given in Appendix I.2. The error terms are controlled by the following lemma.

Lemma B.1. *Assume $\frac{p\lambda^{*2}}{k^4} \to \infty$ and $\max_{a\in\mathbb{Z}/k\mathbb{Z}} \sum_{j=1}^{p} \mathbf{1}_{\{z_j^*=a\}} \leq (1-\alpha)p$ for some constant $\alpha > 0$. Then, for any constant $C' > 0$, there exists a constant $C > 0$ only depending on C' such that*

$$\max_{z:\ell(z,z^*)\leq\tau}\sum_{j=1}^{p}\max_{b\in\mathbb{Z}/k\mathbb{Z}\setminus\{z_j^*\}}\frac{F_j(z_j^*,b;z)^2\left\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\right\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)} \leq C\frac{k^4}{p\lambda^{*2}}, \tag{B.6}$$

$$\begin{aligned}\max_{\{z:\ell(z,z^*)\leq\tau\}} \quad & \max_{T\subset[p]}\frac{\tau}{4\Delta_{\min}^2|T|+\tau}\sum_{j\in T}\max_{b\in\mathbb{Z}/k\mathbb{Z}\setminus\{z_j^*\}}\frac{G_j(z_j^*,b;z)^2\left\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\right\|^2}{\Delta_j(z_j^*,b)^2\ell(z,z^*)} \\ & \leq C\frac{\tau k^6}{p^2\lambda^{*2}}, \end{aligned} \tag{B.7}$$

and

$$\max_{j\in[p]}\max_{a\in\mathbb{Z}/k\mathbb{Z}\setminus\{z_j^*\}}\frac{|H_j(z_j^*,a)|}{\Delta_j(z_j^*,a)^2} \leq C\left(\frac{k^2}{p\lambda^{*2}} + \sqrt{\frac{k^2}{p\lambda^{*2}}}\right), \tag{B.8}$$

with probability at least $1-e^{-C'p}$.

From the bounds (B.6)-(B.8), we can see that a sufficient condition under which Conditions A, B and C hold is $\tau = o\left(\frac{p^2\lambda^{*2}}{k^6}\right)$ and

$$\frac{p\lambda^{*2}}{k^4} \to \infty. \tag{B.9}$$

The signal-to-noise ratio condition (B.9) extends the condition $p\lambda^{*2} \to \infty$ required by $\mathbb{Z}_2$ synchronization.

Next, we need to bound ξ_{ideal} in Condition D. This is given by the following lemma.

Lemma B.2. *Assume $p\lambda^{*2} \to \infty$, $p/k^2 \to \infty$ and $\max_{a\in\mathbb{Z}/k\mathbb{Z}} \sum_{j=1}^{p} \mathbf{1}_{\{z_j^*=a\}} \leq (1-\alpha)p$ for some constant $\alpha > 0$. Then, for any sequence $\delta_p = o(1)$, we have*

$$\xi_{\text{ideal}}(\delta_p) \leq p \exp\left(-(1+o(1))\frac{p\lambda^{*2}}{8}\right),$$

with probability at least $1 - \exp(-\sqrt{p\lambda^{*2}})$.

Thus, under the conditions $p/k^2 \to \infty$ and (B.9), Conditions A, B, C and D hold simultaneously.

B.2 Convergence

With the help of Lemma B.1 and Lemma B.2, we can specialize Theorem 3.1 into the following result.

Theorem B.2. *Assume $p/k^2 \to \infty$, $\frac{p\lambda^{*2}}{k^4} \to \infty$, and $\max_{a\in\mathbb{Z}/k\mathbb{Z}} \sum_{j=1}^{p} \mathbf{1}_{\{z_j^*=a\}} \leq (1-\alpha)p$ for some constant $\alpha > 0$. Suppose $z^{(0)}$ satisfies*

$$\ell(z^{(0)}, z^*) = o\left(\frac{p^2\lambda^{*2}}{k^6}\right), \tag{B.10}$$

with probability at least $1-\eta$. Then, we have

$$\ell(z^{(t)}, z^*) \leq p\exp\left(-(1+o(1))\frac{p\lambda^{*2}}{8}\right) + \frac{1}{2}\ell(z^{(t-1)}, z^*) \quad \text{for all } t \geq 1,$$

with probability at least $1 - \eta - \exp(-\sqrt{p\lambda^{*2}}) - e^{-p}$.

According to the definition of the loss in (B.5), we have the inequality

$$\ell(z, z^*) \geq p^2\lambda^{*2}\frac{1}{p}\sum_{j=1}^{p} \mathbf{1}_{\{z_j \neq z_j^*\}}. \tag{B.11}$$

This immediately implies the following corollary for the Hamming loss.

Corollary B.1. *Assume* $p/k^2 \to \infty$, $\frac{p\lambda^{*2}}{k^4} \to \infty$ *and* $\max_{a \in \mathbb{Z}/k\mathbb{Z}} \sum_{j=1}^p \mathbf{1}_{\{z_j^* = a\}} \leq (1-\alpha)p$ *for some constant* $\alpha > 0$. *Suppose* $z^{(0)}$ *satisfies (B.10) with probability at least* $1 - \eta$. *Then, we have*

$$\frac{1}{p} \sum_{j=1}^p \mathbf{1}_{\left\{z_j^{(t)} \neq z_j^*\right\}} \leq \exp\left(-(1+o(1))\frac{p\lambda^{*2}}{8}\right) + 2^{-t} \quad \text{for all } t \geq 1, \tag{B.12}$$

with probability at least $1 - \eta - \exp(-\sqrt{p\lambda^{*2}}) - e^{-p}$.

By the property of the Hamming loss, the algorithmic error 2^{-t} is negligible after $\lceil 3 \log p \rceil$ iterations, and we have

$$\frac{1}{p} \sum_{j=1}^p \mathbf{1}_{\left\{z_j^{(t)} \neq z_j^*\right\}} \leq \exp\left(-(1+o(1))\frac{p\lambda^{*2}}{8}\right) \quad \text{for all } t \geq 3 \log p.$$

Thus, the minimax rate is achieved given that the initialization condition (B.10) is satisfied.

B.3 Initialization

Unlike the $\mathbb{Z}_2$ synchronization setting, the vector $z^* \in \mathbb{R}^p$ does not correspond to any eigenvector of $\mathbb{E}Y$. However, we observe that the columns of $\mathbb{E}Y$ has only k possibilities, and the k different vectors that each column can take are well separated. We thus propose the following initialization algorithm based on a spectral clustering step.

1. Apply the spectral clustering algorithm (38)-(39) to Y and obtain the column clustering label vector $\bar{z} \in \{0, 1, \cdots, k-1\}^p$.
2. For any $l \in \{1, 2, \cdots, k-1\}$, compute $\bar{Y}_l = \frac{1}{|\bar{\mathcal{Z}}_{l0}|} \sum_{(i,j) \in \bar{\mathcal{Z}}_{l0}} Y_{ij}$, where $\bar{\mathcal{Z}}_{ab} = \{(i,j) \in [p] \times [p] : \bar{z}_i = a, \bar{z}_j = b\}$ for any $a, b \in [k]$.
3. Sort $|\bar{Y}_1|, \cdots, |\bar{Y}_{k-1}|$ into the order statistics $|\bar{Y}|_{(1)} \leq \cdots \leq |\bar{Y}|_{(k-1)}$. Let $\widehat{\pi}$ be a permutation of the labels $l \in \{1, 2, \cdots, k-1\}$ so that $|\bar{Y}|_{(\widehat{\pi}(l))} = |\bar{Y}_l|$ for all $l \in \{1, 2, \cdots, k-1\}$.
4. Output the estimator $z_j^{(0)} = \widehat{\pi}(\bar{z}_j)$ for $j \in [p]$.

Let us give some intuitions why the above algorithm works. By Proposition 4.1, it is clear that there exists some label permutation π such that $\pi(\bar{z}_j)$ recovers the underlying true label z_j^*. This is the purpose of Step 1. We then use Steps 2-4 to recover this unknown label permutation π. Under an appropriate signal-to-noise ratio condition, we can show that $\max_{l \in \{1, \cdots, k-1\}} |\bar{Y}_l - \lambda^*(\pi(l) \circ \pi(0)^{-1})| = o(|\lambda^*|)$. This error bound immediately implies

$$\max_{l \in \{1, \cdots, k-1\}} ||\bar{Y}_l| - |\lambda^*|(\pi(l) \circ \pi(0)^{-1})| = o(|\lambda^*|).$$

By the fact that $\min_{a \neq b} ||\lambda^*|(\pi(a) \circ \pi(0)^{-1}) - |\lambda^*|(\pi(b) \circ \pi(0)^{-1})| \geq |\lambda^*|$, we can deduce the fact that the order of $\{|\bar{Y}_l|\}_{1 \leq l \leq k-1}$ perfectly preserves that of $\{\pi(l) \circ \pi(0)^{-1}\}_{1 \leq l \leq k-1}$. Therefore, the unknown label permutation π can be recovered via sorting $\{|\bar{Y}_l|\}_{1 \leq l \leq k-1}$.

Proposition B.1. *Assume* $\min_{a\in\mathbb{Z}/k\mathbb{Z}}\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}} \geq \frac{\alpha p}{k}$ *for some constant* $\alpha > 0$ *and* $\frac{p\lambda^{*2}}{(M+1)k^3} \to \infty$*. For any constant* $C' > 0$*, there exists a constant* $C > 0$ *only depending on* α *and* C' *such that*

$$\min_{a\in\mathbb{Z}/k\mathbb{Z}} \ell(z^{(0)}, z^* \circ a^{-1}) \leq C(M+1)kp,$$

with probability at least $1 - e^{-C'p}$*. We have used the notation* $z^* \circ a^{-1}$ *for the vector* $\{z_i^* \circ a^{-1}\}_{i\in[p]}$*.*

Proposition B.1 shows that $z^{(0)}$ achieves the rate $O((M+1)kp)$ for estimating z^* up to a group multiplication. This uncertainty cannot be avoided since $z_i \circ z_j^{-1} = (z_i \circ a^{-1}) \circ (z_j \circ a^{-1})^{-1}$. The factor M in the bound comes from the computation of the M-approximation of the k-means objective, and one can take $M = O(\log k)$ when the k-means++ algorithm is used for the approximation. In order that (B.10) is satisfied, we thus require

$$\frac{p\lambda^{*2}}{(M+1)k^7} \to \infty, \tag{B.13}$$

a condition that implies (B.9). Hence, according to Corollary B.1, the iterative algorithm initialized by spectral clustering converges to the minimax error with a linear rate under the condition (B.13).

C Permutation Synchronization

Permutation synchronization is a problem first proposed by [13] in computer vision as an approach to align multiple images from observations of pairwise similarities. Polynomial algorithms for this problem are mostly based on convex relaxation [4, 8, 17, 19] with theoretical guarantees in the form of exact recovery [4] and polynomial convergence rate in Frobenius norm [8]. In this section, we will derive the minimax rate of the problem with exponential convergence and show the optimal rate can be achieved by Algorithm 1 under some signal-to-noise ratio condition.

Let Π_d be the set of permutations on the set $[d] = \{1, \cdots, d\}$. The set of permutation matrices is defined by $\mathcal{P}_d = \{(e_{\pi(1)}, \cdots, e_{\pi(d)}) : \pi \in \Pi_d\}$, where $e_i \in \mathbb{R}^d$ is the ith canonical vector of $\mathbb{R}^d$. In permutation synchronization, we observe $Y_{ij} = \lambda^* Z_i^* Z_j^{*T} + W_{ij} \in \mathbb{R}^{d\times d}$ for $1 \leq i < j \leq p$ with $\lambda^* \in \mathbb{R}$, $Z_1^*, \cdots, Z_p^* \in \mathcal{P}_d$, and W_{ij}'s are independent error matrices with i.i.d. entries following $\mathcal{N}(0, 1)$.

We first present the minimax lower bound for the problem.

Theorem C.1. *If* $p\lambda^{*2} \to \infty$*, we have*

$$\inf_{\widehat{Z}} \sup_{Z^*} \mathbb{E} \min_{U\in\mathcal{P}_d} \frac{1}{p}\sum_{j=1}^{p}\mathbf{1}_{\{\widehat{Z}_j \neq Z_j^* U^T\}} \geq \exp\left(-\frac{(1+o(1))p\lambda^{*2}}{2}\right).$$

Otherwise if $p\lambda^{*2} = O(1)$*, we then have*

$$\inf_{\widehat{Z}} \sup_{Z^*} \mathbb{E} \min_{U\in\mathcal{P}_d} \frac{1}{p}\sum_{j=1}^{p}\mathbf{1}_{\{\widehat{Z}_j \neq Z_j^* U^T\}} \geq c,$$

for some constant $c > 0$.

Computationally efficient algorithms to recover the permutation matrices $Z_1^*, \cdots, Z_p^*$ with minimax error are unknown in the literature. The problem is hard even when the dimension of the permutation d is a constant. We will show our general iterative algorithm leads to a solution of this open problem whenever the signal-to-noise ratio conditoin $\frac{p\lambda^{*2}}{d^2} \to \infty$ is satisfied. We first put the problem into our general framework by organizing all the observations $\{Y_{ij}\}$ into a single matrix $Y = \lambda^* Z^* Z^{*T} + W \in \mathbb{R}^{pd \times pd}$. Here, $Z^{*T} = (Z_1^{*T}, \cdots, Z_p^{*T}) \in \mathbb{R}^{d \times pd}$ is a matrix by concatenating the p permutation matrices together. For each W_{ij}, it can be viewed as the (i,j)th block of W. We have $W_{ij} = W_{ji}^T$ to be independent standard Gaussian matrices for all $1 \leq i < j \leq p$ and $W_{ii} = 0$ for all $i \in [p]$. We identify z^* with Z^* and define $\mathcal{B}_{Z^*} = \{\lambda Z^* : \lambda \in \mathbb{R}\}$ as the space of model parameter. Then, we can write $Y = Z^*(B^*)^T + W$ with $B^* \in \mathcal{B}_{Z^*}$. This is the same strategy that has been used for $\mathbb{Z}_2$ synchronization. To derive an iterative algorithm, let $T_j = Y_j$ and the definition of the matrix $Y_j \in \mathbb{R}^{pd \times d}$ is given by $Y_j^T = (Y_{1j}^T, \cdots, Y_{pj}^T)$. We then have $\nu_j(B^*, U) = \mu_j(B^*, U) = B^* U^T$ for any $U \in \mathcal{P}_d$. The iterative algorithm is

$$B^{(t)} = \underset{B = \lambda Z^{(t-1)} : \lambda \in \mathbb{R}}{\operatorname{argmin}} \|Y - Z^{(t-1)} B^T\|_{\mathrm{F}}^2, \tag{C.1}$$

$$Z_j^{(t)} = \underset{U \in \mathcal{P}_d}{\operatorname{argmin}} \|Y_j - B^{(t)} U^T\|_{\mathrm{F}}^2. \tag{C.2}$$

The computation of (C.1) is a one-dimensional optimization problem, and its solution is given by $B^{(t)} = \widehat{\lambda}(Z^{(t-1)}) Z^{(t-1)}$, with $\widehat{\lambda}(Z) = \frac{\langle Y, ZZ^T \rangle}{p^2 d^2}$. For (C.2), we have the equivalent form

$$Z_j^{(t)} = \underset{U \in \mathcal{P}_d}{\operatorname{argmax}} \left\langle Y_j^T B^{(t)}, U \right\rangle,$$

which can be solved by the Kuhn-Munkres algorithm [6] with $O(d^3)$ complexity.

C.1 Conditions

To analyze the algorithmic convergence of (C.1)-(C.2), we note that

$$\Delta_j(U, V)^2 = \|B^*(U - V)^T\|_{\mathrm{F}}^2 = p\lambda^{*2} \|U - V\|_{\mathrm{F}}^2,$$

for any $U, V \in \mathcal{P}_d$. Therefore, the natural loss function of the problem is

$$\ell(Z, Z^*) = p\lambda^{*2} \sum_{j=1}^p \|Z_j - Z_j^*\|_{\mathrm{F}}^2.$$

To write down the error terms, we introduce the notation $\widehat{B}(Z) = \widehat{\lambda}(Z) Z$. The error terms are

$$\begin{aligned} F_j(U, V; Z) &= \left\langle \epsilon_j, (\widehat{B}(Z^*) - \widehat{B}(Z))(U - V)^T \right\rangle, \\ G_j(U, V; Z) &= \left\langle \widehat{B}(Z^*) - \widehat{B}(Z), B^*(I_d - U^T V) \right\rangle, \\ H_j(U, V) &= \left\langle B^* - \widehat{B}(Z^*), B^*(I_d - U^T V) \right\rangle. \end{aligned}$$

Here $\epsilon_j \in \mathbb{R}^{pd\times d}$ is an error matrix defined by $\epsilon_j^T = (W_{1j}^T, \cdots, W_{pj}^T)$. The error terms are controlled by the following lemma.

Lemma C.1. *For any $C' > 0$, there exists a constant $C > 0$ only depending on C' such that*

$$\max_{\{Z:\ell(Z,Z^*)\leq\tau\}} \quad \sum_{j=1}^{p} \max_{U\neq Z_j^*} \frac{F_j(Z_j^*, U; Z)^2 \left\| \mu_j(B^*, U) - \mu_j(B^*, Z_j^*) \right\|^2}{\Delta_j(Z_j^*, U)^4 \ell(Z, Z^*)} \leq C\left(\frac{d}{p\lambda^{*2}} + \frac{1}{p^2\lambda^{*4}}\right), \tag{C.3}$$

$$\max_{\{Z:\ell(Z,Z^*)\leq\tau\}} \quad \frac{\tau}{4\Delta_{\min}^2 |T| + \tau} \sum_{j\in T} \max_{U\neq Z_j} \frac{G_j(Z_j^*, U; Z)^2 \left\| \mu_j(B^*, U) - \mu_j(B^*, Z_j^*) \right\|^2}{\Delta_j(Z_j^*, U)^4 \ell(Z, Z^*)} \leq C\frac{\tau\left(\lambda^{*2} + \frac{1}{pd}\right)}{(p\lambda^{*2})^2}, \tag{C.4}$$

and

$$\max_{j\in[p]} \max_{U\neq Z_j^*} \frac{\left|H_j(Z_j^*, U)\right|}{\Delta_j(Z_j^*, U)^2} \leq C\sqrt{\frac{d}{p\lambda^{*2}}}, \tag{C.5}$$

with probability at least $1 - e^{-C'pd}$.

From the bounds (C.3)-(C.5), we can see that a sufficient condition under which Conditions A, B and C hold is $\tau = o(p^2\lambda^{*2})$ and $\frac{p\lambda^{*2}}{d} \to \infty$.

Next, we need to bound ξ_{ideal} in Condition D. This is given by the following lemma.

Lemma C.2. *Assume $\frac{p\lambda^{*2}}{d\log d} \to \infty$. Then, for any sequence $\delta_p = o(1)$, we have*

$$\xi_{\text{ideal}}(\delta_p) \leq p\exp\left(-\frac{1+o(1)}{2}p\lambda^{*2}\right),$$

with probability at least $1 - \exp(-\sqrt{p\lambda^{*2}})$.

Thus, under the condition $\frac{p\lambda^{*2}}{d\log d} \to \infty$, Conditions A, B, C and D hold simultaneously.

C.2 Convergence

With the help of Lemma C.1 and Lemma C.2, we can specialize Theorem 3.1 into the following result.

Theorem C.2. *Assume $\frac{p\lambda^{*2}}{d\log d} \to \infty$. Suppose $Z^{(0)}$ satisfies*

$$\ell(Z^{(0)}, Z^*) = o(p^2\lambda^{*2}), \tag{C.6}$$

with probability at least $1 - \eta$. Then, we have

$$\ell(Z^{(t)}, Z^*) \leq p\exp\left(-(1+o(1))\frac{p\lambda^{*2}}{2}\right) + \frac{1}{2}\ell(Z^{(t-1)}, Z^*) \quad \text{for all } t \geq 1,$$

with probability at least $1 - \eta - \exp(-\sqrt{p\lambda^{*2}}) - e^{-pd}$.

According to the definition of the loss $\ell(Z, Z^*)$, we have the inequality

$$\ell(Z, Z^*) \geq 4p^2\lambda^{*2}\frac{1}{p}\sum_{j=1}^{p}\mathbf{1}_{\{Z_j \neq Z_j^*\}}.$$

This immediately implies the following corollary for the Hamming loss.

Corollary C.1. *Assume* $\frac{p\lambda^{*2}}{d\log d} \to \infty$. *Suppose* $z^{(0)}$ *satisfies (C.6) with probability at least* $1-\eta$. *Then, we have*

$$\frac{1}{p}\sum_{j=1}^{p}\mathbf{1}_{\left\{Z_j^{(t)} \neq Z_j^*\right\}} \leq \exp\left(-(1+o(1))\frac{p\lambda^{*2}}{2}\right) + 2^{-t} \quad \text{for all } t \geq 1, \tag{C.7}$$

with probability at least $1 - \eta - \exp(-\sqrt{p\lambda^{*2}}) - e^{-pd}$.

By the property of the Hamming loss, the algorithmic error 2^{-t} is negligible after $\lceil 3\log p\rceil$ iterations, and we have

$$\frac{1}{p}\sum_{j=1}^{p}\mathbf{1}_{\left\{Z_j^{(t)} \neq Z_j^*\right\}} \leq \exp\left(-(1+o(1))\frac{p\lambda^{*2}}{2}\right) \quad \text{for all } t \geq 3\log p.$$

Thus, the minimax rate is achieved given that the initialization condition (C.6) is satisfied.

C.3 Initialization

Since $Y = \lambda^* Z^* Z^{*T} + W$, the matrix $\mathbb{E}Y \in \mathbb{R}^{pd\times pd}$ has a low rank structure, and thus we can recover the information of Z^* via eigenvalue decomposition. Let $\widehat{U} \in \mathbb{R}^{pd\times d}$ collect the leading eigenvectors of Y. We use the notation $\widehat{U}_j \in \mathbb{R}^{d\times d}$ for the jth block of $\widehat{U}$ so that $\widehat{U}^T = (\widehat{U}_1, \cdots, \widehat{U}_p)$. For each $j \in [p]$, find $Z_j^{(0)} = \text{argmin}_{V\in\mathcal{P}_d}\|\sqrt{p}\widehat{U}_j - V\|_{\text{F}}^2$. Again, this optimization is equivalent to

$$Z_j^{(0)} = \underset{V\in\mathcal{P}_d}{\text{argmax}}\left\langle \widehat{U}_j, V\right\rangle,$$

and can be solved by the Kuhn-Munkres algorithm [6] with $O(d^3)$ complexity. The statistical guarantee of $Z^{(0)}$ is given by the following proposition.

Proposition C.1. *For any constant* $C' > 0$, *there exists a constant* $C > 0$ *only depending on* C *such that*

$$\min_{U\in\mathcal{P}_d}\ell(Z^{(0)}, Z^*U^T) \leq Cpd^2,$$

with probability at least $1 - e^{-C'pd}$.

Proposition C.1 show that $Z^{(0)}$ achieves the rate $O(pd^2)$ for estimating Z^* up to a global permutation. This uncertainty cannot be avoided since $Z_iZ_j^T = Z_iU^TUZ_j^T$. In order that the initialization condition (C.6) is satisfied, we thus require $\frac{p\lambda^{*2}}{d^2} \to \infty$, which implies the condition for algorithmic convergence $\frac{p\lambda^{*2}}{d\log d} \to \infty$. By Corollary C.1, we can thus conclude that the iterative algorithm initialized by the spectral method converges to the minimax error with a linear rate under the condition $\frac{p\lambda^{*2}}{d^2} \to \infty$.

D Proof of Theorem 3.1

Suppose $\ell(z^{(t-1)}, z^*) \leq \tau$, and we will show $\ell(z^{(t)}, z^*) \leq 2\xi_{\text{ideal}}(\delta) + \frac{1}{2}\ell(z^{(t-1)}, z^*)$. By the definition of the loss (21), we have

$$\begin{aligned}
\ell(z^{(t)}, z^*) &= \sum_{j=1}^{p} \|\mu_j(B^*, z_j^{(t)}) - \mu_j(B^*, z_j^*)\|^2 \\
&= \sum_{j=1}^{p} \sum_{b\in[k]\setminus\{z_j^*\}} \|\mu_j(B^*, b) - \mu_j(B^*, z_j^*)\|^2 \mathbf{1}_{\left\{z_j^{(t)}=b\right\}}. \qquad \text{(D.1)}
\end{aligned}$$

To bound (D.1), we have

$$\begin{aligned}
& \mathbf{1}_{\left\{z_j^{(t)}=b\right\}} \\
\leq\ & \mathbf{1}_{\left\{\|T_j-\nu_j(\widehat{B}(z^{(t-1)}),b)\|^2\leq\|T_j-\nu_j(\widehat{B}(z^{(t-1)}),z_j^*)\|^2\right\}} \qquad \text{(D.2)} \\
=\ & \mathbf{1}_{\left\{\left\langle\epsilon_j,\nu_j(\widehat{B}(z^*),z_j^*)-\nu_j(\widehat{B}(z^*),b)\right\rangle\leq-\frac{1}{2}\Delta_j(z_j^*,b)^2+F_j(z_j^*,b;z^{(t-1)})+G_j(z_j^*,b;z^{(t-1)})+H_j(z_j^*,b)\right\}} \qquad \text{(D.3)} \\
\leq\ & \mathbf{1}_{\left\{\left\langle\epsilon_j,\nu_j(\widehat{B}(z^*),z_j^*)-\nu_j(\widehat{B}(z^*),b)\right\rangle\leq-\frac{1-\delta}{2}\Delta_j(z_j^*,b)^2\right\}} \qquad \text{(D.4)} \\
& +\mathbf{1}_{\left\{\frac{\delta}{2}\Delta_j(z_j^*,b)^2\leq F_j(z_j^*,b;z^{(t-1)})+G_j(z_j^*,b;z^{(t-1)})+H_j(z_j^*,b)\right\}} \\
\leq\ & \mathbf{1}_{\left\{\left\langle\epsilon_j,\nu_j(\widehat{B}(z^*),z_j^*)-\nu_j(\widehat{B}(z^*),b)\right\rangle\leq-\frac{1-\delta}{2}\Delta_j(z_j^*,b)^2\right\}} \qquad \text{(D.5)} \\
& +\mathbf{1}_{\left\{\frac{\delta}{4}\Delta_j(z_j^*,b)^2\leq F_j(z_j^*,b;z^{(t-1)})+G_j(z_j^*,b;z^{(t-1)})\right\}} \\
\leq\ & \mathbf{1}_{\left\{\left\langle\epsilon_j,\nu_j(\widehat{B}(z^*),z_j^*)-\nu_j(\widehat{B}(z^*),b)\right\rangle\leq-\frac{1-\delta}{2}\Delta_j(z_j^*,b)^2\right\}} \qquad \text{(D.6)} \\
& +\frac{32F_j(z_j^*,b;z^{(t-1)})^2}{\delta^2\Delta_j(z_j^*,b)^4}+\frac{32G_j(z_j^*,b;z^{(t-1)})^2}{\delta^2\Delta_j(z_j^*,b)^4}.
\end{aligned}$$

The inequality (D.2) is due to the definition that $z_j^{(t)} = \text{argmin}_{a\in[k]} \|T_j - \nu_j(\widehat{B}(z^{(t-1)}), a)\|^2$. Then, the equality (D.3) uses the equivalence between (24) and (25). The inequality (D.4) uses a union bound, and (D.5) applies Condition C. Finally, (D.6) follows Markov's inequality.

Apply the bound (D.6) to (D.1), and then $\ell(z^{(t)}, z^*)$ can be bounded by

$$
\begin{aligned}
& \sum_{j=1}^{p} \sum_{b\in[k]\setminus\{z_j^*\}} \|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2 \mathbf{1}_{\left\{\langle \epsilon_j, \nu_j(\widehat{B}(z^*),z_j^*)-\nu_j(\widehat{B}(z^*),b)\rangle \leq -\frac{1-\delta}{2}\Delta_j(z_j^*,b)^2\right\}} \\
& + \sum_{j=1}^{p} \sum_{b\in[k]\setminus\{z_j^*\}} \|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2 \mathbf{1}_{\left\{z_j^{(t)}=b\right\}} \frac{32F_j(z_j^*,b;z^{(t-1)})^2}{\delta^2\Delta_j(z_j^*,b)^4} \\
& + \sum_{j=1}^{p} \sum_{b\in[k]\setminus\{z_j^*\}} \|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2 \mathbf{1}_{\left\{z_j^{(t)}=b\right\}} \frac{32G_j(z_j^*,b;z^{(t-1)})^2}{\delta^2\Delta_j(z_j^*,b)^4} \\
\leq\; & \xi_{\text{ideal}}(\delta) + \sum_{j=1}^{p} \max_{b\in[k]\setminus\{z_j^*\}} \|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2 \frac{32F_j(z_j^*,b;z^{(t-1)})^2}{\delta^2\Delta_j(z_j^*,b)^4} \\
& + \sum_{j=1}^{p} \mathbf{1}_{\left\{z_j^{(t)}\neq z_j^*\right\}} \max_{b\in[k]\setminus\{z_j^*\}} \|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2 \frac{32G_j(z_j^*,b;z^{(t-1)})^2}{\delta^2\Delta_j(z_j^*,b)^4} && \text{(D.7)} \\
\leq\; & \xi_{\text{ideal}}(\delta) + \frac{1}{8}\ell(z^{(t-1)},z^*) + \frac{4\Delta_{\min}^2 h(z^{(t)},z^*)+\tau}{8\tau}\ell(z^{(t-1)},z^*) && \text{(D.8)} \\
\leq\; & \xi_{\text{ideal}}(\delta) + \frac{1}{2}\Delta_{\min}^2 h(z^{(t)},z^*) + \frac{1}{4}\ell(z^{(t-1)},z^*) && \text{(D.9)} \\
\leq\; & \xi_{\text{ideal}}(\delta) + \frac{1}{2}\ell(z^{(t)},z^*) + \frac{1}{4}\ell(z^{(t-1)},z^*), && \text{(D.10)}
\end{aligned}
$$

where we have used Conditions A and B in (D.8). The inequality (D.9) uses the condition $\ell(z^{(t-1)}, z^*) \leq \tau$, and (D.10) is by (22). To summarize, we have obtained

$$
\ell(z^{(t)}, z^*) \leq \xi_{\text{ideal}}(\delta) + \frac{1}{2}\ell(z^{(t)}, z^*) + \frac{1}{4}\ell(z^{(t-1)}, z^*),
$$

which can be rearranged into

$$
\ell(z^{(t)}, z^*) \leq 2\xi_{\text{ideal}}(\delta) + \frac{1}{2}\ell(z^{(t-1)}, z^*).
$$

To prove the conclusion of Theorem 3.1, we use a mathematical induction argument. First, Condition D asserts that $\ell(z^{(0)}, z^*) \leq \tau$. This leads to $\ell(z^{(1)}, z^*) \leq 2\xi_{\text{ideal}}(\delta)+\frac{1}{2}\ell(z^{(0)}, z^*) \leq \tau$, together with Condition C that $\xi_{\text{ideal}}(\delta) \leq \frac{1}{4}\tau$. Suppose $\ell(z^{(t-1)}, z^*) \leq \tau$, we then have $\ell(z^{(t)}, z^*) \leq 2\xi_{\text{ideal}}(\delta)+\frac{1}{2}\ell(z^{(t-1)}, z^*) \leq \tau$. Hence, $\ell(z^{(t-1)}, z^*) \leq \tau$ for all $t \geq 1$, which implies that $\ell(z^{(t)}, z^*) \leq 2\xi_{\text{ideal}}(\delta) + \frac{1}{2}\ell(z^{(t-1)}, z^*)$ for all $t \geq 1$, and the proof is complete.

E Proofs in Section 4

In this section, we present the proofs of Lemma 4.1, Lemma 4.2 and Proposition 4.1. The conclusions of Theorem 4.1 and Corollary 4.1 are direct consequences of Theorem 3.1, and thus their proofs are omitted. We first list some technical lemmas. The following χ^2 tail probability is Lemma 1 of [7].

Lemma E.1. *For any $x > 0$, we have*

$$\begin{aligned}\mathbb{P}\left(\chi_d^2 \geq d + 2\sqrt{dx} + 2x\right) &\leq e^{-x}, \\ \mathbb{P}\left(\chi_d^2 \leq d - 2\sqrt{dx}\right) &\leq e^{-x}.\end{aligned}$$

Lemma E.2. *Consider i.i.d. random vectors $\epsilon_1, \ldots, \epsilon_p \sim \mathcal{N}(0, I_d)$ and some $z^* \in [k]^p$ and $k \in [p]$. Then, for any constant $C' > 0$, there exists some constant $C > 0$ only depending on C' such that*

$$\max_{a\in[k]} \left\| \frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sqrt{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}}} \right\| \leq C\sqrt{d + \log p}, \tag{E.1}$$

$$\max_{T\subset[p]} \left\| \frac{1}{\sqrt{|T|}} \sum_{j\in T} \epsilon_j \right\| \leq C\sqrt{d + p}, \tag{E.2}$$

$$\max_{a\in[k]} \frac{1}{d + \sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}} \left\| \sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j\epsilon_j^T \right\| \leq C, \tag{E.3}$$

with probability at least $1 - p^{-C'}$. We have used the convention that $0/0 = 0$.

Proof. By Lemma E.1, we have $\mathbb{P}(\chi_d^2 \geq d + 2\sqrt{xd} + 2x) \leq e^{-x}$. Then, a union bound argument leads to (E.1). The inequalities (E.2) and (E.3) are Lemmas A.1 and A.2 in [10]. We need to slightly extend Lemma A.2 in [10], but this can be done by a standard union bound argument. □

With the two lemmas above, we are ready to state the proofs of Lemma 4.1 and Lemma 4.2.

Proof of Lemma 4.1. We write $\epsilon_j = Y_j - \theta_{z_j^*}$ and consider the event that the three inequalities (E.1)-(E.3) hold. For any $z \in [k]^p$ such that $\ell(z, z^*) \leq \tau \leq \frac{\Delta_{\min}^2 \alpha p}{2k}$, we have

$$\begin{aligned}\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}} &\geq \sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}} - \sum_{j=1}^p \mathbf{1}_{\{z_j\neq z_j^*\}} \\ &\geq \sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}} - \frac{\ell(z, z^*)}{\Delta_{\min}^2} \\ &\geq \frac{\alpha p}{k} - \frac{\alpha p}{2k} \\ &= \frac{\alpha p}{2k},\end{aligned}$$

which implies

$$\min_{a\in[k]} \sum_{j=1}^p \mathbf{1}_{\{z_j=a\}} \geq \frac{\alpha p}{2k}. \tag{E.4}$$

We then introduce more notation. We write $\theta_a(z) = \mathbb{E}\widehat{\theta}_a(z)$ and

$$\bar{\epsilon}_a(z) = \frac{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}}.$$

We first derive bounds for $\max_{a\in[k]}\|\widehat{\theta}_a(z^*)-\theta_a^*\|$, $\max_{a\in[k]}\|\theta_a(z)-\theta_a(z^*)\|$ and $\max_{a\in[k]}\|\bar{\epsilon}_a(z)-\bar{\epsilon}_a(z^*)\|$. By (E.1) and (E.4), we have

$$\begin{aligned}\max_{a\in[k]}\|\widehat{\theta}_a(z^*)-\theta_a^*\| &= \max_{a\in[k]}\left\|\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}}\right\| \\ &\leq \sqrt{\frac{k}{\alpha p}}\max_{a\in[k]}\left\|\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sqrt{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}}}\right\| \\ &\lesssim \sqrt{\frac{k(d+\log p)}{p}}. \end{aligned} \tag{E.5}$$

By (E.4), we have

$$\begin{aligned}\max_{a\in[k]}\|\theta_a(z)-\theta_a(z^*)\| &= \left\|\frac{1}{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}}\sum_{j=1}^p\sum_{b\in[k]\backslash\{a\}}\mathbf{1}_{\{z_j=a,z_j^*=b\}}(\theta_b^*-\theta_a^*)\right\| \\ &\leq \frac{2k}{\alpha p}\sum_{j=1}^p\sum_{b\in[k]\backslash\{a\}}\|\theta_b^*-\theta_a^*\|\mathbf{1}_{\{z_j=a,z_j^*=b\}} \\ &\leq \frac{2k}{\alpha p\Delta_{\min}}\ell(z,z^*). \end{aligned} \tag{E.6}$$

By (E.4), we have

$$\begin{aligned}&\max_{a\in[k]}\|\bar{\epsilon}_a(z)-\bar{\epsilon}_a(z^*)\| \\ &= \max_{a\in[k]}\left\|\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}}-\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}}\right\| \\ &\leq \max_{a\in[k]}\left\|\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}}-\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}}\right\|+\max_{a\in[k]}\left\|\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j=a\}}}-\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}}\right\| \\ &\leq \frac{2k}{\alpha p}\max_{a\in[k]}\left\|\sum_{j=1}^p(\mathbf{1}_{\{z_j=a\}}-\mathbf{1}_{\{z_j^*=a\}})\epsilon_j\right\| \\ &\quad+\frac{2k}{\alpha p}\sqrt{\frac{k}{\alpha p}}\left|\sum_{j=1}^p\mathbf{1}_{\{z_j=a\}}-\sum_{j=1}^p\mathbf{1}_{\{z_j^*=a\}}\right|\max_{a\in[k]}\left\|\frac{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sqrt{\sum_{j=1}^p \mathbf{1}_{\{z_j^*=a\}}}}\right\|, \end{aligned}$$

where the first term in the above bound can be bounded by

$$\begin{aligned}
&\frac{2k}{\alpha p}\max_{a\in[k]}\left\|\sum_{j=1}^{p}\mathbf{1}_{\{z_j=a,z_j^*\neq a\}}\epsilon_j\right\|+\frac{2k}{\alpha p}\max_{a\in[k]}\left\|\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a,z_j\neq a\}}\epsilon_j\right\| \\
\lesssim\ &\frac{k\sqrt{d+p}}{p}\sqrt{\frac{\ell(z,z^*)}{\Delta_{\min}^2}},
\end{aligned}$$

because of the facts that $\max_{a\in[k]}\sum_{j=1}^{p}\mathbf{1}_{\{z_j=a,z_j^*\neq a\}}\leq\frac{\ell(z,z^*)}{\Delta_{\min}^2}$, $\max_{a\in[k]}\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a,z_j\neq a\}}\leq\frac{\ell(z,z^*)}{\Delta_{\min}^2}$, and the inequality (E.2), and the second term can be bounded by

$$\begin{aligned}
&\frac{2k}{\alpha p}\sqrt{\frac{k}{\alpha p}}\max_{a\in[k]}\left\|\frac{\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sqrt{\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}}}}\right\|\left(\max_{a\in[k]}\sum_{j=1}^{p}\mathbf{1}_{\{z_j=a,z_j^*\neq a\}}+\max_{a\in[k]}\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a,z_j\neq a\}}\right) \\
\leq\ &\frac{4k}{\alpha p}\sqrt{\frac{k}{\alpha p}}\frac{\ell(z,z^*)}{\Delta_{\min}^2}\max_{a\in[k]}\left\|\frac{\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}}\epsilon_j}{\sqrt{\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}}}}\right\| \\
\lesssim\ &\frac{k\sqrt{k}\ell(z,z^*)\sqrt{d+\log p}}{p\sqrt{p}\Delta_{\min}^2}.
\end{aligned}$$

Under the condition that $\ell(z,z^*)\leq\tau\leq\frac{\Delta_{\min}^2\alpha p}{2k}$, we have

$$\begin{aligned}
&\max_{a\in[k]}\|\bar{\epsilon}_a(z)-\bar{\epsilon}_a(z^*)\| \\
\lesssim\ &\frac{k\sqrt{d+p}}{p}\sqrt{\frac{\ell(z,z^*)}{\Delta_{\min}^2}}+\frac{k\sqrt{k}\ell(z,z^*)\sqrt{d+\log p}}{p\sqrt{p}\Delta_{\min}^2} \\
\lesssim\ &\frac{k\sqrt{d+p}}{p}\sqrt{\frac{\ell(z,z^*)}{\Delta_{\min}^2}}. \qquad \text{(E.7)}
\end{aligned}$$

Combining the two bounds (E.6) and (E.7) and using triangle inequality, we also have

$$\begin{aligned}
&\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\| \\
\leq\ &\max_{a\in[k]}\|\theta_a(z)-\theta_a(z^*)\|+\max_{a\in[k]}\|\bar{\epsilon}_a(z)-\bar{\epsilon}_a(z^*)\| \\
\lesssim\ &\frac{k}{p\Delta_{\min}}\ell(z,z^*)+\frac{k\sqrt{d+p}}{p\Delta_{\min}}\sqrt{\ell(z,z^*)}. \qquad \text{(E.8)}
\end{aligned}$$

Now we proceed to prove (31)-(33). For (31), we have

$$
\begin{aligned}
& \sum_{j=1}^{p} \max_{b\in[k]\backslash\{z_j^*\}} \frac{F_j(z_j^*,b;z)^2\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)} \\
\leq\ & \sum_{j=1}^{p}\sum_{b=1}^{k} \frac{\left|\left\langle \epsilon_j, \widehat{\theta}_{z_j^*}(z^*)-\widehat{\theta}_{z_j^*}(z)-\widehat{\theta}_b(z^*)+\widehat{\theta}_b(z)\right\rangle\right|^2}{\|\theta^*_{z_j^*}-\theta^*_b\|^2\ell(z,z^*)} \\
\leq\ & \sum_{b=1}^{k}\sum_{a\in[k]\backslash\{b\}}\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}} \frac{\left|\left\langle \epsilon_j, \widehat{\theta}_a(z^*)-\widehat{\theta}_a(z)-\widehat{\theta}_b(z^*)+\widehat{\theta}_b(z)\right\rangle\right|^2}{\|\theta^*_a-\theta^*_b\|^2\ell(z,z^*)} \\
\leq\ & \sum_{b=1}^{k}\sum_{a\in[k]\backslash\{b\}} \frac{\left\|\widehat{\theta}_a(z^*)-\widehat{\theta}_a(z)-\widehat{\theta}_b(z^*)+\widehat{\theta}_b(z)\right\|^2}{\|\theta^*_a-\theta^*_b\|^2\ell(z,z^*)} \left\|\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}}\epsilon_j\epsilon_j^T\right\| \\
\lesssim\ & \frac{k^2(kd/p+1)}{\Delta_{\min}^2}\left(1+\frac{k(d/p+1)}{\Delta_{\min}^2}\right)
\end{aligned}
$$

where we have used (E.3), (E.8) and the condition that $\ell(z,z^*)\leq\tau\leq\frac{\Delta_{\min}^2\alpha p}{2k}$. Next, for (32), we have

$$
\begin{aligned}
|G_j(a,b;z)| \leq\ & \frac{1}{2}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^2+\frac{1}{2}\|\widehat{\theta}_b(z)-\widehat{\theta}_b(z^*)\|^2 \\
& +\|\widehat{\theta}_a(z^*)-\theta^*_a\|\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|+\|\widehat{\theta}_a(z^*)-\theta^*_b\|\|\widehat{\theta}_b(z)-\widehat{\theta}_b(z^*)\| \\
\leq\ & \max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^2+2\left(\max_{a\in[k]}\|\widehat{\theta}_a(z^*)-\theta^*_a\|\right)\left(\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|\right) \\
& +\|\theta^*_a-\theta^*_b\|\left(\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|\right).
\end{aligned}
$$

This implies for any subset $T \subset [p]$, we have

$$
\begin{aligned}
& \frac{\tau}{4\Delta_{\min}^2|T|+\tau}\sum_{j\in T}\max_{b\in[k]\backslash\{z_j^*\}}\frac{G_j(z_j^*,b;z)^2\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)} \\
\leq\ & \frac{\tau}{4\Delta_{\min}^2|T|}\sum_{j\in T}\frac{3\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^4}{\Delta_{\min}^2\ell(z,z^*)} \\
& +\frac{\tau}{4\Delta_{\min}^2|T|}\sum_{j\in T}\frac{12\left(\max_{a\in[k]}\|\widehat{\theta}_a(z^*)-\theta_a^*\|^2\right)\left(\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^2\right)}{\Delta_{\min}^2\ell(z,z^*)} \\
& +\frac{\tau}{4\Delta_{\min}^2|T|}\sum_{j\in T}\frac{3\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^2}{\ell(z,z^*)} \\
=\ & \frac{3\tau\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^4}{4\Delta_{\min}^4\ell(z,z^*)}+\frac{3\tau\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^2}{4\Delta_{\min}^2\ell(z,z^*)} \\
& +\frac{3\tau\left(\max_{a\in[k]}\|\widehat{\theta}_a(z^*)-\theta_a^*\|^2\right)\left(\max_{a\in[k]}\|\widehat{\theta}_a(z)-\widehat{\theta}_a(z^*)\|^2\right)}{\Delta_{\min}^4\ell(z,z^*)} \\
\lesssim\ & \frac{k\tau}{p\Delta_{\min}^2}+\frac{k(d+p)}{p\Delta_{\min}^2}+\frac{k^2(d+p)^2}{p^2\Delta_{\min}^4},
\end{aligned}
$$

where we have used (E.5), (E.8), and the condition that $\ell(z,z^*)\leq\tau\leq\frac{\Delta_{\min}^2\alpha p}{2k}$. Finally, for (33), the bound (E.5) leads to

$$
\begin{aligned}
\frac{|H_j(a,b)|}{\Delta_j(a,b)^2} &\leq \frac{\frac{1}{2}\|\widehat{\theta}_a(z^*)-\theta_a^*\|^2+\frac{1}{2}\|\widehat{\theta}_b(z^*)-\theta_b^*\|^2+\|\theta_a^*-\theta_b^*\|\|\widehat{\theta}_b(z^*)-\theta_b^*\|}{\|\theta_a^*-\theta_b^*\|^2} \\
&\lesssim \frac{k(d+\log p)}{p\Delta_{\min}^2}+\sqrt{\frac{k(d+\log p)}{p\Delta_{\min}^2}}.
\end{aligned}
$$

By taking maximum, we have obtained (31)-(33). The proof is complete. □

Proof of Lemma 4.2. Note that

$$
\begin{aligned}
& \mathbb{P}\left(\left\langle\epsilon_j,\widehat{\theta}_a(z^*)-\widehat{\theta}_b(z^*)\right\rangle\leq-\frac{1-\delta}{2}\|\theta_a^*-\theta_b^*\|^2\right) \\
\leq\ & \mathbb{P}\left(\langle\epsilon_j,\theta_a^*-\theta_b^*\rangle\leq-\frac{1-\delta-\bar{\delta}}{2}\|\theta_a^*-\theta_b^*\|^2\right) \\
& +\mathbb{P}\left(\left\langle\epsilon_j,\widehat{\theta}_a(z^*)-\theta_a^*\right\rangle\leq-\frac{\bar{\delta}}{4}\|\theta_a^*-\theta_b^*\|^2\right) \\
& +\mathbb{P}\left(-\left\langle\epsilon_j,\widehat{\theta}_b(z^*)-\theta_b^*\right\rangle\leq-\frac{\bar{\delta}}{4}\|\theta_a^*-\theta_b^*\|^2\right),
\end{aligned}
$$

where $\bar{\delta}=\bar{\delta}_p$ is some sequence to be chosen later, and we need to bound the three terms on the right hand side of the above inequality respectively. For the first term, a standard

Gaussian tail bound gives

$$\mathbb{P}\left(\langle \epsilon_j, \theta_a^* - \theta_b^* \rangle \leq -\frac{1-\delta-\bar{\delta}}{2}\|\theta_a^* - \theta_b^*\|^2\right) \leq \exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|\theta_a^* - \theta_b^*\|^2\right).$$

To bound the second term, we note that

$$\left\langle \epsilon_j, \widehat{\theta}_a(z^*) - \theta_a^* \right\rangle = \frac{\mathbf{1}_{\{z_j^*=a\}}\|\epsilon_j\|^2}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=a\}}} + \frac{\sum_{l\in[p]\setminus\{j\}} \mathbf{1}_{\{z_l^*=a\}}\epsilon_j^T\epsilon_l}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=a\}}} \geq \frac{\sum_{l\in[p]\setminus\{j\}} \mathbf{1}_{\{z_l^*=a\}}\epsilon_j^T\epsilon_l}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=a\}}}.$$

This implies

$$\begin{aligned}
& \mathbb{P}\left(\left\langle \epsilon_j, \widehat{\theta}_a(z^*) - \theta_a^* \right\rangle \leq -\frac{\bar{\delta}}{4}\|\theta_a^* - \theta_b^*\|^2\right) \\
\leq\ & \mathbb{P}\left(\frac{\sum_{l\in[p]\setminus\{j\}} \mathbf{1}_{\{z_l^*=a\}}\epsilon_j^T\epsilon_l}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=a\}}} \leq -\frac{\bar{\delta}}{4}\|\theta_a^* - \theta_b^*\|^2\right) \\
\leq\ & \mathbb{P}\left(\frac{\sum_{l\in[p]\setminus\{j\}} \mathbf{1}_{\{z_l^*=a\}}\epsilon_j^T\epsilon_l}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=a\}}} \leq -\frac{\bar{\delta}}{4}\|\theta_a^* - \theta_b^*\|^2 \middle| \|\epsilon_l\|^2 < d + 2\sqrt{xd} + 2x\right) \\
& +\mathbb{P}\left(\|\epsilon_l\|^2 > d + 2\sqrt{xd} + 2x\right) \\
\leq\ & \mathbb{E}\left(\exp\left(-\frac{\bar{\delta}^2\|\theta_a^* - \theta_b^*\|^4 \sum_{l=1}^p \mathbf{1}_{\{z_l^*=a\}}}{32\|\epsilon_l\|^2}\right) \middle| \|\epsilon_l\|^2 < d + 2\sqrt{xd} + 2x\right) \\
& +\mathbb{P}\left(\|\epsilon_l\|^2 > d + 2\sqrt{xd} + 2x\right) \\
\leq\ & \exp\left(-\frac{\bar{\delta}^2\|\theta_a^* - \theta_b^*\|^4\alpha p}{32k\left(d + 2\sqrt{xd} + 2x\right)}\right) + \exp(-x).
\end{aligned}$$

Choosing $x = \bar{\delta}\|\theta_a^* - \theta_b^*\|^2\sqrt{\alpha p/k}$, we have

$$\begin{aligned}
& \mathbb{P}\left(\left\langle \epsilon_j, \widehat{\theta}_a(z^*) - \theta_a^* \right\rangle \leq -\frac{\bar{\delta}}{4}\|\theta_a^* - \theta_b^*\|^2\right) \\
\leq\ & \exp\left(-C\frac{\bar{\delta}^2\|\theta_a^* - \theta_b^*\|^4 p}{kd}\right) + \exp\left(-C\frac{\bar{\delta}\|\theta_a^* - \theta_b^*\|^2\sqrt{p}}{\sqrt{k}}\right).
\end{aligned} \tag{E.9}$$

To bound the third term, we note that

$$-\left\langle \epsilon_j, \widehat{\theta}_b(z^*) - \theta_b^* \right\rangle = -\frac{\mathbf{1}_{\{z_j^*=b\}}\|\epsilon_j\|^2}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=b\}}} - \frac{\sum_{l\in[p]\setminus\{j\}} \mathbf{1}_{\{z_l^*=b\}}\epsilon_j^T\epsilon_l}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=b\}}},$$

and we then have

$$\begin{aligned}
&\mathbb{P}\left(-\left\langle \epsilon_j, \widehat{\theta}_b(z^*) - \theta_b^* \right\rangle \leq -\frac{\bar{\delta}}{4}\|\theta_a^* - \theta_b^*\|^2\right) \\
\leq\ & \mathbb{P}\left(-\frac{\mathbf{1}_{\{z_j^*=b\}}\|\epsilon_j\|^2}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=b\}}} \leq -\frac{\bar{\delta}}{8}\|\theta_a^* - \theta_b^*\|^2\right) \\
&+\mathbb{P}\left(-\frac{\sum_{l\in[p]\backslash\{j\}} \mathbf{1}_{\{z_l^*=b\}}\epsilon_j^T\epsilon_l}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=b\}}} \leq -\frac{\bar{\delta}}{8}\|\theta_a^* - \theta_b^*\|^2\right).
\end{aligned}$$

The second term on the right hand side of the above inequality can be bounded in the same way as (E.9). For the first term, we have

$$\begin{aligned}
&\mathbb{P}\left(-\frac{\mathbf{1}_{\{z_j^*=b\}}\|\epsilon_j\|^2}{\sum_{l=1}^p \mathbf{1}_{\{z_l^*=b\}}} \leq -\frac{\bar{\delta}}{8}\|\theta_a^* - \theta_b^*\|^2\right) \\
\leq\ & \mathbb{P}\left(\|\epsilon_j\|^2 > \frac{\bar{\delta}}{8}\|\theta_a^* - \theta_b^*\|^2\frac{\alpha p}{k}\right) \\
\leq\ & \exp\left(-C\bar{\delta}\|\theta_a^* - \theta_b^*\|^2\frac{p}{k}\right),
\end{aligned}$$

under the condition $\frac{\Delta_{\min}^2}{\log k + kd/p} \to \infty$. Combining the bounds above, we have

$$\begin{aligned}
&\mathbb{P}\left(\left\langle \epsilon_j, \widehat{\theta}_a(z^*) - \widehat{\theta}_b(z^*) \right\rangle \leq -\frac{1-\delta}{2}\|\theta_a^* - \theta_b^*\|^2\right) \\
\leq\ & \exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|\theta_a^* - \theta_b^*\|^2\right) + \exp\left(-C\bar{\delta}\|\theta_a^* - \theta_b^*\|^2\frac{p}{k}\right) \\
&+2\exp\left(-C\frac{\bar{\delta}^2\|\theta_a^* - \theta_b^*\|^4 p}{kd}\right) + 2\exp\left(-C\frac{\bar{\delta}\|\theta_a^* - \theta_b^*\|^2\sqrt{p}}{\sqrt{k}}\right) \\
\leq\ & 6\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|\theta_a^* - \theta_b^*\|^2\right),
\end{aligned}$$

where the last inequality above is obtained under the condition that $\frac{\Delta_{\min}^2}{\log k + kd/p} \to \infty$ and $p/k \to \infty$, so that we can choose some $\bar{\delta} = \bar{\delta}_p = o(1)$ that is slowly diverging to zero.

Now we are ready to bound $\xi_{\text{ideal}}(\delta)$. We first bound its expectation. We have

$$\begin{aligned}
\mathbb{E}\xi_{\text{ideal}}(\delta) &= \sum_{j=1}^p \sum_{b\in[k]\backslash\{z_j^*\}} \|\theta_b^* - \theta_{z_j^*}^*\|^2 \mathbb{P}\left(\left\langle \epsilon_j, \widehat{\theta}_{z_j^*}(z^*) - \widehat{\theta}_b(z^*) \right\rangle \leq -\frac{1-\delta}{2}\|\theta_{z_j^*}^* - \theta_b^*\|^2\right) \\
&\leq 6\sum_{j=1}^p \sum_{b\in[k]\backslash\{z_j^*\}} \|\theta_b^* - \theta_{z_j^*}^*\|^2 \exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|\theta_{z_j^*}^* - \theta_b^*\|^2\right).
\end{aligned}$$

With $\delta = \delta_p = o(1)$, we then have

$$\mathbb{E}\xi_{\text{ideal}}(\delta_p) \leq \sum_{j=1}^p \sum_{b\in[k]\backslash\{z_j^*\}} \exp\left(-(1+o(1))\frac{\|\theta_{z_j^*}^* - \theta_b^*\|^2}{8}\right) \leq p\exp\left(-(1+o(1))\frac{\Delta_{\min}^2}{8}\right),$$

under the condition that $\frac{\Delta_{\min}^2}{\log k + kd/p} \to \infty$. Finally, by Markov's inequality, we have

$$\mathbb{P}\left(\xi_{\text{ideal}}(\delta_p) > \mathbb{E}\xi_{\text{ideal}}(\delta_p)\exp\left(\Delta_{\min}\right)\right) \leq \exp\left(-\Delta_{\min}\right).$$

In other words, with probability at least $1 - \exp\left(-\Delta_{\min}\right)$, we have

$$\xi_{\text{ideal}}(\delta_p) \leq \mathbb{E}\xi_{\text{ideal}}(\delta_p)\exp\left(\Delta_{\min}\right).$$

By the fact that $\Delta_{\min} \to \infty$, we have

$$\mathbb{E}\xi_{\text{ideal}}(\delta_p)\exp\left(\Delta_{\min}\right) \leq p\exp\left(-(1+o(1))\frac{\Delta_{\min}^2}{8}\right),$$

and thus the proof is complete. □

Finally, we prove Proposition 4.1.

Proof of Proposition 4.1. We divide the proof into three steps.

Step 1. Define $\widehat{P} = \widehat{U}\widehat{U}^T Y \in \mathbb{R}^{d\times p}$ with $\widehat{P}_j$ being the jth column of $\widehat{P}$. Since $\widehat{P}_j = \widehat{U}\widehat{\mu}_j$ for all $j \in [p]$, we have $\|\widehat{P}_j - \widehat{P}_{j'}\| = \|\widehat{\mu}_j - \widehat{\mu}_{j'}\|$ for all $j, j' \in [p]$. This implies

$$\min_{\substack{\theta_1,\ldots,\theta_k\in\mathbb{R}^k \\ z\in[k]^p}} \sum_{j=1}^p \|\widehat{P}_j - \theta_{z_j}\|^2 = \min_{\substack{\beta_1,\ldots,\beta_k\in\mathbb{R}^k \\ z\in[k]^p}} \sum_{j=1}^p \|\widehat{\mu}_j - \beta_{z_j}\|^2.$$

Similarly, define $\theta_a^{(0)} = \widehat{U}\beta_a^{(0)}$ for all $a \in [k]$, we have

$$\sum_{j=1}^p \|\widehat{P}_j - \theta^{(0)}_{z_j^{(0)}}\|^2 = \sum_{j=1}^p \|\widehat{U}\widehat{\mu}_j - \widehat{U}\beta^{(0)}_{z_j^{(0)}}\|^2 = \sum_{j=1}^p \|\widehat{\mu}_j - \beta^{(0)}_{z_j^{(0)}}\|^2.$$

Thus, (39) leads to

$$\sum_{j=1}^p \|\widehat{P}_j - \theta^{(0)}_{z_j^{(0)}}\|^2 \leq M \min_{\substack{\theta_1,\ldots,\theta_k\in\mathbb{R}^k \\ z\in[k]^p}} \sum_{j=1}^p \|\widehat{P}_j - \theta_{z_j}\|^2. \tag{E.10}$$

That is, any $z^{(0)} \in [k]^p$ that satisfies (39) with some $\beta_1^{(0)}, \ldots \beta_k^{(0)}$ also satisfies (E.10) with some $\theta_1^{(0)}, \ldots \theta_k^{(0)}$.

Step 2. It is sufficient to study any $\theta_1^{(0)}, \ldots, \theta_k^{(0)} \in \mathbb{R}^d$ and $z^{(0)} \in [k]^p$ that satisfies (E.10). Let us define $P^* = \mathbb{E}Y$, and we have $P_j^* = \theta^*_{z_j^*}$ according to the model assumption. We first give an error bound for $\|\widehat{P} - P^*\|_{\mathrm{F}}^2$. Since $\widehat{P}$ is the rank-k approximation of Y, we have $\|Y - \widehat{P}\|_{\mathrm{F}}^2 \leq \|Y - P^*\|_{\mathrm{F}}^2$, which implies that $\|\widehat{P} - P^*\|_{\mathrm{F}}^2 \leq 4\max_{\{A\in\mathbb{R}^{d\times p}:\|A\|_{\mathrm{F}}\leq 1, \mathsf{rank}(A)\leq 2k\}} |\langle A, Y - P^*\rangle|^2$. Use a standard random matrix theory result [16], we have $\|Y - P^*\|^2 \lesssim p + d$ with probability at least $1 - e^{-C'(d+p)}$. For any A such that $\|A\|_{\mathrm{F}} \leq 1$ and $\mathsf{rank}(A) \leq 2k$, its singular value decomposition can be written as $A = \sum_{l=1}^{2k} d_l u_l v_l^T$, where $\sum_{l=1}^{2k} d_l^2 \leq 1$. Thus, we have

$$|\langle A, Y - P^* \rangle|^2 = \left| \sum_{l=1}^{2k} d_l u_l^T (Y - P^*) v_l \right|^2 \leq \sum_{l=1}^{2k} \left| u_l^T (Y - P^*) v_l \right|^2 \leq 2k \|Y - P^*\|^2 \lesssim k(p+d).$$

Taking maximum over A, we have $\|\widehat{P} - P^*\|_{\mathrm{F}}^2 \lesssim k(p+d)$ with probability at least $1 - e^{-C'(d+p)}$.

By (E.10), we have

$$\sum_{j=1}^{p} \|\widehat{P}_j - \theta^{(0)}_{z_j^{(0)}}\|^2 \leq M \|\widehat{P} - P^*\|_{\mathrm{F}}^2 \lesssim Mk(p+d),$$

and as a consequence,

$$\sum_{j=1}^{p} \|\theta^*_{z_j^*} - \theta^{(0)}_{z_j^{(0)}}\|^2 \leq 2 \sum_{j=1}^{p} \left(\|\widehat{P}_j - \theta^{(0)}_{z_j^{(0)}}\|^2 + \|\theta^*_{z_j^*} - \widehat{P}_j\|^2 \right) \lesssim (M+1)k(p+d). \tag{E.11}$$

Define

$$S = \left\{ j \in [p] : \|\theta^*_{z_j^*} - \theta^{(0)}_{z_j^{(0)}}\| \geq \frac{\Delta_{\min}}{2} \right\},$$

and we have

$$|S| \leq \frac{\sum_{j=1}^{p} \|\theta^*_{z_j^*} - \theta^{(0)}_{z_j^{(0)}}\|^2}{\left(\frac{\Delta_{\min}}{2}\right)^2} \lesssim \frac{(M+1)k(p+d)}{\Delta_{\min}^2}.$$

We are now going to show that all the data points in S^{c} are all correctly clustered. We define

$$\mathcal{C}_a = \left\{ j \in [p] : z_j^* = a, j \in S^{\mathsf{c}} \right\},$$

for all $a \in [k]$. Under the assumption $\Delta_{\min}^2 / ((M+1)k^2(1 + d/p)) \to \infty$, we have

$$|S| = o(p/k). \tag{E.12}$$

We have the following arguments:

- For each $a \in [k]$, $\mathcal{C}_a$ cannot be empty, as

$$|\mathcal{C}_a| \geq |\{j \in [p] : z_j^* = a\}| - |S| \geq \frac{|\{j \in [p] : z_j^* = a\}|}{2} \geq \frac{\alpha p}{2k}. \tag{E.13}$$

- For each pair $a, b \in [k], a \neq b$, there cannot exist some $j \in \mathcal{C}_a, j' \in \mathcal{C}_b$ such that $z_j^{(0)} = z_{j'}^{(0)}$. Otherwise $\theta^{(0)}_{z_j^{(0)}} = \theta^{(0)}_{z_{j'}^{(0)}}$ would imply

$$\|\theta_a^* - \theta_b^*\| = \left\| \theta^*_{z_j^*} - \theta^*_{z_{j'}^*} \right\| \leq \left\| \theta^*_{z_j^*} - \theta^{(0)}_{z_j^{(0)}} \right\| + \left\| \theta^{(0)}_{z_j^{(0)}} - \theta^{(0)}_{z_{j'}^{(0)}} \right\| + \left\| \theta^{(0)}_{z_{j'}^{(0)}} - \theta^*_{z_{j'}^*} \right\| < \Delta_{\min},$$

  contradicting the definition of $\Delta_{\min}$.

Since $z_j^{(0)}$ can only take values in $[k]$, we conclude that $\{z_j^{(0)} : j \in \mathcal{C}_a\}$ contains only one and different element for all $a \in [k]$. That is, there exists a permutation $\pi_0 \in \Pi_k$, such that

$$z_j^{(0)} = \pi_0(z_j^*), \tag{E.14}$$

for all $j \in S^{\mathsf{c}}$.

Step 3. The last step is to establish an upper bound for $\ell(\pi_0^{-1} \circ z^{(0)}, z^*)$. By (E.11), (E.13) and (E.14), we have

$$\left\|\theta_a^* - \theta_{\pi_0(a)}^{(0)}\right\|^2 = \frac{\sum_{j \in \mathcal{C}_a} \left\|\theta_{z_j^*}^* - \theta_{z_j^{(0)}}^{(0)}\right\|^2}{|\mathcal{C}_a|} \le \frac{\sum_{j=1}^p \left\|\theta_{z_j^*}^* - \theta_{z_j^{(0)}}^{(0)}\right\|^2}{|\mathcal{C}_a|} \lesssim (M+1)k^2\left(1 + \frac{d}{p}\right),$$

for all $a \in [k]$. As a result, together with (E.11), (E.12) and (E.14), we have

$$\begin{aligned}
\ell\left(\pi_0^{-1} \circ z^{(0)}, z^*\right) &= \sum_{j \in [p]} \left\|\theta_{z_j^*}^* - \theta_{\pi_0^{-1}\left(z_j^{(0)}\right)}^*\right\|^2 = \sum_{j \in [p]} \left\|\theta_{z_j^*}^* - \theta_{\pi_0^{-1}\left(z_j^{(0)}\right)}^*\right\|^2 \mathbf{1}_{\left\{z_j^* \neq \pi_0^{-1}\left(z_j^{(0)}\right)\right\}} \\
&\le 2\sum_{j \in [p]} \left(\left\|\theta_{z_j^*}^* - \theta_{z_j^{(0)}}^{(0)}\right\|^2 + \left\|\theta_{z_j^{(0)}}^{(0)} - \theta_{\pi_0^{-1}\left(z_j^{(0)}\right)}^*\right\|^2\right) \mathbf{1}_{\left\{z_j^* \neq \pi_0^{-1}\left(z_j^{(0)}\right)\right\}} \\
&\le 2\sum_{j \in [p]} \left\|\theta_{z_j^*}^* - \theta_{z_j^{(0)}}^{(0)}\right\|^2 + \max_{a \in [k]} \left\|\theta_a^{(0)} - \theta_{\pi_0^{-1}(a)}^*\right\|^2 \sum_{j=1}^p \mathbf{1}_{\left\{z_j^* \neq \pi_0^{-1}\left(z_j^{(0)}\right)\right\}} \\
&\le 2\sum_{j \in [p]} \left\|\theta_{z_j^*}^* - \theta_{z_j^{(0)}}^{(0)}\right\|^2 + |S| \max_{a \in [k]} \left\|\theta_{\pi_0(a)}^{(0)} - \theta_a^*\right\|^2 \\
&\lesssim (M+1)k\left(p + d\right).
\end{aligned}$$

The proof is complete. □

F Proofs in Appendix A

This section collects the proofs of Lemma A.1, Lemma A.2, and Proposition A.1. The conclusions of Theorem A.1 and Corollary A.1 are direct consequences of Theorem 3.1, and thus we omit their proofs. We first need the following technical lemma.

Lemma F.1. *Consider i.i.d. random variables $w_{ij} \sim \mathcal{N}(0,1)$ for $1 \le i \ne j \le p$. Then, for* *any constant $C' > 0$, there exists some constant $C > 0$ only depending on C' such that*

$$\max_{a\in\mathbb{R}^p}\left|\frac{\sum_{1\leq i\neq j\leq p}(a_i-a_j)w_{ij}}{\sqrt{\sum_{1\leq i\neq j\leq p}(a_i-a_j)^2}}\right| \leq C\sqrt{p}, \tag{F.1}$$

$$\sum_{j=1}^{p}\left(\frac{1}{\sqrt{2(p-1)}}\sum_{i\in[p]\backslash\{j\}}(w_{ji}-w_{ij})\right)^2 \leq Cp, \tag{F.2}$$

$$\max_{j\in[p]}\left|\frac{1}{\sqrt{2(p-1)}}\sum_{i\in[p]\backslash\{j\}}(w_{ji}-w_{ij})\right| \leq C\sqrt{\log p}, \tag{F.3}$$

with probability at least $1-(C'p)^{-1}$. We have used the convention that $0/0=0$.

Proof. To bound the first inequality, we define

$$\mathcal{A}=\left\{A=\{a_{ij}\}_{(i,j)\in[p]^2}: a_{ij}=a_i-a_j \text{ for some } a\in\mathbb{R}^p, \|A\|_{\mathrm{F}}\leq 1\right\},$$

and

$$\mathcal{B}=\left\{B=\{b_{ij}\}_{(i,j)\in[p]^2}: \mathsf{rank}(B)\leq 2, \|B\|_{\mathrm{F}}\leq 1\right\}.$$

Then, we have $\mathcal{A}\subset\mathcal{B}$, and

$$\max_{a\in\mathbb{R}^p}\left|\frac{\sum_{1\leq i\neq j\leq p}(a_i-a_j)w_{ij}}{\sqrt{\sum_{1\leq i\neq j\leq p}(a_i-a_j)^2}}\right| = \max_{A\in\mathcal{A}}|\langle A,W\rangle|.$$

By Lemma 3.1 of [2], the covering number of the low-rank set $\mathcal{B}$ is bounded by $e^{O(p)}$, which further implies the same covering number bound for $\mathcal{A}$ by the fact that $\mathcal{A}\subset\mathcal{B}$. In other words, there exists $A_1,\ldots,A_m\in\mathcal{A}$, such that $m\leq e^{C_1p}$, and for any $A\in\mathcal{A}$, $\min_{1\leq l\leq m}\|A_l-A\|_{\mathrm{F}}\leq 1/2$. Let us choose any $A\in\mathcal{A}$, and then let A_l be the matrix in the covering set that satisfies $\|A_l-A\|_{\mathrm{F}}\leq 1/2$. We then have

$$|\langle A,W\rangle|\leq\|A-A_l\|_{\mathrm{F}}\left|\left\langle\frac{A-A_l}{\|A-A_l\|_{\mathrm{F}}},W\right\rangle\right|+|\langle A_l,W\rangle|\leq\frac{1}{2}\max_{A\in\mathcal{A}}|\langle A,W\rangle|+|\langle A_l,W\rangle|,$$

which implies

$$\max_{A\in\mathcal{A}}|\langle A,W\rangle|\leq\frac{1}{2}\max_{A\in\mathcal{A}}|\langle A,W\rangle|+\max_{1\leq l\leq m}|\langle A_l,W\rangle|.$$

After rearrangement, we get $\max_{A\in\mathcal{A}}|\langle A,W\rangle|\leq 2\max_{1\leq l\leq m}|\langle A_l,W\rangle|$. Then, the conclusion follows by a standard union bound argument.

For the second inequality, we use the notation $r_j=\frac{1}{\sqrt{2(p-1)}}\sum_{i\in[p]\backslash\{j\}}(w_{ji}-w_{ij})$. It is clear that $r_j\sim\mathcal{N}(0,1)$ for all $j\in[p]$, and thus we have $\mathbb{E}\left(\sum_{j=1}^p r_j^2\right)=p$. We then calculate the variance. We have

$$\mathsf{Var}\left(\sum_{j=1}^p r_j^2\right)=\sum_{j=1}^p\sum_{l=1}^p\mathbb{E}(r_j^2-1)(r_l^2-1).$$

For $j = l$, we get $\mathbb{E}(r_j^2 - 1)^2 = 2$. For $j \neq l$, we have $\mathbb{E}(r_j^2 - 1)(r_l^2 - 1) = \mathbb{E}r_j^2 r_l^2 - 1$, and

$$\mathbb{E}r_j^2 r_l^2 = \frac{1}{4(p-1)^2}\mathbb{E}\left(\sum_{i\in[p]\setminus\{l\}}(w_{ji} - w_{ij}) + (w_{jl} - w_{lj})\right)^2 \left(\sum_{i\in[p]\setminus\{j\}}(w_{li} - w_{il}) + (w_{lj} - w_{jl})\right)^2.$$

Since the three terms $\sum_{i\in[p]\setminus\{l\}}(w_{ji}-w_{ij})$, $\sum_{i\in[p]\setminus\{j\}}(w_{li}-w_{il})$ and $(w_{jl}-w_{lj})$ are independent, we can expand the above display and calculate the expectation of each term in the expansion, and we get

$$\mathbb{E}r_j^2 r_l^2 = \frac{4(p-2)^2 + 4 + 8(p-2)}{4(p-1)^2} = 1.$$

Therefore, $\mathsf{Var}\left(\sum_{j=1}^p r_j^2\right) = 2p$, and the desired conclusion is obtained by Chebyshev's inequality. Finally, the last inequality is a direct consequence of a union bound argument. $\square$

Now we are ready to state the proofs of Lemma A.1 and Lemma A.2. Note that under the setting of approximate ranking, the error terms are

$$\begin{aligned}
F_j(a,b;z) &= \epsilon_j \frac{2p}{\sqrt{2(p-1)}}(\widehat{\beta}(z^*) - \widehat{\beta}(z))(a-b), \\
G_j(a,b;z) &= \frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \widehat{\beta}(z)\left(a - \frac{p+1}{2}\right)\right)^2 \\
&\quad -\frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \widehat{\beta}(z^*)\left(a - \frac{p+1}{2}\right)\right)^2 \\
&\quad -\frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \widehat{\beta}(z)\left(b - \frac{p+1}{2}\right)\right)^2 \\
&\quad +\frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \widehat{\beta}(z^*)\left(b - \frac{p+1}{2}\right)\right)^2, \\
H_j(a,b) &= \frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \widehat{\beta}(z^*)\left(a - \frac{p+1}{2}\right)\right)^2 \\
&\quad -\frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \beta^*\left(a - \frac{p+1}{2}\right)\right)^2 \\
&\quad -\frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \widehat{\beta}(z^*)\left(b - \frac{p+1}{2}\right)\right)^2 \\
&\quad +\frac{p^2}{p-1}\left(\beta^*\left(a - \frac{1}{p}\sum_{j=1}^p z_j^*\right) - \beta^*\left(b - \frac{p+1}{2}\right)\right)^2.
\end{aligned}$$

Proof of Lemma A.1. For any $z \in [p]^p$ such that $\ell(z, z^*) \leq \tau = o(p^2(\beta^*)^2)$, we have

$$\sum_{1 \leq i \neq j \leq p} (z_i - z_i^* - z_j + z_j^*)^2 \leq 4(p-1) \sum_{j=1}^{p} (z_j - z_j^*)^2 \leq \frac{2}{(\beta^*)^2} \ell(z, z^*) = o(p^2). \tag{F.4}$$

For any $z^* \in \mathcal{R}$, we have $\sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*)^2 \leq p^4$. Moreover, by the definition of $\mathcal{R}$, there exists a $\widetilde{z} \in \Pi_p$ such that $\|z^* - \widetilde{z}\|^2 \leq c_p$. This implies

$$\left(\frac{1}{p} \sum_{j=1}^{p} z_j^* - \frac{p-1}{2} \right)^2 = \left(\frac{1}{p} \sum_{j=1}^{p} z_j^* - \sum_{j=1}^{p} \widetilde{z}_j \right)^2 \leq \frac{1}{p} \|z^* - \widetilde{z}\|^2 \leq \frac{c_p}{p} = o(1), \tag{F.5}$$

and

$$\left| \sum_{j=1}^{p} (z_j^*)^2 - \sum_{j=1}^{p} j^2 \right| = \left| \|z^*\|^2 - \|\widetilde{z}\|^2 \right| \leq \|z^* - \widetilde{z}\| (\|z^*\| + \|\widetilde{z}\|) \lesssim c_p^{1/2} p^{1.5} = o(p^2).$$

Thus,

$$\begin{aligned} \sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*)^2 &= 2p \sum_{j=1}^{p} (z_j^*)^2 - 2 \left(\sum_{j=1}^{p} z_j^* \right)^2 \\ &\geq 2p \left(\sum_{j=1}^{p} j^2 - o(p^2) \right) - (1 + o(1)) 2 \left(\sum_{j=1}^{p} j \right)^2 \\ &\geq \frac{p^4}{12}. \end{aligned} \tag{F.6}$$

Therefore,

$$\begin{aligned} \sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2 &\geq \frac{1}{2} \sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*)^2 - \sum_{1 \leq i \neq j \leq p} (z_i - z_i^* - z_j + z_j^*)^2 \\ &\geq \frac{p^4}{24} - o(p^2) \\ &\geq \frac{p^4}{25}, \end{aligned} \tag{F.7}$$

where the last inequality assumes p is sufficiently large. We then introduce more notations. We define

$$\beta(z) = \beta^* \frac{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)(z_i^* - z_j^*)}{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2},$$

and

$$\bar{w}(z) = \frac{\sum_{1 \leq i \neq j \leq p} (z_i - z_j) w_{ij}}{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2}.$$

We write $w_{ij} = Y_{ij} - \beta^*(z_i^* - z_j^*)$ so that $\epsilon_j = \frac{1}{\sqrt{2(p-1)}} \sum_{i \in [p] \setminus \{j\}} (w_{ji} - w_{ij})$. We consider the event that the three inequalities (F.1)-(F.3) hold. We first derive bounds for $|\widehat{\beta}(z^*) - \beta^*|$, $|\beta(z) - \beta^*|$ and $|\bar{w}(z) - \bar{w}(z^*)|$. By (F.1), we have

$$|\widehat{\beta}(z^*) - \beta^*| = \left| \frac{\sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*) w_{ij}}{\sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*)^2} \right| \lesssim p^{-1.5}. \tag{F.8}$$

By (F.4) and (F.7), we have

$$\begin{aligned} |\beta(z) - \beta^*| &= \left| \beta^* \frac{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)(z_i^* - z_j^* - z_i + z_j)}{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2} \right| \\ &\leq \frac{25|\beta^*|}{p^4} \sqrt{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2} \sqrt{\sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^* - z_i + z_j)^2} \\ &\leq \frac{25|\beta^*|}{p^2} \sqrt{\frac{2}{(\beta^*)^2} \ell(z, z^*)} \\ &\lesssim \frac{\sqrt{\ell(z, z^*)}}{p^2}. \end{aligned} \tag{F.9}$$

Next, we bound $|\bar{w}(z) - \bar{w}(z^*)|$. We have

$$\begin{aligned} |\bar{w}(z) - \bar{w}(z^*)| \leq & \left| \frac{\sum_{1 \leq i \neq j \leq p} (z_i - z_j - z_i^* + z_j^*) w_{ij}}{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2} \right| \\ & + \left| \frac{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2 - \sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*)^2}{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2 \sqrt{\sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*)^2}} \right| \left| \frac{\sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*) w_{ij}}{\sqrt{\sum_{1 \leq i \neq j \leq p} (z_i^* - z_j^*)^2}} \right|. \end{aligned}$$

We bound the two terms on the right hand side of the above inequality separately. The first term can be bounded by

$$\begin{aligned} & \frac{\sqrt{\sum_{1 \leq i \neq j \leq p} (z_i - z_i^* - z_j + z_j^*)^2}}{\sum_{1 \leq i \neq j \leq p} (z_i - z_j)^2} \left| \frac{\sum_{1 \leq i \neq j \leq p} (z_i - z_j - z_i^* + z_j^*) w_{ij}}{\sqrt{\sum_{1 \leq i \neq j \leq p} (z_i - z_i^* - z_j + z_j^*)^2}} \right| \\ \leq & \frac{25}{p^4} \sqrt{\frac{2}{(\beta^*)^2} \ell(z, z^*)} \left| \frac{\sum_{1 \leq i \neq j \leq p} (z_i - z_j - z_i^* + z_j^*) w_{ij}}{\sqrt{\sum_{1 \leq i \neq j \leq p} (z_i - z_i^* - z_j + z_j^*)^2}} \right| \\ \lesssim & \frac{\sqrt{p \ell(z, z^*)}}{|\beta^*| p^4}, \end{aligned}$$

where we have used the inequalities (F.1), (F.4), and (F.7). By (F.1) and (F.7), the second term can be bounded by

$$\begin{aligned}
& C_1 p^{-5.5}\left|\sum_{1\leq i\neq j\leq p}(z_i-z_j)^2-\sum_{1\leq i\neq j\leq p}(z_i^*-z_j^*)^2\right| \\
\leq\ & C_1 p^{-5.5}\left|\sum_{1\leq i\neq j\leq p}(z_i-z_j)(z_i^*-z_j^*-z_i+z_j)\right| \\
& +C_1 p^{-5.5}\left|\sum_{1\leq i\neq j\leq p}(z_i^*-z_j^*)(z_i^*-z_j^*-z_i+z_j)\right| \\
\leq\ & C_1 p^{-5.5}\left(\sqrt{\sum_{1\leq i\neq j\leq p}(z_i-z_j)^2}+\sqrt{\sum_{1\leq i\neq j\leq p}(z_i^*-z_j^*)^2}\right)\sqrt{\sum_{1\leq i\neq j\leq p}(z_i-z_i^*-z_j+z_j^*)^2} \\
\lesssim\ & \frac{\sqrt{p\ell(z,z^*)}}{|\beta^*|p^4},
\end{aligned}$$

where we have used (F.4) in the last inequality. Combining the two bounds, we obtain

$$|\bar{w}(z)-\bar{w}(z^*)|\lesssim\frac{\sqrt{p\ell(z,z^*)}}{|\beta^*|p^4}. \tag{F.10}$$

From (F.8), (F.9) and (F.10), we can further derive

$$|\widehat{\beta}(z)-\widehat{\beta}(z^*)| \ \leq\ |\beta(z)-\beta^*|+|\bar{w}(z)-\bar{w}(z^*)|\lesssim\frac{\sqrt{\ell(z,z^*)}}{p^2}, \tag{F.11}$$

under the condition $p(\beta^*)^2\geq 1$.

We are ready to prove (A.7)-(A.9). Recall that $\epsilon_j=\frac{1}{\sqrt{2(p-1)}}\sum_{i\in[p]\setminus\{j\}}(w_{ji}-w_{ij})$, and we have

$$\begin{aligned}
\sum_{j=1}^p\epsilon_j^2 &\leq \sum_{j=1}^p\left(\frac{1}{\sqrt{2(p-1)}}\sum_{i\in[p]\setminus\{j\}}(w_{ji}-w_{ij})\right)^2 \\
&\lesssim p,
\end{aligned} \tag{F.12}$$

by (F.2). Moreover, from (A.5), since $z^*\in\mathcal{R}$, we have

$$\Delta_j(a,b)^2=(1+o(1))\frac{2p^2(\beta^*)^2}{p-1}(a-b)^2. \tag{F.13}$$

Thus,

$$\begin{aligned}
& \sum_{j=1}^p\max_{b\in[k]\setminus\{z_j^*\}}\frac{F_j(z_j^*,b;z)^2\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)} \\
=\ & \frac{|\widehat{\beta}(z)-\widehat{\beta}(z^*)|^2}{|\beta^*|^2\ell(z,z^*)}\sum_{j=1}^p\epsilon_j^2 \\
\lesssim\ & \frac{1}{p^2},
\end{aligned}$$

where we have used (F.11), (F.12), (F.13) and the condition $p(\beta^*)^2 \geq 1$ in the last inequality. Taking maximum, we obtain (A.7). For (A.8), we note that

$$\begin{aligned}
|G_j(a,b;z)| &\leq \frac{p^2}{p-1}\left|\left(a-\frac{p+1}{2}\right)^2-\left(b-\frac{p+1}{2}\right)^2\right||\widehat{\beta}(z)-\widehat{\beta}(z^*)|^2 \\
&+\frac{2p^2}{p-1}\left(a-\frac{p+1}{2}\right)^2|\widehat{\beta}(z)-\widehat{\beta}(z^*)||\widehat{\beta}(z^*)-\beta^*| \\
&+\frac{2p^2}{p-1}\left|b-\frac{p+1}{2}\right|\left|\widehat{\beta}(z)-\widehat{\beta}(z^*)\right|\left|\left(b-\frac{p+1}{2}\right)\widehat{\beta}(z^*)-\left(a-\frac{p+1}{2}\right)\beta^*\right| \\
&+\frac{2p^3}{p-1}|\beta^*||\widehat{\beta}(z)-\widehat{\beta}(z^*)|\left|\frac{1}{p}\sum_{j=1}^p z_j^*-\frac{p+1}{2}\right| \\
&\leq \frac{p^4}{p-1}|\widehat{\beta}(z)-\widehat{\beta}(z^*)|^2+\frac{4p^4}{p-1}|\widehat{\beta}(z)-\widehat{\beta}(z^*)||\widehat{\beta}(z^*)-\beta^*| \\
&+\frac{4p^3}{p-1}|a-b||\beta^*||\widehat{\beta}(z)-\widehat{\beta}(z^*)|.
\end{aligned}$$

Therefore, for any subset $T \subset [p]$, we have

$$\begin{aligned}
&\frac{\tau}{4\Delta_{\min}^2|T|+\tau}\sum_{j\in T}\max_{b\in[k]\backslash\{z_j^*\}}\frac{G_j(z_j^*,b;z)^2\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)} \\
\lesssim\;& \frac{\tau p^4|\widehat{\beta}(z)-\widehat{\beta}(z^*)|^4}{|\beta^*|^4\ell(z,z^*)}+\frac{\tau p^4|\widehat{\beta}(z)-\widehat{\beta}(z^*)|^2|\widehat{\beta}(z^*)-\beta^*|^2}{|\beta^*|^4\ell(z,z^*)}+\frac{\tau p^2|\widehat{\beta}(z)-\widehat{\beta}(z^*)|^2}{|\beta^*|^2\ell(z,z^*)} \\
\lesssim\;& \frac{\tau}{p^2|\beta^*|^2}+\frac{1}{p|\beta^*|^2},
\end{aligned}$$

where we have used (F.8), (F.11), and $\ell(z,z^*) \leq \tau = o(p^2(\beta^*)^2)$. Taking maximum, we thus obtain (A.8). Finally, for (A.9), we have

$$\begin{aligned}
\frac{|H_j(a,b)|}{\Delta_j(a,b)^2} &\leq \frac{1}{2|\beta^*|^2}(\widehat{\beta}(z^*)-\beta^*)^2\left(a-\frac{p+1}{2}\right)^2+\frac{1}{2|\beta^*|^2}(\widehat{\beta}(z^*)-\beta^*)^2\left(b-\frac{p+1}{2}\right)^2 \\
&+2p\frac{|\widehat{\beta}(z^*)-\beta^*|}{|\beta^*|} \\
&\lesssim \frac{1}{\sqrt{p}|\beta^*|},
\end{aligned}$$

where we have used (F.8). We thus obtain (A.9) by taking maximum. The proof is complete.

□

Proof of Lemma A.2. By (F.13), there exists some $\delta' = \delta'_p = o(1)$, such that

$$\begin{aligned}
& \mathbf{1}_{\left\{\left\langle \epsilon_j, \nu_j(\widehat{B}(z^*), z_j^*) - \nu_j(\widehat{B}(z^*), b)\right\rangle \leq -\frac{1-\delta}{2}\Delta_j(z_j^*, b)^2\right\}} \\
\leq\ & \mathbf{1}_{\left\{\epsilon_j \frac{2p}{\sqrt{2(p-1)}}\widehat{\beta}(z^*)(z_j^*-b) \leq -\frac{1-\delta-\delta'}{2}\frac{2p^2(\beta^*)^2}{p-1}(z_j^*-b)^2\right\}} \\
\leq\ & \mathbf{1}_{\left\{\epsilon_j \frac{2p}{\sqrt{2(p-1)}}\beta^*(z_j^*-b) \leq -\frac{1-\delta-\delta'-\bar{\delta}}{2}\frac{2p^2(\beta^*)^2}{p-1}(z_j^*-b)^2\right\}} \\
& + \mathbf{1}_{\left\{\epsilon_j \frac{2p}{\sqrt{2(p-1)}}(\widehat{\beta}(z^*)-\beta^*)(z_j^*-b) \leq -\frac{\bar{\delta}}{2}\frac{2p^2(\beta^*)^2}{p-1}(z_j^*-b)^2\right\}}.
\end{aligned}$$

By (F.3), (F.8), and $p(\beta^*)^2 \to \infty$, we have

$$\max_{j\in[p]} \frac{\left|\epsilon_j \frac{2p}{\sqrt{2(p-1)}}(\widehat{\beta}(z^*) - \beta^*)(z_j^* - b)\right|}{\frac{2p^2(\beta^*)^2}{p-1}(z_j^* - b)^2} = o\left(\frac{\sqrt{\log p}}{p}\right),$$

with probability at least $1-p^{-1}$. Therefore, we can set $\bar{\delta} = \bar{\delta}_p$ for some sequence $\bar{\delta}_p \to 0$ and $\bar{\delta}_p \gtrsim \frac{\sqrt{\log p}}{p}$, and then

$$\mathbf{1}_{\left\{\epsilon_j \frac{2p}{\sqrt{2(p-1)}}(\widehat{\beta}(z^*)-\beta^*)(z_j^*-b) \leq -\frac{\bar{\delta}}{2}\frac{2p^2(\beta^*)^2}{p-1}(z_j^*-b)^2\right\}} = 0,$$

for all $j \in [p]$ with probability at least $1 - p^{-1}$. This immediately implies that $\xi_{\text{ideal}}(\delta_p) \leq \widetilde{\xi}_{\text{ideal}}(\delta_p + \delta'_p + \bar{\delta}_p)$ with high probability, where

$$\widetilde{\xi}_{\text{ideal}}(\delta_p+\delta'_p+\bar{\delta}_p) = \frac{2p^2(\beta^*)^2}{p-1}\sum_{j=1}^{p}\sum_{b\in[p]\backslash\{z_j^*\}}(z_j^*-b)^2\mathbf{1}_{\left\{\epsilon_j \frac{2p}{\sqrt{2(p-1)}}\beta^*(z_j^*-b) \leq -\frac{1-\delta_p-\delta'_p-\bar{\delta}_p}{2}\frac{2p^2(\beta^*)^2}{p-1}(z_j^*-b)^2\right\}}.$$

A standard Gaussian tail bound implies

$$\begin{aligned}
& \mathbb{E}\widetilde{\xi}_{\text{ideal}}(\delta_p + \delta'_p + \bar{\delta}_p) \\
=\ & \frac{2p^2(\beta^*)^2}{p-1}\sum_{j=1}^{p}\sum_{b\in[p]\backslash\{z_j^*\}}(z_j^* - b)^2\mathbb{P}\left(\mathcal{N}(0,1) \leq -\frac{1-\delta_p-\delta'_p-\bar{\delta}_p}{2}\sqrt{\frac{2p^2(\beta^*)^2}{p-1}(z_j^*-b)^2}\right) \\
\leq\ & \sum_{j=1}^{p}\sum_{l=1}^{\infty}\frac{4p^2(\beta^*)^2}{p-1}l^2\exp\left(-\left(\frac{1-\delta_p-\delta'_p-\bar{\delta}_p}{2}\right)^2\frac{p^2(\beta^*)^2}{p-1}l^2\right) \\
\leq\ & p\exp\left(-(1+o(1))\frac{p(\beta^*)^2}{4}\right),
\end{aligned}$$

where we have used the conditions $p(\beta^*)^2 \to \infty$ and $\delta_p + \delta'_p + \bar{\delta}_p = o(1)$ in the last inequality. Finally, by Markov's inequality, with probability at least $1 - \exp\left(-\sqrt{p(\beta^*)^2}\right)$, we have

$$\begin{aligned}
\widetilde{\xi}_{\text{ideal}}(\delta_p + \delta'_p + \bar{\delta}_p) &\leq \mathbb{E}\widetilde{\xi}_{\text{ideal}}(\delta_p + \delta'_p + \bar{\delta}_p)\exp\left(\sqrt{p(\beta^*)^2}\right) \\
&\leq p\exp\left(-(1+o(1))\frac{p(\beta^*)^2}{4}\right),
\end{aligned}$$

as $p(\beta^*)^2 \to \infty$. Since $\xi_{\text{ideal}}(\delta_p) \leq \widetilde{\xi}_{\text{ideal}}(\delta_p + \delta'_p + \bar{\delta}_p)$, the proof is complete. □

Finally, we state the proof of Proposition A.1.

Proof of Proposition A.1. Note that we have the following fact. Consider any $x = (x_1, x_2, \ldots, x_m)^T \in \mathbb{R}^m$. Let $y = \operatorname{argmin}_{z\in\Pi_p} \sum_{i=1}^{p}(x_i - z_i)^2$. Then for any pair (i,j) such that $x_i < x_j$, we must have $y_i < y_j$. Otherwise if $y_i > y_j$, since $((x_i - y_i)^2 + (x_j - y_j)^2) - ((x_i - y_j)^2 + (x_j - y_i)^2) = -2(x_i - x_j)(y_i - y_j) > 0$, we can always swap y_i and y_j to make $\sum_{i=1}^{p}(x_i - y_i)^2$ strictly smaller. This indicates that y preserves the order of x.

As a result, since a linear transformation does not change the rank, we can write $z^{(0)}$ as

$$z^{(0)} = \operatorname*{argmin}_{z\in\Pi_p} \sum_{j=1}^{p}\left(\frac{\sqrt{2(p-1)}}{2p\beta^*}T_j + \frac{p+1}{2} - z_j\right)^2. \tag{F.14}$$

Since $z^* \in \mathcal{R}$, there exists some $\widetilde{z} \in \Pi_p$ such that $L_2(\widetilde{z}, z^*) = o(1)$. By (F.14),

$$\sum_{j=1}^{p}\left(\frac{\sqrt{2(p-1)}}{2p\beta^*}T_j + \frac{p+1}{2} - z_j^{(0)}\right)^2 \le \sum_{j=1}^{p}\left(\frac{\sqrt{2(p-1)}}{2p\beta^*}T_j + \frac{p+1}{2} - \widetilde{z}_j\right)^2.$$

We then have

$$\begin{aligned}
&L_2(z^{(0)}, \widetilde{z}) = p^{-1}\sum_{j=1}^{p}(z_j^{(0)} - \widetilde{z}_j)^2 \\
&\le 2p^{-1}\sum_{j=1}^{p}\left(\frac{\sqrt{2(p-1)}}{2p\beta^*}T_j + \frac{p+1}{2} - z_j^{(0)}\right)^2 + 2p^{-1}\sum_{j=1}^{p}\left(\frac{\sqrt{2(p-1)}}{2p\beta^*}T_j + \frac{p+1}{2} - \widetilde{z}_j\right)^2 \\
&\le 4p^{-1}\sum_{j=1}^{p}\left(\frac{\sqrt{2(p-1)}}{2p\beta^*}T_j + \frac{p+1}{2} - \widetilde{z}_j\right)^2 \\
&\le 12p^{-1}\sum_{j=1}^{p}\left(\frac{\sqrt{2(p-1)}}{2p\beta^*}T_j + \frac{1}{p}\sum_{j=1}^{p} z_j^* - z_j^*\right)^2 + 12p^{-1}\sum_{j=1}^{p}(\widetilde{z}_j - z_j^*)^2 + 12\left(\frac{1}{p}\sum_{j=1}^{p} z_j^* - \frac{p+1}{2}\right)^2 \\
&= \frac{6\,(p-1)}{p^3\beta^{*2}}\sum_{j=1}^{p}\epsilon_j^2 + 12L_2(\widetilde{z}, z^*) + 12\left(\frac{1}{p}\sum_{j=1}^{p} z_j^* - \frac{p+1}{2}\right)^2,
\end{aligned}$$

where the last equation is due to the fact that $T_j = \mu_j(B^*, z_j^*) + \epsilon_j$. By (F.12), (F.5), and $L_2(\widetilde{z}, z^*) = o(1)$, we have

$$L_2(z^{(0)}, \widetilde{z}) \lesssim \frac{1}{p(\beta^*)^2} + o(1),$$

with probability at least $1 - p^{-1}$. By $L(z^{(0)}, z^*) \le 2L(z^{(0)}, \widetilde{z}) + 2L_2(\widetilde{z}, z^*)$, we have

$$L_2(z^{(0)}, z^*) \lesssim \frac{1}{p(\beta^*)^2} + o(1),$$

with high probability. When $p(\beta^*)^2 \to \infty$, we clearly have $L_2(z^{(0)}, z^*) = o(1)$. When $p(\beta^*)^2 = O(1)$, we have $L_2(z^{(0)}, z^*) \lesssim \min\left(p^2, \frac{1}{p(\beta^*)^2}\right)$, where $L_2(z^{(0)}, z^*) \lesssim p^2$ is by the definition of the loss. $\square$

G Proofs in Section 5

In this section, we will prove results presented in Section 5. Most efforts will be devoted to the proofs of Lemma 5.1, Lemma 5.2, and Proposition 5.1. With these results established, the conclusions of Theorem 5.1, Theorem 5.2, and Corollary 5.1 easily follow.

Let us introduce some more notation to facilitate the proofs. Given some vector $v \in \mathbb{R}^d$, some matrix $A \in \mathbb{R}^{d' \times d}$ and some set $S \subset [d]$, we use $v_S \in \mathbb{R}^{|S|}$ for the sub-vector $(v_j : j \in S)$ and $A_S \in \mathbb{R}^{d' \times |S|}$ for the sub-matrix $(A_{ij} : i \in [d'], j \in S)$. We denote $\operatorname{span}(A)$ to be the space spanned by the columns of A. For any $j \in [d]$, we denote $[v]_j = v_j$ to be the jth coordinate of v. We also write $\phi_S : [d] \to [|S|]$ for the map that satisfies $v_j = [v_S]_{\phi_S(j)}$. The domain of the map ϕ_S can also be extended to sets so that for any $S' \subset S$, we can write $v_{S'} = [v_S]_{\phi_S(S')}$. For any $j \in S$, we write $S_{-j} = S \backslash \{j\}$. We use I_d for the $d \times d$ identity matrix, and sometimes just write I for simplicity if the dimension is clear from the context. Given any square matrix $A \in \mathbb{R}^{d \times d}$, we use $\operatorname{diag}\{A\}$ for the diagonal matrix whose diagonal entries are identical to those of A. For two random elements X and Y, we write $X \stackrel{d}{=} Y$ if their distributions are identical, and $X \perp Y$ if they are independent of each other.

We first state and prove three technical lemmas.

Lemma G.1. *Assume $s \log p \le n$. Consider a random matrix $X \in \mathbb{R}^{n \times p}$ with i.i.d. entries $X_{ij} \sim \mathcal{N}(0,1)$, an independent $w \sim \mathcal{N}(0, I_n)$, some $S^* \subset [p]$ satisfying $|S^*| = s$, and some $\beta^* \in \mathbb{R}^p$. For any $S \subset [p]$, denote $P_S = X_S \left(X_S^T X_S\right)^{-1} X_S^T$ to be the projection matrix onto the subspace* $\operatorname{span}(X_S)$. *We also use the notation $P_j = X_j X_j^T / \|X_j\|^2$, where X_j represents the jth column of X. Then, for any constants $C_0, C' > 0$, there exists some constant $C > 0$*

only depending on C_0, C' such that

$$\max_{j\in[p]} \left| \|X_j\|^2 - n \right| \le C\sqrt{n\log p} \le n/2, \tag{G.1}$$

$$\max_{S\subset[p]:|S|\le 2C_0 s} \left\| \left(X_S^T X_S\right)^{-1} \right\| \le \frac{C}{n}, \tag{G.2}$$

$$\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \left\| \left(X_S^T X_S\right)^{-1} X_S^T X_T \right\|^2 \le C\frac{s\log p}{n}, \tag{G.3}$$

$$\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \frac{1}{|T|} \left\| \left(X_T^T (I-P_S) X_T\right)^{-1} X_T^T (I-P_S) w \right\|^2 \le \frac{C\log p}{n}, \tag{G.4}$$

$$\max_{S\subset[p]:|S|\le 2C_0 s} \max_{j\in S^*\cap S^{\mathsf{c}}} \left| X_j^T P_S X_j \right| \le Cs\log p, \tag{G.5}$$

$$\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \left\| X_T^T (I-P_S) X_T - \operatorname{diag}\left\{ X_T^T (I-P_S) X_T \right\} \right\|^2 \le Cns\log p, \tag{G.6}$$

$$\max_{S\subset[p]:|S|\le 2C_0 s} \frac{1}{|S^*\cap S^{\mathsf{c}}| + |S^{*\mathsf{c}}\cap S|} \sum_{j\in S^*\cap S^{\mathsf{c}}} \left(X_j^T P_S w\right)^2 \le Cs\log^2 p, \tag{G.7}$$

$$\max_{j\notin S^*} \left(X_j^T P_{S^*} w\right)^2 \le Cs\log p, \tag{G.8}$$

$$\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \frac{1}{|T|} \left\| \left(X_T^T (I-P_S) X_T\right)^{-1} X_T^T P_S w \right\|^2 \le \frac{Cs\log p}{n^2}, \tag{G.9}$$

$$\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \left\| X_T^T (I-P_S) X_T - (n-|S|) I_{|T|} \right\|^2 \le Cns\log p, \tag{G.10}$$

$$\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \left\| \left(X_T^T (I-P_S) X_T\right)^{-1} \right\| \le \frac{C}{n}, \tag{G.11}$$

$$\max_{S\subset[p]:|S|\le 2C_0 s} \frac{1}{|S|} \left\| X_S^T w \right\|^2 \le Cn\log p, \tag{G.12}$$

$$\max_{S,T\subset[p]:S\cap T=\emptyset,|S|\le 2C_0 s} \frac{1}{\left\|\beta^*_{S^{\mathsf{c}}\cap S^*}\right\|^2} \frac{1}{|T|\vee s} \left\| X_T^T (I-P_S) X_{S^{\mathsf{c}}\cap S^*} \beta^*_{S^{\mathsf{c}}\cap S^*} \right\|^2 \le Cn\log p, \tag{G.13}$$

$$\max_{S,T\subset[p]:T\cap(S\cup S^*)=\emptyset,|S|\le 2C_0 s} \frac{1}{|S^*\cap S^{\mathsf{c}}| + |S^{*\mathsf{c}}\cap S|} \frac{1}{|T|} \left\| X_T^T (P_{S^*} - P_S) w \right\|^2 \le C\log^2 p, \tag{G.14}$$

$$\max_{j\in S^*} \left| \|X_j\|^{-1} X_j^T X_{S^*_{-j}} \left(X_{S^*_{-j}}^T (I-P_j) X_{S^*_{-j}} \right)^{-1} X_{S^*_{-j}}^T (I-P_j) w \right| \le \sqrt{\frac{Cs\log^2 p}{n}}, \tag{G.15}$$

with probability at least $1-\exp(-C'\log p)$. We have used the convention that 0/0 = 0.

Proof. We first present a fact that will be used repeatedly in the proof. For two independent $\xi_1, \xi_2 \sim \mathcal{N}(0, I_d)$, we have $\xi_1^T\xi_2 = \|\xi_1\| (\|\xi_1\|^{-1}\xi_1)^T \xi_2 \overset{d}{=} \|\xi_1\|\zeta$, where $\zeta \sim \mathcal{N}(0,1)$ and $\zeta \perp \xi_1$. Throughout the proof, we will use $c, c', c_1, c_2, \ldots$ as generic constants whose values may change from place to place. We refer to Lemma E.1 for the χ^2 tail probability bound.

Equation (G.1): We have $\|X_j\|^2 \sim \chi^2_n$. Then the χ^2 tail bound and a union bound argument over $j \in [p]$ lead to the desired bound.

Equation (G.2): It is sufficient to study the smallest eigenvalue of $X_S^T X_S$. For a fixed S and $\theta \in \mathbb{R}^{|S|}$ such that $\|\theta\| = 1$, we have $\theta^T X_S^T X_S \theta \sim \chi^2_n$. Thus $\mathbb{P}\left(\left|\theta^T X_S^T X_S\theta - n\right| \le \frac{n}{2}\right) \le$

$2\exp(-n/16)$. By a standard ϵ-net argument [16], we can obtain

$$\mathbb{P}\left(\min_{\theta\in\mathbb{R}^{|S|}:\|\theta\|=1}\theta^T X_S^T X_S\theta \le c_1 n\right) \le c_2^{|S|}\exp(-n/16).$$

We then take a union bound over S to obtain

$$\mathbb{P}\left(\min_{S\subset[p]:|S|\le 2C_0 s}\min_{\theta\in\mathbb{R}^{|S|}:\|\theta\|=1}\theta^T X_S^T X_S\theta \le c_1 n\right) \le 2\binom{p}{2C_0 s}c_2^{|S|}\exp(-n/16).$$

Thus

$$\mathbb{P}\left(\max_{S\subset[p]:|S|\le 2C_0 s}\left\|\left(X_S^T X_S\right)^{-1}\right\| \ge \frac{1}{c_1 n}\right) \le \exp(-c_3 n),$$

for some constant $c_3 > 0$.

Equation (G.3): Conditioning on X_S, we have $\left(X_S^T X_S\right)^{-1} X_S^T X_T \overset{d}{=} \left(X_S^T X_S\right)^{-\frac{1}{2}}\zeta$, where $\zeta \sim \mathcal{N}(0, I_{|S|})$. Thus $\left\|\left(X_S^T X_S\right)^{-1} X_S^T X_T\right\|^2 \overset{d}{=} \zeta^T\left(X_S^T X_S\right)^{-1}\zeta \le \left\|\left(X_S^T X_S\right)^{-1}\right\|\|\zeta\|^2$. We have

$$\mathbb{P}\left(\left\|\left(X_S^T X_S\right)^{-1} X_S^T X_T\right\|^2 \ge 4c\left\|\left(X_S^T X_S\right)^{-1}\right\| s\log p \,\middle|\, X_S\right) \le \exp(-cs\log p).$$

A union bound over T gives

$$\mathbb{P}\left(\max_{T\subset[p]:|T|\le 2C_0 s}\left\|\left(X_S^T X_S\right)^{-1} X_S^T X_T\right\|^2 \ge 4c\left\|\left(X_S^T X_S\right)^{-1}\right\| s\log p \,\middle|\, X_S\right) \le \binom{p}{2C_0 s}\exp(-cs\log p).$$

Consequently, for a fixed S,

$$\begin{aligned}&\mathbb{P}\left(\max_{T\subset[p]:|T|\le 2C_0 s}\left\|\left(X_S^T X_S\right)^{-1} X_S^T X_T\right\|^2 \ge \frac{4cs\log p}{c_1 n}\right)\\ &\le \binom{p}{2C_0 s}\exp(-cs\log p) + \mathbb{P}\left(\left\|\left(X_S^T X_S\right)^{-1}\right\| \ge \frac{1}{c_1 n}\right).\end{aligned}$$

Using the result established above when proving (G.2), together with a union bound over S, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s}\left\|\left(X_S^T X_S\right)^{-1} X_S^T X_T\right\|^2 \ge \frac{4cs\log p}{c_1 n}\right)\\ &\le \binom{p}{2C_0 s}\binom{p}{2C_0 s}\exp(-cs\log p) + \mathbb{P}\left(\max_{S\subset[p]:|S|\le 2C_0 s}\left\|\left(X_S^T X_S\right)^{-1}\right\| \ge \frac{1}{c_1 n}\right)\\ &\le \binom{p}{2C_0 s}\binom{p}{2C_0 s}\exp(-cs\log p) + \exp(-c_2 n)\\ &\le \exp(-c_3 s\log p).\end{aligned}$$

Equation (G.4): The proof of (G.4) is very similar to that of (G.3). Conditioning on X_S, X_T, we have $\left(X_T^T(I-P_S)X_T\right)^{-1} X_T^T(I-P_S)w \overset{d}{=} \left(X_T^T(I-P_S)X_T\right)^{-1}\zeta$ where $\zeta \sim$

$\mathcal{N}(0, I_{|T|})$. Consequently, $\left\| \left(X_T^T(I-P_S)X_T\right)^{-1} X_T^T(I-P_S)w \right\|^2$ is stochastically dominated by $\left\| \left(X_T^T(I-P_S)X_T\right)^{-1} \right\| \|\zeta\|^2$. Similar to the proof of (G.3), we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \left\| \frac{1}{|T|} \left(X_T^T(I-P_S)X_T\right)^{-1} X_T^T(I-P_S)w \right\|^2 \ge \frac{4c\log p}{c_1 n} \right) \\
&\le \sum_{S\subset[p]:|S|\le 2C_0 s} \sum_{m=0}^{2C_0 s} \sum_{T\subset[p]:|T|=m} \mathbb{P}\left(\chi_m^2 \ge 4cm\log p\right) \\
&\quad + \mathbb{P}\left(\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s} \left\| \left(X_T^T(I-P_S)X_T\right)^{-1} \right\| \ge \frac{1}{c_1 n} \right) \\
&\le \binom{p}{m} \sum_{m=0}^{2C_0 s} \binom{p}{2C_0 s} \exp\left(-cm\log p\right) + \exp\left(-cn\right) \\
&\le \exp\left(-c'\log p\right),
\end{aligned}$$

where in the second to the last inequality, we use (G.11), which will be proved later.

Equation (G.5): First we fix some S. Then $X_j^T P_S X_j$ is stochastically dominated by a $\chi^2_{|S|}$. We have $\mathbb{P}\left(X_j^T P_S X_j \ge 4cs\log p\right) \le \exp\left(-cs\log p\right)$. A union bound over S and j leads to the desired result.

Equation (G.6): For any pair (S,T), we have

$$\begin{aligned}
&\left\|X_T^T(I-P_S)X_T - \operatorname{diag}\left\{X_T^T(I-P_S)X_T\right\}\right\| \\
&\le \left\|X_T^T(I-P_S)X_T - (n-|S|)\, I\right\| + \left\|(n-|S|)\, I - \operatorname{diag}\left\{X_T^T(I-P_S)X_T\right\}\right\|.
\end{aligned}$$

The first term can be controlled by (G.10), to be proved later. For the second term, we have

$$\begin{aligned}
\left\|(n-|S|)\, I - \operatorname{diag}\left\{X_T^T(I-P_S)X_T\right\}\right\| &= \max_{j\in T} \left|X_j^T\left(I-P_S\right)X_j - (n-|S|)\right| \\
&\le \max_{j\in T} \left|\|X_j\|^2 - n\right| + \max_{j\in T}\left|X_j^T P_S X_j - |S|\right|,
\end{aligned}$$

which can be bounded by (G.1) and (G.5). Combining the two terms together gives the desired result.

Equation (G.7): For a fixed S and any $j \in S^* \cap S^{\mathsf{c}}$, using the fact we give at the beginning of the proof, we have $X_j^T P_S w \overset{d}{=} \|P_S w\|\,\xi_j$ where $\xi_j \sim \mathcal{N}(0,1)$ and $\xi_j \perp \|P_S w\|$. Since ξ_j only depends on $X_j^T(\|P_S w\|^{-1} P_S w)$, we have the independence among $\{\xi_j\}_{j\in S^*\cap S^{\mathsf{c}}}$. As a result, we have $\sum_{j\in S^*\cap S^{\mathsf{c}}} \left(X_j^T P_S w\right)^2 \overset{d}{=} \zeta\xi$, where $\zeta \sim \chi^2_{|S|}$, $\xi \sim \chi^2_{|S^*\cap S^{\mathsf{c}}|}$ and $\zeta \perp \xi$. Similar arguments will also be used later to prove (G.8)-(G.9) and (G.12)-(G.15) and will be omitted there. Then

$$\begin{aligned}
&\mathbb{P}\left(\sum_{j\in S^*\cap S^{\mathsf{c}}} \left(X_j^T P_S w\right)^2 \ge 16c^2 s\log^2 p\left(|S^*\cap S^{\mathsf{c}}| + |S^{*\mathsf{c}}\cap S|\right) \right) \\
&\le \mathbb{P}\left(\zeta \ge 4cs\log p\right) + \mathbb{P}\left(\xi \ge 4c\left(|S^*\cap S^{\mathsf{c}}| + |S^{*\mathsf{c}}\cap S|\right)\log p\right) \\
&\le \exp\left(-cs\log p\right) + \exp\left(-c\left(|S^*\cap S^{\mathsf{c}}| + |S^{*\mathsf{c}}\cap S|\right)\log p\right).
\end{aligned}$$

After applying union bound, we get

$$\begin{aligned}
&\mathbb{P}\left(\max_{S\subset[p]:|S|\le 2C_0 s}\frac{1}{|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|}\sum_{j\in S^*\cap S^{\mathsf{c}}}\left(X_j^T P_S w\right)^2\ge 16c^2 s\log^2 p\right)\\
&\le\sum_{m=0}^{2C_0 s}\binom{p}{m}\left(\exp\left(-cs\log p\right)+\exp\left(-cm\log p\right)\right)\\
&\le\exp\left(-c'\log p\right).
\end{aligned}$$

Equation (G.8): For each $j\notin S^*$, we have $(X_j^T P_{S^*}w)^2$ stochastically dominated by $\xi\zeta$ where $\xi\sim\chi_s^2$, $\zeta\sim\chi_1^2$ and $\xi\perp\zeta$. We get the desired result by the χ^2 tail bound and a union bound over $j\notin S^*$.

Equation (G.9): By (G.11) to be proved later, it is sufficient to establish

$$\mathbb{P}\left(\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s}\frac{1}{|T|}\left\|X_T^T P_S w\right\|^2\ge cs\log p\right)\le\exp\left(-c'\log p\right).$$

Note that for any fixed S,T, we have $\left\|X_T^T P_S w\right\|^2$ stochastically dominated by $\xi\zeta$ where $\xi\sim\chi_{2C_0 s}^2$, $\zeta\sim\chi_{|T|}^2$ and $\xi\perp\zeta$. Then we have $\mathbb{P}\left(\left\|X_T^T P_S w\right\|^2\ge c^2 s\,|T|\log p\right)\le\mathbb{P}\left(\zeta\ge c\,|T|\right)+\mathbb{P}\left(\xi\ge cs\log p\right)$ which can be controlled by the χ^2 tail bound. A union bound is then sufficient to complete the proof.

Equation (G.10): For any fixed S,T, and any $\theta\in\mathbb{R}^{|T|}$ such that $\|\theta\|=1$, we have

$$\theta^T\left(X_T^T(I-P_S)X_T\right)\theta\sim\chi_{n-|S|}^2,$$

and $\theta^T(n-|S|)I_{|T|}\theta=(n-|S|)$. By a standard ϵ-net argument [16], the χ^2 tail bound, and a union bound over S,T, we conclude its proof.

Equation (G.11): Its proof is similar to that of (G.2). We can show

$$\mathbb{P}\left(\max_{S,T\subset[p]:S\cap T=\emptyset,|S|,|T|\le 2C_0 s}\left\|\left(X_T^T(I-P_S)X_T\right)^{-1}\right\|\ge\frac{1}{cn}\right)\le\exp\left(-c'n\right)$$

for some c,c'. Its proof is omitted here.

Equation (G.12): We have $\left\|X_S^T w\right\|^2\overset{d}{=}\xi\zeta$ where $\xi\sim\chi_n^2$, $\zeta\sim\chi_{|S|}^2$ and $\xi\perp\zeta$. Thus,

$$\mathbb{P}\left(\left\|X_S^T w\right\|^2\ge c^2 n\,|S|\log p\right)\le\mathbb{P}\left(\xi\ge cn\right)+\mathbb{P}\left(\zeta\ge c\,|S|\log p\right).$$

A union bound over integers $0\le m\le 2C_0 s$ and over all sets $\{S\subset[p]:|S|=m\}$ leads to the desired result.

Equation (G.13): For a fixed pair S,T, we have $\|\beta^*_{S^{\mathsf{c}}\cap S^*}\|^{-1}X_{S^{\mathsf{c}}\cap S^*}\beta^*_{S^{\mathsf{c}}\cap S^*}\sim\mathcal{N}(0,I_n)$, and consequently $\|\beta^*_{S^{\mathsf{c}}\cap S^*}\|^{-2}\left\|X_T^T(I-P_S)X_{S^{\mathsf{c}}\cap S^*}\beta^*_{S^{\mathsf{c}}\cap S^*}\right\|$ is stochastically dominated by $\xi\zeta$ where $\xi\sim\chi_n^2$, $\zeta\sim\chi_{|T|}^2$ and $\xi\perp\zeta$. Note that ξ only depends on $S^{\mathsf{c}}\cap S^*$ and ζ only depends on T. For a fixed S, in order to take a union bound over T, we add a subscript to ζ as in ζ_T to make the dependence explicit. We have

$$\begin{aligned}
&\mathbb{P}\left(\max_{T\subset[p]} \frac{1}{|T|\vee s} \|\beta^*_{S^{\mathsf{c}}\cap S^*}\|^{-2} \left\|X_T^T(I-P_S)X_{S^{\mathsf{c}}\cap S^*}\beta^*_{S^{\mathsf{c}}\cap S^*}\right\|^2 \geq 16c^2 n\log p\right)\\
&\leq \mathbb{P}\left(\xi\geq 4cn\right)+\mathbb{P}\left(\max_{T\subset[p]}\frac{1}{|T|\vee s}\zeta_T\geq 4c\log p\right)\\
&\leq \exp\left(-cn\right)+\sum_{m=0}^{p}\sum_{T\subset[p]:|T|=m}\exp\left(-c\left(m\wedge s\right)\log p\right)\\
&\leq \exp\left(-c' s\log p\right).
\end{aligned}$$

The proof is completed by an additional union bound argument over S.

Equation (G.14): Consider a fixed pair S, T. For any $x\in\mathbb{R}^n$, we have $(P_{S^*}-P_S)\,x = P_{S,1}x - P_{S,2}x$, where $P_{S,1}$ is the projection matrix onto the space $\operatorname{span}(X_{S^*})\backslash(\operatorname{span}(X_{S^*})\cap\operatorname{span}(X_S))$, and $P_{S,2}$ is the projection matrix onto the space $\operatorname{span}(X_S)\setminus(\operatorname{span}(X_{S^*})\cap\operatorname{span}(X_S))$. Then we have

$$\left\|X_T^T(P_{S^*}-P_S)w\right\|^2 = \left\|X_T^T P_{S,1}w - X_T^T P_{S,2}w\right\|^2 \leq 2\left(\left\|X_T^T P_{S,1}w\right\|^2+\left\|X_T^T P_{S,2}w\right\|^2\right).$$

Note that $\operatorname{span}(X_{S^*\cap S})\subset\operatorname{span}(X_{S^*})\cap\operatorname{span}(X_S)$, and thus the rank of $P_{S,1}$ is bounded by $|S^*\cap S^{\mathsf{c}}|$. Hence, $\left\|X_T^T P_{S,1}w\right\|^2$ is stochastically dominated by $\xi\zeta$ where $\xi\sim\chi^2_{|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|}$, $\zeta\sim\chi^2_{|T|}$ and $\xi\perp\zeta$. Note that ξ only depends on S and ζ only depends on T. For a fixed S, in order to take a union bound over T, we add a subscript to ζ as in ζ_T to make the dependence explicit. We have

$$\begin{aligned}
&\mathbb{P}\left(\max_{T\subset[p]:T\cap(S\cup S^*)=\emptyset}\frac{1}{|T|\vee s}\left\|X_T^T P_{S,1}w\right\|^2\geq 16c^2\left(|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|\right)\log^2 p\right)\\
&\leq \mathbb{P}\left(\xi\geq 4c\left(|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|\right)\log p\right)+\mathbb{P}\left(\max_{T\subset[p]}\frac{1}{|T|\vee s}\zeta_T\geq cs\log p\right)\\
&\leq \exp\left(-c\left(|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|\right)\log p\right)+\sum_{m=0}^{p}\sum_{T\subset[p]:|T|=m}\exp\left(-c\left(m\wedge s\right)\log p\right)\\
&\leq \exp\left(-c\left(|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|\right)\log p\right)+\exp\left(-c' s\log p\right).
\end{aligned}$$

Then we take a union bound of S.

$$\begin{aligned}
&\mathbb{P}\left(\max_{S\subset[p]:|S|\leq 2C_0 s}\max_{T\subset[p]:T\cap(S\cup S^*)=\emptyset}\frac{1}{|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|}\frac{1}{|T|\vee s}\left\|X_T^T P_{S,1}w\right\|^2\geq 16c^2\log^2 p\right)\\
&\leq \sum_{m'=0}^{2C_0 s}\sum_{S\subset[p]:|S^*\cap S^{\mathsf{c}}|+|S^{*\mathsf{c}}\cap S|=m'}\left(\exp\left(-cm'\log p\right)+\exp\left(-c' s\log p\right)\right)\\
&\leq \exp\left(-c''\log p\right).
\end{aligned}$$

A similar result holds for the term related to $P_{S,2}$. Putting them together, we complete the proof.

Equation (G.15): Define $B_j = \|X_j\|^{-1} X_j^T X_{S^*_{-j}} \left(X^T_{S^*_{-j}} (I - P_j) X_{S^*_{-j}} \right)^{-1} X^T_{S^*_{-j}} X_j \|X_j\|^{-1}$ for all $j \in S^*$. Note that $\|X_j\|^{-1} X_j^T X_{S^*_{-j}} \left(X^T_{S^*_{-j}} (I - P_j) X_{S^*_{-j}} \right)^{-1} X^T_{S^*_{-j}} (I - P_j) w$ is identically distributed by $\sqrt{B_j}\xi_j$ with $\xi_j \sim \mathcal{N}(0,1)$ and $\xi_j \perp B_j$. Here we have the subscript for both ξ_j and B_j to make their dependence on j explicit. Then,

$$\mathbb{P}\left(\|X_j\|^{-1} X_j^T X_{S^*_{-j}} \left(X^T_{S^*_{-j}} (I - P_j) X_{S^*_{-j}} \right)^{-1} X^T_{S^*_{-j}} (I - P_j) w \geq \sqrt{\frac{c^2 s \log^2 p}{n}} \right)$$
$$\leq \mathbb{P}\left(\xi_j \geq \sqrt{c \log p} \right) + \mathbb{P}\left(B_j \geq \frac{cs \log p}{n} \right),$$

and thus

$$\mathbb{P}\left(\max_{j \in S^*} \|X_j\|^{-1} X_j^T X_{S^*_{-j}} \left(X^T_{S^*_{-j}} (I - P_j) X_{S^*_{-j}} \right)^{-1} X^T_{S^*_{-j}} (I - P_j) w \geq \sqrt{\frac{c^2 s \log^2 p}{n}} \right)$$
$$\leq \mathbb{P}\left(\max_{j \in S^*} \xi_j \geq \sqrt{c \log p} \right) + \mathbb{P}\left(\max_{j \in S^*} B_j \geq \frac{cs \log p}{n} \right).$$

The first term can be easily bounded by $s \exp\left(-2^{-1} c \log p\right) \leq \exp\left(-c' \log p\right)$. For the second term, we have

$$B_j \leq \left\| \left(X^T_{S^*_{-j}} (I - P_j) X_{S^*_{-j}} \right)^{-1} \right\| \left\| X^T_{S^*_{-j}} X_j \|X_j\|^{-1} \right\|^2,$$

for all $j \in S^*$. By a similar analysis as in (G.11), we can show $\max_{j \in S^*} \left\| \left(X^T_{S^*_{-j}} (I - P_j) X_{S^*_{-j}} \right)^{-1} \right\| \leq c_1/n$ with probability at least $1 - \exp(-c_2 n)$. Note that $\left\| X^T_{S^*_{-j}} X_j \|X_j\|^{-1} \right\|^2 \sim \chi^2_{s-1}$. Easily we can show $\max_{j \in S^*} \left\| X^T_{S^*_{-j}} X_j \|X_j\|^{-1} \right\|^2 \leq 4 c_3 s \log p$ with probability at least $1 - \exp(-c_4 s \log p)$. As a result,

$$\mathbb{P}\left(\max_{j \in S^*} B_j \geq \frac{cs \log p}{n} \right) \leq \exp(-c_2 n) + \exp(-c_4 s \log p),$$

which completes the proof. □

Lemma G.2. *Define*

$$\widetilde{\psi}(n,p,s,\lambda,\delta,C) = s\mathbb{P}\left(\epsilon > (1-\delta)\|\zeta\|(\lambda - t(\zeta)) \ \& \ \left| \|\zeta\|^2 - n \right| \leq C\sqrt{n \log p} \right)$$
$$+ (p-s)\mathbb{P}\left(\epsilon > (1-\delta)\|\zeta\| t(\zeta) \ \& \ \left| \|\zeta\|^2 - n \right| \leq C\sqrt{n \log p} \right), \quad \text{(G.16)}$$

where $\epsilon \sim \mathcal{N}(0,1)$, $\zeta \sim \mathcal{N}(0, I_n)$, and they are independent of each other. Assume $s \log p \leq n$, $\limsup s/p < 1/2$*, and $\mathsf{SNR} \to \infty$. For any $\delta \leq 1/\log p$ and any constant $C > 0$, we have*

$$\widetilde{\psi}(n,p,s,\lambda,\delta,C) = s \exp\left(-\frac{(1+o(1))\mathsf{SNR}^2}{2} \right).$$

Proof. Throughout the proof, we use $g(x)$ and $G(x)$ for the density and survival functions of $\mathcal{N}(0,1)$. A standard Gaussian tail analysis gives

$$\frac{1}{2x}g(x) \le G(x) \le \frac{1}{x}g(x), \tag{G.17}$$

for all $x \ge 2$. With a slight abuse of notation, we also use the notation

$$t(u) = \frac{\lambda}{2} + \frac{\log\frac{p-s}{s}}{\lambda u^2}, \tag{G.18}$$

for all $u > 0$. We first focus on deriving an upper bound for $\widetilde{\psi}(n,p,s,\lambda,\delta,C)$. For any $u > 0$, we define

$$m(u) = u\left(\lambda - \frac{\log\frac{p-s}{s}}{\lambda u^2}\right) = \lambda u - \frac{\log\frac{p-s}{s}}{\lambda u}.$$

Recall the definition of $t(u)$ in (G.18). Define $u_{\min} = \sqrt{n - C\sqrt{n\log p}}$ and $u_{\max} = \sqrt{n + C\sqrt{n\log p}}$, and $U = [u_{\min}, u_{\max}]$. Since $u\left(\lambda - t(u)\right)$ is an increasing function of $u > 0$. We have

$$m(u_{\min}) \le u\left(\lambda - t(u)\right) \le m(u_{\max}),$$

for all $u \in U$. This gives

$$\begin{aligned}
&s\mathbb{P}\left(\frac{\epsilon}{1-\delta} \ge \|\zeta\|\left(\lambda - t(\zeta)\right)\ \&\ \left|\|\zeta\| - n\right| \le C\sqrt{n\log p}\right)\\
&\le s\mathbb{P}\left(\frac{\epsilon}{1-\delta} \ge m(u_{\min})\right)\\
&\le \frac{s}{\sqrt{2\pi}\left(1-\delta\right)m(u_{\min})}\exp\left(-\frac{1}{2}\left(1-\delta\right)^2 m^2(u_{\min})\right),
\end{aligned}$$

where the last inequality is by (G.17). In addition, we have the identity $\left(ut(u)\right)^2 = 2\log\frac{p-s}{s} + m^2(u)$. This leads to

$$2\log\frac{p-s}{s} + m^2(u_{\min}) \le \left(ut(u)\right)^2 \le 2\log\frac{p-s}{s} + m^2(u_{\max}),$$

for all $u \in U$. As a result, using (G.17), we have

$$\begin{aligned}
&(p-s)\,\mathbb{P}\left(\frac{\epsilon}{1-\delta} \ge \|\zeta\|\,t(\zeta)\ \&\ \left|\|\zeta\| - n\right| \le C\sqrt{n\log p}\right)\\
&= (p-s)\,\mathbb{E}_{u^2\sim\chi^2_n}\left[\mathbb{P}\left(\frac{\epsilon}{1-\delta} \ge |u|t(u)\middle|u\right)\mathbf{1}_{\left\{|u^2-n|\le C\sqrt{n\log p}\right\}}\right]\\
&\le \frac{p-s}{\sqrt{2\pi}}\mathbb{E}_{u^2\sim\chi^2_n}\left[\frac{1}{ut(u)}\exp\left(-\frac{1}{2}\left(1-\delta\right)^2\left(ut(u)\right)^2\right)\mathbf{1}_{\left\{|u^2-n|\le C\sqrt{n\log p}\right\}}\right]\\
&\le \frac{p-s}{\sqrt{2\pi}}\mathbb{E}_{u^2\sim\chi^2_n}\left[\frac{1}{\min_{u\in U}ut(u)}\exp\left(-\frac{1}{2}\left(1-\delta\right)^2\left(2\log\frac{p-s}{s} + m^2(u_{\min})\right)\right)\mathbf{1}_{\left\{|u^2-n|\le C\sqrt{n\log p}\right\}}\right]\\
&\le \frac{p-s}{\sqrt{2\pi}\min_{u\in U}ut(u)}\exp\left(-\left(1-\delta\right)^2\log\frac{p-s}{s} - \frac{1}{2}\left(1-\delta\right)^2 m^2(u_{\min})\right)\\
&= \frac{s}{\sqrt{2\pi}\min_{u\in U}ut(u)}\exp\left(\left(2\delta-\delta^2\right)\log\frac{p-s}{s} - \frac{1}{2}\left(1-\delta\right)^2 m^2(u_{\min})\right).
\end{aligned}$$

Combining the above results together, we have

$$\widetilde{\psi}(n,p,s,\lambda,\delta,C) \leq \frac{s}{\sqrt{2\pi}\,(1-\delta)\,m(u_{\min})}\exp\left(-\frac{1}{2}(1-\delta)^2 m^2(u_{\min})\right) + \frac{s}{\sqrt{2\pi}\min_{u\in U} ut(u)}\exp\left(\left(2\delta-\delta^2\right)\log\frac{p-s}{s} - \frac{1}{2}(1-\delta)^2 m^2(u_{\max})\right).$$

Now we derivative a lower bound for $\widetilde{\psi}(n,p,s,\lambda,\delta,C)$. Note that $\mathbb{P}\left(|\|\zeta\|-n| \leq C\sqrt{n\log p}\right) \geq 1/2$. We therefore have

$$\begin{aligned} & s\mathbb{P}\left(\epsilon \geq \|\zeta\|\left(\lambda - t(\zeta)\right) \ \&\ |\|\zeta\|-n| \leq C\sqrt{n\log p}\right) \\ & \geq \frac{s}{2}\mathbb{P}\left(\epsilon \geq m(u_{\max})\right) \\ & \geq \frac{s}{4\sqrt{2\pi}m(u_{\max})}\exp\left(-\frac{1}{2}m^2(u_{\max})\right), \end{aligned}$$

and

$$\begin{aligned} & (p-s)\,\mathbb{P}\left(\epsilon \geq \|\zeta\|\,t(\zeta)\ \&\ |\|\zeta\|-n| \leq C\sqrt{n\log p}\right) \\ & \geq \frac{s}{4\sqrt{2\pi}\max_{u\in U} ut(u)}\exp\left(\left(2\delta-\delta^2\right)\log\frac{p-s}{s} - \frac{1}{2}(1-\delta)^2 m^2(u_{\max})\right). \end{aligned}$$

Consequently,

$$\widetilde{\psi}(n,p,s,\lambda,\delta,C) \geq \frac{s}{4\sqrt{2\pi}m(u_{\max})}\exp\left(-\frac{1}{2}m^2(u_{\max})\right) + \frac{s}{4\sqrt{2\pi}\max_{u\in U} ut(u)}\exp\left(\left(2\delta-\delta^2\right)\log\frac{p-s}{s} - \frac{1}{2}(1-\delta)^2 m^2(u_{\max})\right).$$

Since $\delta \leq 1/\log p$ and $\mathsf{SNR} \to \infty$, with the same arguments used in the proof of Lemma G.3, we can show for all $u \in U$, we have $\frac{\lambda u}{2} - \frac{\log\frac{p-s}{s}}{\lambda u} = (1+o(1))\,\mathsf{SNR}$ and $\frac{\lambda u}{2} + \frac{\log\frac{p-s}{s}}{\lambda u} \to \infty$. This leads to $\widetilde{\psi}(n,p,s,\lambda,\delta,C) = s\exp\left(-\frac{(1+o(1))\mathsf{SNR}^2}{2}\right)$ as desired, which completes the proof. □

Lemma G.3. *Consider some $\beta^* \in \mathbb{R}^p$ that satisfies either $|\beta_j^*| \geq \lambda$ or $\beta_j^* = 0$ for all $j \in [p]$. Assume $\limsup s/p < \frac{1}{2}$ and $\mathsf{SNR} \to \infty$. Then, for i.i.d. $X_1, \cdots, X_p \sim \mathcal{N}(0, I_n)$, we have*

$$\begin{aligned} \min_{j\in[p]} \sqrt{n}\left||\beta_j^*| - t(X_j)\right| &> 1, \\ \max_{j\in[p]} \frac{|\beta_j^*|}{\left||\beta_j^*| - t(X_j)\right|} &\leq \sqrt{\log p}, \end{aligned}$$

with probability at least $1 - e^{-p}$.

Proof. We first show that under the assumption that $\limsup s/p < 1/2$, the condition $\mathsf{SNR} \to \infty$ is equivalent to

$$\frac{n\lambda^2 - 2\log\frac{p-s}{s}}{\sqrt{\log\frac{p-s}{s}}} \to \infty. \tag{G.19}$$

A direct calculation gives

$$\left(\frac{\lambda\sqrt{n}}{2} - \frac{\log\frac{p-s}{s}}{\lambda\sqrt{n}}\right)^2 = \frac{\left(n\lambda^2 - 2\log\frac{p-s}{s}\right)^2}{4n\lambda^2} = \left(\frac{n\lambda^2 - 2\log\frac{p-s}{s}}{\sqrt{\log\frac{p-s}{s}}}\right)^2 \frac{\log\frac{p-s}{s}}{4n\lambda^2}.$$

If $\mathsf{SNR} \to \infty$ holds, we have $n\lambda^2 \geq 2\log\frac{p-s}{s}$, which leads to (G.19). For the other direction, if (G.19) holds, there exists some $A \to \infty$ such that $n\lambda^2 = 2\log\frac{p-s}{s} + A\sqrt{\log\frac{p-s}{s}}$. By the above identity, we have

$$\left(\frac{\lambda\sqrt{n}}{2} - \frac{\log\frac{p-s}{s}}{\lambda\sqrt{n}}\right)^2 = A^2\frac{\log\frac{p-s}{s}}{2\left(2\log\frac{p-s}{s} + A\sqrt{\log\frac{p-s}{s}}\right)} \to \infty.$$

Thus we have shown that $\mathsf{SNR} \to \infty$ and (G.19) are equivalent.

Now we are going to prove the proposition under the high-probability event (G.1). Note that for any $j \in [p]$ such that $\beta_j^* = 0$, we have $\sqrt{n}|\beta_j^*| - t(X_j)| = \sqrt{n}t(X_j) \geq \sqrt{n}\lambda/2 \to \infty$ by (G.19) and $|\beta_j^*|/||\beta_j^*| - t(X_j)| = 0$. Thus, we only need to consider the remaining $j \in [p]$ such that $\beta_j^* \neq 0$. It is sufficient to prove $\min_{j\in[p]:z_j^*\neq 0}\sqrt{n}\,(\lambda - t(X_j)) > 1$ and $\max_{j\in[p]:z_j^*\neq 0}\frac{\lambda}{\lambda - t(X_j)} \leq \sqrt{\log p}$. Consider any $j \in [p]$ such that $z_j^* \neq 0$. We have

$$\sqrt{n}\,(\lambda - t(X_j)) = \frac{\sqrt{n}}{2\lambda}\left(\lambda^2 - 2\frac{\log\frac{p-s}{s}}{\|X_j\|^2}\right) \geq \frac{\sqrt{n}}{2\lambda}\left(\lambda^2 - 2\frac{\log\frac{p-s}{s}}{n - C\sqrt{n\log p}}\right),$$

where the last inequality is by (G.1). By (G.19), there exists an $A \to \infty$, such that

$$n\lambda^2 = 2\log\frac{p-s}{s} + A\sqrt{\log\frac{p-s}{s}}.$$

Then, we have

$$\begin{aligned}
\sqrt{n}\,(\lambda - t(X_j)) &\geq \frac{1}{2\sqrt{n}\lambda}\left(2\log\frac{p-s}{s} + A\sqrt{\log\frac{p-s}{s}} - \frac{2\log\frac{p-s}{s}}{1 - C\sqrt{\frac{\log p}{n}}}\right) \\
&\geq \frac{1}{2\sqrt{n}\lambda}\left(A\sqrt{\log\frac{p-s}{s}} - C'\sqrt{\frac{\log p}{n}}\log\frac{p-s}{s}\right) \\
&\geq \frac{C''A\sqrt{\log\frac{p-s}{s}}}{\sqrt{n}\lambda},
\end{aligned}$$

for some constants $C', C'' > 0$. Starting from here, first we have

$$\sqrt{n}\,(\lambda - t(X_j)) \geq C'' \sqrt{\frac{A^2 \log \frac{p-s}{s}}{n\lambda^2}} = C'' \sqrt{\frac{A^2 \log \frac{p-s}{s}}{2\log \frac{p-s}{s} + A\sqrt{\log \frac{p-s}{s}}}} \to \infty,$$

as $A \to \infty$. Second, we have

$$\begin{aligned} \frac{\lambda}{\lambda - t(X_j)} &= \frac{\sqrt{n}\lambda}{\sqrt{n}\,(\lambda - t(X_j))} \leq \frac{\sqrt{n}\lambda}{\frac{C'' A\sqrt{\log \frac{p-s}{s}}}{\sqrt{n}\lambda}} = \frac{n\lambda^2}{C'' A\sqrt{\log \frac{p-s}{s}}} \\ &= \frac{2\log \frac{p-s}{s} + A\sqrt{\log \frac{p-s}{s}}}{C'' A\sqrt{\log \frac{p-s}{s}}} \leq o\left(\sqrt{\log \frac{p-s}{s}}\right). \end{aligned}$$

Hence, the proof is complete. □

Now we are ready to prove Theorem 5.1, Theorem 5.2, and Corollary 5.1.

Proof of Theorem 5.1. By [12], we have

$$\inf_{\widehat{z}} \sup_{z^* \in \mathcal{Z}_s} \sup_{\beta^* \in \mathcal{B}_{z^*,\lambda}} \mathbb{E}\mathsf{H}_{(s)}(\widehat{z}, z^*) \geq \frac{1}{2s}\psi(n,p,s,\lambda,0) - 4e^{-s/8},$$

where $\psi(n,p,s,\lambda,0)$ is defined in (50). By Lemma G.2,

$$\frac{1}{2s}\psi(n,p,s,\lambda,0) \geq \frac{1}{2s}\widetilde{\psi}(n,p,s,\lambda,0,C) = \exp\left(-\frac{(1+o(1))\mathsf{SNR}^2}{2}\right),$$

and we obtain the desired conclusion. □

Proof of Theorem 5.2. The condition of Theorem 5.2 allows us to apply Lemma G.3 to the conclusion of Lemma 5.1. This implies that the right hand sides of (47) and (48) can be bounded by $o((\log p)^{-1})$, which then implies Conditions A-C hold with some $\delta = o((\log p)^{-1})$. Then, the desired conclusion is a special case of Theorem 3.1. □

Proof of Corollary 5.1. By (45) and (G.1), we have

$$\mathsf{H}_{(s)}(z, z^*) \leq \frac{\ell(z,z^*)}{s\lambda^2 \min_{j\in[p]} \|X_j\|^2} \leq \frac{2\ell(z,z^*)}{sn\lambda^2},$$

with high probability. Then, the conclusion is a direct consequence of Theorem 5.2. □

Finally, we present the proofs of Lemma 5.1, Lemma 5.2, and Proposition 5.1.

Proof of Lemma 5.1. The proof will be established under the high-probability events (G.1)-(G.15). First we present a few important quantities closely related to $\ell(z,z^*)$. By $h(z,z^*) \leq \frac{\ell(z,z^*)}{\lambda^2 \min_j \|X_j\|^2}$ and (G.1), we have

$$h(z,z^*) \leq \frac{2\ell(z,z^*)}{n\lambda^2}. \tag{G.20}$$

By the definition of $\ell(z, z^*)$, it is obvious

$$\sum_{j=1}^{p} \beta_j^{*2} \mathbf{1}_{\{|z_j| \neq |z_j^*|\}} \leq \frac{\ell(z, z^*)}{\min_j \|X_j\|^2} \leq \frac{2\ell(z, z^*)}{n}, \tag{G.21}$$

where we have used (G.1) again. For any $z \in \{0, 1, -1\}^p$ such that $\ell(z, z^*) \leq \tau \leq C_0 s n \lambda^2$, (G.20) implies

$$h(z, z^*) \leq 2C_0 s. \tag{G.22}$$

We will first prove the easier conclusion (48) and then prove (47).

Proof of (48). According to the definition, we have

$$\begin{aligned}
\left| \frac{1}{\|X_j\|} \sum_{l \in [p] \setminus \{j\}} \left(\widehat{\beta}_l(z^*) - \beta^* \right) X_j^T X_l \right| &= \left| \frac{1}{\|X_j\|} X_j^T X \widehat{\beta}(z^*) - \frac{1}{\|X_j\|} X_j^T X \beta^* - \|X_j\| \left(\widehat{\beta}_j(z^*) - \beta_j^* \right) \right| \\
&= \left| \frac{1}{\|X_j\|} X_j^T \left(X \widehat{\beta}(z^*) - X \beta^* \right) - \|X_j\| \left(\widehat{\beta}_j(z^*) - \beta_j^* \right) \right|.
\end{aligned}$$

By the fact that $\widehat{\beta}_{S^*}(z^*) = \beta_{S^*} + \left(X_{S^*}^T X_{S^*} \right)^{-1} X_{S^*}^T w$ and $\widehat{\beta}_{S^{*c}}(z^*) = 0$, we have

$$\left| \frac{1}{\|X_j\|} \sum_{l \in [p] \setminus \{j\}} \left(\widehat{\beta}_l(z^*) - \beta^* \right) X_j^T X_l \right| = \begin{cases} \frac{1}{\|X_j\|} X_j^T w - \|X_j\| \left[\left(X_{S^*}^T X_{S^*} \right)^{-1} X_{S^*}^T w \right]_{\phi_{S^*}(j)} & j \in S^* \\ \frac{1}{\|X_j\|} X_j^T P_{S^*} w & j \notin S^*. \end{cases} \tag{G.23}$$

- We first consider $j \notin S^*$. By (G.1), (G.8) and (G.23), we have

$$\max_{j \notin S^*} \left| \frac{1}{\|X_j\|} \sum_{l \in [p] \setminus \{j\}} \left(\widehat{\beta}_l(z^*) - \beta^* \right) X_j^T X_l \right| \leq \sqrt{\frac{2Cs \log p}{n}}.$$

- Next, we consider $j \in S^*$. Writing X_{S^*} into a block matrix form $X_{S^*} = (X_j, X_{S^*_{-j}})$, we have a block matrix inverse formula

$$\left(X_{S^*}^T X_{S^*} \right)^{-1} = \begin{pmatrix} \|X_j\|^2 & X_j^T X_{S^*_{-j}} \\ X_{S^*_{-j}}^T X_j & X_{S^*_{-j}}^T X_{S^*_{-j}} \end{pmatrix}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \tag{G.24}$$

with

$$\begin{aligned}
B_{11} &= \|X_j\|^{-2} + \|X_j\|^{-2} X_j^T X_{S^*_{-j}} \left(X_{S^*_{-j}}^T (I - P_j) X_{S^*_{-j}} \right)^{-1} X_{S^*_{-j}}^T X_j \|X_j\|^{-2}, \\
B_{12} &= - \|X_j\|^{-2} X_j^T X_{S^*_{-j}} \left(X_{S^*_{-j}}^T (I - P_j) X_{S^*_{-j}} \right)^{-1},
\end{aligned}$$

and the explicit expressions of B_{21}, B_{22} not displayed since they are irrelevant to our proof. We have $[\left(X_{S^*}^T X_{S^*}\right)^{-1} X_{S^*}^T w]_{\phi_{S^*}(j)} = B_{11} X_j^T w + B_{12} X_{S^*_{-j}}^T w$. From (G.23), some algebra leads to

$$
\begin{aligned}
&\max_{j\in S^*}\left|\frac{1}{\|X_j\|}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z^*)-\beta^*\right)X_j^T X_l\right| \\
&= \max_{j\in S^*}\left|\frac{1}{\|X_j\|}X_j^T w - \|X_j\|\left[\left(X_{S^*}^T X_{S^*}\right)^{-1} X_{S^*}^T w\right]_{\phi_{S^*}(j)}\right| \\
&= \max_{j\in S^*}\left|\|X_j\|^{-1} X_j^T X_{S^*_{-j}}\left(X_{S^*_{-j}}^T\left(I-P_j\right)X_{S^*_{-j}}\right)^{-1} X_{S^*_{-j}}^T\left(I-P_j\right)w\right| \\
&\le \sqrt{\frac{Cs\log^2 p}{n}},
\end{aligned}
\tag{G.25}
$$

where the last inequality is by (G.15).

Combining the two cases, we have

$$
\max_{j\in[p]}\left|\frac{1}{\|X_j\|}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z^*)-\beta^*\right)X_j^T X_l\right| \le \sqrt{\frac{2Cs\log^2 p}{n}}.
$$

Using (46), we get $\left|\Delta_j(z_j^*, b)^2\right| \ge 4t(X_j)\left|\left|\beta_j^*\right| - t(X_j)\right|\|X_j\|^2$ for all $j\in[p]$. Then,

$$
\begin{aligned}
\max_{j\in[p]}\left|\frac{H_j(z_j^*, b)}{\Delta_j(z_j^*, b)^2}\right| &\le \max_{j\in[p]}\frac{4\|X_j\|\,t(X_j)\left|\|X_j\|^{-1}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z^*)-\beta^*\right)X_j^T X_l\right|}{4t(X_j)\left|\left|\beta_j^*\right|-t(X_j)\right|\|X_j\|^2} \\
&\le \max_{j\in[p]}\frac{\left|\|X_j\|^{-1}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z^*)-\beta^*\right)X_j^T X_l\right|}{\left|\left|\beta_j^*\right|-t(X_j)\right|\|X_j\|} \\
&\le \sqrt{\frac{2Cs\log^2 p}{n}}\frac{1}{\min_{j\in[p]}\left|\left|\beta_j^*\right|-t(X_j)\right|\|X_j\|} \\
&\le \sqrt{\frac{4Cs\log^2 p}{n}}\frac{1}{\min_{j\in[p]}\sqrt{n}\left|\left|\beta_j^*\right|-t(X_j)\right|},
\end{aligned}
$$

where in the last inequality we use (G.1).

Proof of (47). By (44) and (46), we have

$$\frac{G_j(z_j^*, b; z)^2 \|\mu_j(B^*, b) - \mu_j(B^*, z_j^*)\|^2}{\Delta_j(z_j^*, b)^4 \ell(z, z^*)}$$

$$\leq \frac{\left(4\|X_j\| t(X_j)\left(\|X_j\|^{-1}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z) - \widehat{\beta}_l(z^*)\right)X_j^T X_l\right)\right)^2 \left(4\left|\beta_j^*\right|^2 \|X_j\|^2 \mathbf{1}_{\{z_j^*=\pm 1\}} + \lambda^2 \|X_j\|^2 \mathbf{1}_{\{z_j^*=0\}}\right)}{\left(\left(4t(X_j)\left(\left|\beta_j^*\right| - t(X_j)\right)\|X_j\|^2\right)^2 \mathbf{1}_{\{z_j^*=\pm 1\}} + \left(4t(X_j)^2\|X_j\|^2\right)^2 \mathbf{1}_{\{z_j^*=0\}}\right)\ell(z, z^*)}$$

$$= \frac{\left(\|X_j\|^{-1}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z) - \widehat{\beta}_l(z^*)\right)X_j^T X_l\right)^2}{\ell(z, z^*)}\left(4\left(\frac{\left|\beta_j^*\right|}{\left|\beta_j^*\right| - t(X_j)}\right)^2 \mathbf{1}_{\{z_j^*=\pm 1\}} + \left(\frac{\lambda}{t(X_j)}\right)^2 \mathbf{1}_{\{z_j^*=0\}}\right)$$

$$\leq \frac{\left(\|X_j\|^{-1}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z) - \widehat{\beta}_l(z^*)\right)X_j^T X_l\right)^2}{\ell(z, z^*)}\max\left\{4\left(\frac{\left|\beta_j^*\right|}{\left|\beta_j^*\right| - t(X_j)}\right)^2, \left(\frac{\lambda}{t(X_j)}\right)^2\right\}.$$

Define

$$\widetilde{\beta}_j(z) = \|X_j\|^{-2} X_j^T \left(Y - \sum_{l\in[p]\setminus\{j\}} X_l \widehat{\beta}_l(z)\right).$$

We then have

$$\|X_j\|^{-1}\sum_{l\in[p]\setminus\{j\}}\left(\widehat{\beta}_l(z) - \widehat{\beta}_l(z^*)\right)X_j^T X_l$$
$$= -\|X_j\|\left(\|X_j\|^{-2}X_j^T\left(Y - \sum_{l\in[p]\setminus\{j\}} X_l\widehat{\beta}_l(z)\right) - \|X_j\|^{-2}X_j^T\left(Y - \sum_{l\in[p]\setminus\{j\}} X_l\widehat{\beta}_l(z^*)\right)\right)$$
$$= -\|X_j\|\left(\widetilde{\beta}_j(z) - \widetilde{\beta}_j(z^*)\right).$$

Hence, we have

$$\max_{b\in\{-1,0,1\}\setminus\{z_j^*\}} \frac{G_j(z_j^*, b; z)^2 \|\mu_j(B^*, b) - \mu_j(B^*, z_j^*)\|^2}{\Delta_j(z_j^*, b)^4 \ell(z, z^*)}$$
$$\leq \max\left\{4\left(\frac{\left|\beta_j^*\right|}{\left|\beta_j^*\right| - t(X_j)}\right)^2, \left(\frac{\lambda}{t(X_j)}\right)^2\right\}\frac{\|X_j\|^2\left(\widetilde{\beta}_j(z) - \widetilde{\beta}_j(z^*)\right)^2}{\ell(z, z^*)}$$
$$\leq \max\left\{4\left(\frac{\left|\beta_j^*\right|}{\left|\beta_j^*\right| - t(X_j)}\right)^2, \left(\frac{\lambda}{t(X_j)}\right)^2\right\}\frac{2n\left(\widetilde{\beta}_j(z) - \widetilde{\beta}_j(z^*)\right)^2}{\ell(z, z^*)},$$

where the last inequality is by (G.1). Therefore,

$$\max_{T\subset[p]}\frac{1}{s+|T|}\sum_{j\in T}\max_{b\in\{-1,0,1\}\backslash\{z_j^*\}}\frac{G_j(z_j^*,b;z)^2\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)}$$
$$\leq\max_{j\in[p]}\max\left\{4\left(\frac{\left|\beta_j^*\right|}{\left|\beta_j^*\right|-t(X_j)}\right)^2,\left(\frac{\lambda}{t(X_j)}\right)^2\right\}\frac{2n}{\ell(z,z^*)}\max_{T\subset[p]}\frac{1}{s+|T|}\sum_{j\in T}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2, \tag{G.26}$$

which is all about studying $\max_{T\subset[p]}\frac{1}{s+|T|}\sum_{j\in T}(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*))^2$.

In the following, we will focus on understanding $\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)$ for different $j\in[p]$. Define

$$S^*=\left\{j\in[p]:z_j^*\neq 0\right\},\quad\text{and}\quad S(z)=\{j\in[p]:z_j\neq 0\}.$$

For simplicity of notation, we just write S instead of $S(z)$ from now on. Recall that $\widehat{\beta}(z)$ is the least square estimator on the support S. That is,

$$\widehat{\beta}_S(z)=\left(X_S^TX_S\right)^TX_S^Ty,\quad\text{and}\quad\widehat{\beta}_{S^c}(z)=0.$$

Thus, the explicit expression of $\widetilde{\beta}_j(z)$ is given by

$$\widetilde{\beta}_j(z)=\begin{cases}\beta_j^*+\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta_{S_1}^*\right]_{\phi_S(j)}+\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)} & j\in S\\ \frac{1}{\|X_j\|^2}X_j^T\left[(I-P_S)X_{S_1}\beta_{S_1}^*+(I-P_S)w\right] & j\notin S.\end{cases}$$

Similarly, we also have

$$\widetilde{\beta}_j(z^*)=\begin{cases}\beta_j^*+\left[\left(X_{S^*}^TX_{S^*}\right)^{-1}X_{S^*}^Tw\right]_{\phi_{S^*}(j)} & j\in S^*\\ \frac{1}{\|X_j\|^2}X_j^T(I-P_{S^*})w & j\notin S^*.\end{cases}$$

The analysis of $\widetilde{\beta}_j(z^*)-\widetilde{\beta}_j(z)$ will be studied in four different regimes. We divide $[p]$ into four disjoint sets,

$$S_1=S^*\cap S^{\mathsf{c}},\quad S_2=S^*\cap S,\quad S_3=S^{*\mathsf{c}}\cap S,\quad\text{and}\quad S_4=S^{*\mathsf{c}}\cap S^{\mathsf{c}}.$$

Note that by (G.22), we have

$$|S_1|+|S_3|=h(z,z^*)\leq 2C_0s. \tag{G.27}$$

We denote $\mathsf{X}_l=X_{S_l},l=1,2,3,4$ for simplicity. We also denote $\mathsf{P}_l=\mathsf{X}_l\left(\mathsf{X}_l^T\mathsf{X}_l\right)^{-1}\mathsf{X}_l^T$ to be the projection matrix onto the subspace span(X_l), for $l=1,2,3,4$.

(1) Regime $j\in S_1$. In this case, we have $\widetilde{\beta}_j(z^*)=\beta_j^*+\left[\left(X_{S^*}^TX_{S^*}\right)^{-1}X_{S^*}^Tw\right]_{\phi_{S^*}(j)}$. We can

also write

$$\begin{aligned}
\widetilde{\beta}_j(z) &= \frac{1}{\|X_j\|^2} X_j^T \left[(I-P_S)X_{S_1}\beta^*_{S_1} + (I-P_S)w\right] \\
&= \frac{1}{\|X_j\|^2} X_j^T (I-P_S) X_j \beta^*_j + \sum_{l\in S_1, l\neq j} \frac{1}{\|X_j\|^2} X_j^T (I-P_S) X_l \beta^*_l + \frac{1}{\|X_j\|^2} X_j^T (I-P_S) w \\
&= \beta^*_j - \frac{1}{\|X_j\|^2} X_j^T P_S X_j \beta^*_j + \sum_{l\in S_1, l\neq j} \frac{1}{\|X_j\|^2} X_j^T (I-P_S) X_l \beta^*_l + \frac{1}{\|X_j\|^2} X_j^T (I-P_S) w.
\end{aligned}$$

This leads to the decomposition

$$\begin{aligned}
\widetilde{\beta}_j(z) - \widetilde{\beta}_j(z^*) = &- \frac{1}{\|X_j\|^2} X_j^T P_S X_j \beta^*_j + \sum_{l\in S_1, l\neq j} \frac{1}{\|X_j\|^2} X_j^T (I-P_S) X_l \beta^*_l - \frac{1}{\|X_j\|^2} X_j^T P_S w \\
&+ \left(\frac{1}{\|X_j\|^2} X_j^T w - \left[\left(X_{S^*}^T X_{S^*}\right)^{-1} X_{S^*}^T w \right]_{\phi_{S^*}(j)} \right).
\end{aligned}$$

We will bound each term on the right hand side of the above equation.

(1.1). First, we have

$$\begin{aligned}
\left| - \frac{1}{\|X_j\|^2} X_j^T P_S X_j \beta^*_j \right| &\leq \frac{2}{\min_j \|X_j\|^2} \max_{j\in S_1} \left| X_j^T P_S X_j \right| \left| \beta^*_j \right| \\
&\leq \frac{2}{n} C s \left| \beta^*_j \right| \log p,
\end{aligned}$$

where the last inequality is by (G.1) and (G.5). Then

$$\sum_{j\in S_1} \left(- \frac{1}{\|X_j\|^2} X_j^T P_S X_j \beta^*_j \right)^2 \leq \frac{4C^2 s^2 \log^2 p}{n^2} \left\| \beta^*_{S_1} \right\|^2.$$

(1.2). For the second term, we have a matrix representation,

$$\begin{aligned}
&\sum_{j\in S_1} \left(\sum_{l\in S_1, l\neq j} \frac{1}{\|X_j\|^2} X_j^T (I-P_S) X_l \beta^*_l \right)^2 \\
&\leq \frac{1}{\min_j \|X_j\|^4} \left\| \left(X_{S_1}^T (I-P_S) X_{S_1} - \operatorname{diag}\left\{ X_{S_1}^T (I-P_S) X_{S_1} \right\} \right) \beta^*_{S_1} \right\|^2 \\
&\leq \frac{1}{\min_j \|X_j\|^4} \left\| X_{S_1}^T (I-P_S) X_{S_1} - \operatorname{diag}\left\{ X_{S_1}^T (I-P_S) X_{S_1} \right\} \right\|^2 \left\| \beta^*_{S_1} \right\|^2 \\
&\leq \frac{2}{n^2} C s \log p \left\| \beta^*_{S_1} \right\|^2.
\end{aligned}$$

where the last inequality is by (G.1) and (G.6).

(1.3). For the third term, we have

$$\sum_{j\in S_1}\left(-\frac{1}{\|X_j\|^2}X_j^T P_S w\right)^2 \le \frac{1}{\min_j \|X_j\|^4}\sum_{j\in S_1}\left(X_j^T P_S w\right)^2$$
$$\le \frac{2}{n^2}Cs\log^2 p\,(|S_1|+|S_3|)$$
$$=\frac{2}{n^2}Csh(z,z^*)\log^2 p,$$

where the second to the last inequality is by (G.1) and (G.7).

(1.4). For the last term, using (G.1) and (G.25), we have

$$\max_{j\in S^*}\left|\frac{1}{\|X_j\|^2}X_j^T w-\left[\left(X_{S^*}^T X_{S^*}\right)^{-1}X_{S^*}^T w\right]_{\phi_{S^*}(j)}\right|$$
$$\le \max_{j\in S}\frac{1}{\|X_j\|}\max_{j\in S^*}\left|\frac{1}{\|X_j\|}X_j^T w-\|X_j\|\left[\left(X_{S^*}^T X_{S^*}\right)^{-1}X_{S^*}^T w\right]_{\phi_{S^*}(j)}\right|$$
$$\le\sqrt{\frac{2}{n}}\sqrt{\frac{Cs\log^2 p}{n}}.$$

Hence,

$$\sum_{j\in S_1}\left(\frac{1}{\|X_j\|^2}X_j^T w-\left[\left(X_{S^*}^T X_{S^*}\right)^{-1}X_{S^*}^T w\right]_{\phi_{S^*}(j)}\right)^2$$
$$\le |S_1|\max_{j\in S^*}\left(\frac{1}{\|X_j\|^2}X_j^T w-\left[\left(X_{S^*}^T X_{S^*}\right)^{-1}X_{S^*}^T w\right]_{\phi_{S^*}(j)}\right)^2$$
$$\le\frac{2Cs\,|S_1|\log^2 p}{n^2}$$
$$\le\frac{2Csh(z,z^*)\log^2 p}{n^2},$$

where the last inequality is due to (G.27).

(1.5). Combining the above results, we have

$$\sum_{j\in S_2}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2$$
$$\le 4\left(\frac{4C^2s^2\log^2 p}{n^2}\left\|\beta_{S_1}^*\right\|^2+\frac{2}{n^2}Cs\log p\left\|\beta_{S_1}^*\right\|^2+\frac{2}{n^2}Csh(z,z^*)\log^2 p+\frac{2Csh(z,z^*)\log^2 p}{n^2}\right)$$
$$\le 4\left(\frac{16C^2s^2\log^2 p}{n^2}+\frac{16s\log p}{\lambda^2 n^2}\right)\frac{\ell(z,z^*)}{n}$$

where the last inequality is by (G.20) and (G.21).

(2) Regime $j\in S_2$. In this case, $\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)$ can be written as

$$\left[\left(X_S^T X_S\right)^{-1}X_S^T X_{S_1}\beta_{S_1}^*\right]_{\phi_S(j)}+\left[\left(X_S^T X_S\right)^{-1}X_S^T w\right]_{\phi_S(j)}-\left[\left(X_{S^*}^T X_{S^*}\right)^{-1}X_{S^*}^T w\right]_{\phi_{S^*}(j)}.$$

We will bound the first term, and then the second and the third term will be analyzed together.

(2.1). For the first term, we have

$$\begin{aligned}\sum_{j\in S_2}\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta^*_{S_1}\right]^2_{\phi_S(j)} &\le \sum_{j\in S}\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta^*_{S_1}\right]^2_{\phi_S(j)}\\ &= \left\|\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta^*_{S_1}\right\|^2\\ &\le \left\|\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\right\|^2\left\|\beta^*_{S_1}\right\|^2\\ &\le C\frac{s\log p}{n}\left\|\beta^*_{S_1}\right\|^2,\end{aligned}$$

where the last inequality is due to (G.3) in Lemma G.1.

(2.2). Note that

$$\begin{aligned}&\sum_{j\in S_2}\left(\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}-\left[\left(X_{S^*}^TX_{S^*}\right)^{-1}X_{S^*}^Tw\right]_{\phi_{S^*}(j)}\right)^2\\ &=\left\|\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(S_2)}-\left[\left(X_{S^*}^TX_{S^*}\right)^{-1}X_{S^*}^Tw\right]_{\phi_{S^*}(S_2)}\right\|^2.\end{aligned}$$

Since S is close to S^*, the two length-$|S_2|$ vectors on the right hand side of the above equation should also be close to each other. Applying block matrix inverse formula, we have

$$\left(X_SX_S^T\right)^{-1}=\begin{pmatrix}\mathsf{X}_2^T\mathsf{X}_2 & \mathsf{X}_2^T\mathsf{X}_3\\ \mathsf{X}_3^T\mathsf{X}_2 & \mathsf{X}_3^T\mathsf{X}_3\end{pmatrix}^{-1}\triangleq\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}, \tag{G.28}$$

where

$$\begin{aligned}A_{11}&=\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}+\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_3\left(\mathsf{X}_3^T\left(I-\mathsf{P}_2\right)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{X}_2\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1},\\ A_{12}&=-\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_3\left(\mathsf{X}_3^T\left(I-\mathsf{P}_2\right)\mathsf{X}_3\right)^{-1},\\ A_{21}&=-\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{X}_2\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1},\\ A_{22}&=\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}.\end{aligned}$$

With these notation, we have

$$\begin{aligned}\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(S_2)}&=A_{11}\mathsf{X}_2^Tw+A_{12}\mathsf{X}_3^Tw\\ &=\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^Tw-\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_3\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T(I-\mathsf{P}_2)w,\end{aligned}$$

and

$$\left[\left(X_{S^*}^TX_{S^*}\right)^{-1}X_{S^*}^Tw\right]_{\phi_{S^*}(S_2)}=\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^Tw-\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_1\left(\mathsf{X}_1^T(I-\mathsf{P}_2)\mathsf{X}_1\right)^{-1}\mathsf{X}_1^T(I-\mathsf{P}_2)w.$$

Thus

$$
\begin{aligned}
&\sum_{j\in S_2}\left(\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}-\left[\left(X_{S^*}^TX_{S^*}\right)^{-1}X_{S^*}^Tw\right]_{\phi_{S^*}(j)}\right)^2\\
&=\left\|\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_1\left(\mathsf{X}_1^T(I-\mathsf{P}_2)\mathsf{X}_1\right)^{-1}\mathsf{X}_1^T(I-\mathsf{P}_2)w-\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_3\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T(I-\mathsf{P}_2)w\right\|^2\\
&\leq\left\|\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_1\right\|^2\left\|\left(\mathsf{X}_1^T(I-\mathsf{P}_2)\mathsf{X}_1\right)^{-1}\mathsf{X}_1^T(I-\mathsf{P}_2)w\right\|^2\\
&\quad+\left\|\left(\mathsf{X}_2^T\mathsf{X}_2\right)^{-1}\mathsf{X}_2^T\mathsf{X}_3\right\|^2\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T(I-\mathsf{P}_2)w\right\|^2\\
&\leq\left(C\frac{s\log p}{n}\right)\left(C\frac{|S_1|\log p}{n}+C\frac{|S_3|\log p}{n}\right)\\
&=C^2\frac{sh(z,z^*)\log^2 p}{n^2},
\end{aligned}
$$

where the second to the last inequality is due to (G.3) and (G.4).

(2.3). Combining the above results, we have

$$
\begin{aligned}
\sum_{j\in S_2}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2&\leq 2\left(C\frac{s\log p}{n}\left\|\beta_{S_1}^*\right\|^2+C^2\frac{sh(z,z^*)\log^2 p}{n^2}\right)\\
&\leq\left(\frac{4Cs\log p}{n}+\frac{4C^2s\log^2 p}{\lambda^2n^2}\right)\frac{\ell(z,z^*)}{n},
\end{aligned}
$$

where the last inequality is by (G.20) and (G.21).

(3) Regime $j\in S_3$. Since

$$
\begin{aligned}
\widetilde{\beta}_j(z)&=\beta_j^*+\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta_{S_1}^*\right]_{\phi_S(j)}+\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)},\\
\widetilde{\beta}_j(z^*)&=\frac{1}{\|X_j\|^2}X_j^T(I-P_{S^*})w,
\end{aligned}
$$

and $\beta_j^*=0$, we can write $\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)$ as

$$
\begin{aligned}
&\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta_{S_1}^*\right]_{\phi_S(j)}+\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}-\frac{1}{\|X_j\|^2}X_j^T(I-P_{S^*})w\\
&=\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta_{S_1}^*\right]_{\phi_S(j)}+\frac{1}{\|X_j\|^2}X_j^TP_{S^*}w+\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}-\frac{1}{\|X_j\|^2}X_j^Tw\\
&=\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta_{S_1}^*\right]_{\phi_S(j)}+\frac{1}{\|X_j\|^2}X_j^TP_{S^*}w+\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}-\frac{1}{\|X_j\|^2}X_j^Tw.
\end{aligned}
$$

We are going to bound each term separately. The last two terms will be analyzed together.

(3.1). Note that the first term here is identical to the first term in the regime $j\in S_2$. By the same argument, we have

$$
\sum_{j\in S_3}\left[\left(X_S^TX_S\right)^{-1}X_S^TX_{S_1}\beta_{S_1}^*\right]_{\phi_S(j)}^2\leq C\frac{s\log p}{n}\left\|\beta_{S_1}^*\right\|^2.
$$

(3.2). For the second term, we have

$$\sum_{j\in S_3}\left(\frac{1}{\|X_j\|^2}X_j^TP_{S^*}w\right)^2\le\frac{1}{\min_j\|X_j\|^4}\left\|X_3^TP_{S^*}w\right\|^2\le\frac{4}{n^2}Cs\,|S_3|\log p\le\frac{4}{n^2}Csh(z,z^*)\log p,$$

where the second to last inequality is due to (G.1) and (G.8).

(3.3). For the last two terms, we can again apply block matrix inverse formula to simplify them. Using (G.28), we have

$$\begin{aligned}\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}&=\left[\begin{pmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{pmatrix}\begin{pmatrix}\mathsf{X}_2^T\\\mathsf{X}_3^T\end{pmatrix}w\right]_{\phi_S(j)}\\&=\left[A_{21}\mathsf{X}_2^Tw+A_{22}\mathsf{X}_3^Tw\right]_{\phi_{S_3}(j)}\\&=-\left[\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{P}_2w\right]_{\phi_S(j)}+\left[\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^Tw\right]_{\phi_S(j)}.\end{aligned}$$

Then

$$\begin{aligned}\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}-\frac{1}{\|X_j\|^2}X_j^Tw=&-\left[\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{P}_2w\right]_{\phi_S(j)}\\&+\left(\left[\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^Tw\right]_{\phi_S(j)}-\frac{1}{\|X_j\|^2}\mathsf{X}_j^Tw\right).\end{aligned}$$

Consequently,

$$\begin{aligned}&\sum_{j\in S_3}\left(\left[\left(X_S^TX_S\right)^{-1}X_S^Tw\right]_{\phi_S(j)}-\frac{1}{\|X_j\|^2}X_j^Tw\right)^2\\&=\left\|-\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{P}_2w+\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^Tw-D^{-1}\mathsf{X}_3^T\right\|^2\\&\le2\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{P}_2w\right\|^2+2\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^Tw-D^{-1}\mathsf{X}_3^Tw\right\|^2\\&\le2\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{P}_2w\right\|^2+2\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}-D^{-1}\right\|^2\left\|\mathsf{X}_3^Tw\right\|^2\\&\le2\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{P}_2w\right\|^2+4\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}-(n-|S_2|)^{-1}I_{|S_3|}\right\|^2\left\|\mathsf{X}_3^Tw\right\|^2\\&\quad+4\left\|(n-|S_2|)^{-1}I_{|S_3|}-D^{-1}\right\|^2\left\|\mathsf{X}_3^Tw\right\|^2\end{aligned}$$

where $D\in\mathbb{R}^{|S_3|\times|S_3|}$ is a diagonal matrix with diagonal entries $\{1/\|X_j\|^2\}_{j\in S_3}$ and off-diagonal entries being 0. By (G.9), we have

$$\left\|\left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1}\mathsf{X}_3^T\mathsf{P}_2w\right\|^2\le Cn^{-2}s\,|S_3|\log p.$$

By (G.10), we have

$$\left\|\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3-(n-|S_2|)\,I_{|S_3|}\right\|^2\le Cns\log p.$$

Together with (G.11), we have

$$\begin{aligned}
&\left\| \left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1} - (n-|S_2|)^{-1} I_{|S_3|} \right\|^2 \\
&\le \left\| \left(\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3\right)^{-1} \right\|^2 \left\| I_{|S_3|} - \frac{\mathsf{X}_3^T(I-\mathsf{P}_2)\mathsf{X}_3}{n-|S_2|} \right\|^2 \\
&\le \frac{2C^3 s\log p}{n^3},
\end{aligned}$$

where we have used the assumption $|S_3| \le 2C_0 s$. By (G.1), we have

$$\begin{aligned}
\left\| (n-|S_2|)^{-1} I_{|S_3|} - D^{-1} \right\|^2 &\le \max_{j\in[p]} \left| \frac{1}{n-|S_2|} - \frac{1}{\|X_j\|^2} \right|^2 \\
&\le \max\left\{ \left| \frac{1}{n} - \frac{1}{n - C\sqrt{n\log p}} \right|^2, \left| \frac{1}{n-2C_0 s} - \frac{1}{n} \right|^2 \right\} \\
&\le \frac{2C\log p}{n^3} + \frac{8s^2}{n^4}.
\end{aligned}$$

By (G.12), we have

$$\left\| X_3^T w \right\|^2 \le Cn\,|S_3|\log p.$$

As a consequence, we have

$$\begin{aligned}
&\sum_{j\in S_3} \left(\left[\left(X_S^T X_S\right)^{-1} X_S^T w \right]_{\phi_S(j)} - \frac{1}{\|X_j\|^2} X_j^T w \right)^2 \\
&\le 2Cn^{-2} s\,|S_3|\log p + \left(\frac{4C^3 s\log p}{n^3} + \frac{8s^2}{n^4} \right) Cn\,|S_3|\log p \\
&\le 8C^3 \frac{s\log p}{n^2}\,|S_3|\log p \\
&\le 8C^3 \frac{s\log p}{n^2} h(z,z^*)\log p.
\end{aligned}$$

(3.4). Combining the above results, we have

$$\begin{aligned}
&\sum_{j\in S_3} \left(\widetilde{\beta}_j(z) - \widetilde{\beta}_j(z^*) \right)^2 \\
&\le 3\left(C\frac{s\log p}{n} \left\| \beta_{S_1}^* \right\|^2 + \frac{4}{n^2} Csh(z,z^*)\log p + 8C^3 \frac{s\log p}{n^2} h(z,z^*)\log p \right) \\
&\le 3\left(\frac{2Cs\log p}{n} + \frac{32C^3 s\log^2 p}{\lambda^2 n^2} \right) \frac{\ell(z,z^*)}{n},
\end{aligned}$$

where the last inequality is by (G.20) and (G.21).

(4) Regime $j \in S_4$. In this case, we have

$$\widetilde{\beta}_j(z) - \widetilde{\beta}_j(z^*) = \frac{1}{\|X_j\|^2} X_j^T (I - P_S) X_{S_1}\beta_{S_1}^* + \frac{1}{\|X_j\|^2} X_j^T (P_{S^*} - P_S) w.$$

Then,

$$\max_{T\subset S_4}\frac{1}{|T|\vee s}\sum_{j\in T}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2$$

$$\leq \frac{1}{\min_j \|X_j\|^4}\left(\max_{T\subset S_4}\frac{1}{|T|\vee s}\sum_{j\in T}\left(X_j^T(I-P_S)X_{S_1}\beta^*_{S_1}\right)^2+\max_{T\subset S_4}\frac{1}{|T|\vee s}\sum_{j\in T}\left(X_j^T(P_{S^*}-P_S)w\right)^2\right)$$

$$=\frac{1}{\min_j \|X_j\|^4}\left(\max_{T\subset S_4}\frac{1}{|T|\vee s}\left\|X_T^T(I-P_S)X_{S_1}\beta^*_{S_1}\right\|^2+\max_{T\subset S_4}\frac{1}{|T|\vee s}\left\|X_T^T(P_{S^*}-P_S)w\right\|^2\right)$$

$$\leq \frac{4}{n^2}\left(Cn\log p\left\|\beta^*_{S_1}\right\|^2+C\left(|S_1|+|S_3|\right)\log^2 p\right)$$

$$=\frac{4}{n^2}\left(Cn\log p\left\|\beta^*_{S_1}\right\|^2+Ch(z,z^*)\log^2 p\right).$$

where the last inequality is by (G.1), (G.13) and (G.14). Then by (G.20) and (G.21), we have

$$\max_{T\subset S_4}\frac{1}{|T|\vee s}\sum_{j\in T}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2\leq\left(\frac{8C\log p}{n}+\frac{8C\log^2 p}{\lambda^2n^2}\right)\frac{\ell(z,z^*)}{n}.$$

(5) Combining the bounds. Now we are ready to combine the bounds obtained in the four regimes. Let $T\subset[p]$ be any set. We have

$$\sum_{j\in T}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2$$

$$\leq\sum_{j\in S_2}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2+\sum_{j\in S_1}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2+\sum_{j\in S_3}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2+\sum_{j\in S_4\cap T}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2$$

$$\leq\left(\left(\frac{4Cs\log p}{n}+\frac{4C^2s\log^2 p}{\lambda^2n^2}\right)+4\left(\frac{16C^2s^2\log^2 p}{n^2}+\frac{16s\log p}{\lambda^2n^2}\right)+3\left(\frac{2Cs\log p}{n}+\frac{32C^3s\log^2 p}{\lambda^2n^2}\right)\right)\frac{\ell(z,z^*)}{n}$$

$$+\left(\frac{8C\log p}{n}+\frac{8C\log^2 p}{\lambda^2n^2}\right)\left(|T|\vee s\right)\frac{\ell(z,z^*)}{n}$$

$$\leq\left(\frac{128C^2\log p}{n}+\frac{256C^3\log^2 p}{\lambda^2n^2}\right)(s+|T|)\frac{\ell(z,z^*)}{n}.$$

Thus

$$\max_{T\subset[p]}\frac{1}{s+|T|}\sum_{j\in T}\left(\widetilde{\beta}_j(z)-\widetilde{\beta}_j(z^*)\right)^2\leq\left(\frac{128C^2\log p}{n}+\frac{256C^3\log^2 p}{\lambda^2n^2}\right)\frac{\ell(z,z^*)}{n}.$$

Together with (G.26), we have

$$\max_{T\subset[p]}\frac{1}{s+|T|}\sum_{j\in T}\max_{b\in\{-1,1,0\}\setminus\{z_j^*\}}\frac{G_j(z_j^*,b;z)^2\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)}$$

$$\leq 2\left(\frac{128C^2\log p}{n}+\frac{256C^3\log^2 p}{\lambda^2n^2}\right)\max_{j\in[p]}\max\left\{4\left(\frac{\left|\beta_j^*\right|}{\left|\beta_j^*\right|-t(X_j)}\right)^2,\left(\frac{\lambda}{t(X_j)}\right)^2\right\}.$$

Recall that $\Delta_{\min}^2 = \lambda^2 \min_{j\in[p]} \|X_j\|^2 \ge n\lambda^2/2$. For any $T \subset [p]$, we have $\tau/(\tau + 4\Delta_{\min}^2 |T|) \le \tau/(\tau + 2n\lambda^2 |T|) \le C_0/(C_0 s + |T|)$, since $\tau \le C_0 s n \lambda^2$. This gives us

$$\begin{aligned}
&\max_{T\subset[p]} \frac{\tau}{\tau + 4\Delta_{\min}^2 |T|} \sum_{j\in T} \max_{b\in\{-1,1,0\}\setminus\{z_j^*\}} \frac{G_j(z_j^*, b; z)^2 \|\mu_j(B^*, b) - \mu_j(B^*, z_j^*)\|^2}{\Delta_j(z_j^*, b)^4 \ell(z, z^*)} \\
&\le C' s \left(\frac{\log^2 p}{n} + \frac{\log^2 p}{\lambda^2 n^2} \right) \max_{j\in[p]} \max \left\{ 4 \left(\frac{\left|\beta_j^*\right|}{\left|\beta_j^*\right| - t(X_j)} \right)^2, \left(\frac{\lambda}{t(X_j)} \right)^2 \right\},
\end{aligned}$$

for some constant C'. The proof is complete. □

Proof of Lemma 5.2. Recall for any $j \in [p]$, T_j the local test to recover z_j^* is defined in (13). We have the decomposition $T_j = \mu_j(B^*, z_j^*) + \epsilon_j$, where $\epsilon_j = \|X_j\|^{-1} X_j^T w \sim \mathcal{N}(0,1)$. Since $\nu_j(\widehat{B}(z^*), z_j^*) - \nu_j(\widehat{B}(z^*), b) = 2(z_j^* - b)\|X_j\| t(X_j)$, by (46), for any $0 < \delta < 1$, we have

$$\begin{aligned}
&\mathbf{1}_{\left\{\left\langle \epsilon_j, \nu_j(\widehat{B}(z^*), z_j^*) - \nu_j(\widehat{B}(z^*), b)\right\rangle \le -\frac{1-\delta}{2}\Delta_j(z_j^*, b)^2\right\}} \\
&\le \begin{cases} \mathbf{1}_{\left\{z_j^*\epsilon_j \le -(1-\delta)\|X_j\|\left(\left|\beta_j^*\right| - t(X_j)\right)\right\}} & z_j^* \ne 0 \text{ and } b \ne z_j^* \\ \mathbf{1}_{\left\{-b\epsilon_j \le -(1-\delta)\|X_j\| t(X_j)\right\}} & z_j^* = 0 \text{ and } b \ne 0. \end{cases}
\end{aligned}$$

Together with (44), for $b \ne z_j^*$, we have

$$\begin{aligned}
&\|\mu_j(B^*, b) - \mu_j(B^*, z_j^*)\|^2 \mathbf{1}_{\left\{\left\langle \epsilon_j, \nu_j(\widehat{B}(z^*), z_j^*) - \nu_j(\widehat{B}(z^*), b)\right\rangle \le -\frac{1-\delta}{2}\Delta_j(z_j^*, b)^2\right\}} \\
&\le \begin{cases} 4\left|\beta_j^*\right|^2 \|X_j\|^2 \mathbf{1}_{\left\{z_j^*\epsilon_j \le -(1-\delta)\|X_j\|\left(\left|\beta_j^*\right| - t(X_j)\right)\right\}} & z_j^* \ne 0 \\ 4\lambda^2 \|X_j\|^2 \mathbf{1}_{\left\{-b\epsilon_j \le -(1-\delta)\|X_j\| t(X_j)\right\}} & z_j^* = 0. \end{cases}
\end{aligned}$$

As a consequence

$$\xi_{\text{ideal}}(\delta) \le 8 \sum_{j\in S^*} \left|\beta_j^*\right|^2 \|X_j\|^2 \mathbf{1}_{\left\{z_j^*\epsilon_j \le -(1-\delta)\|X_j\|\left(\left|\beta_j^*\right| - t(X_j)\right)\right\}} + 4 \sum_{j\notin S^*} \lambda^2 \|X_j\|^2 \mathbf{1}_{\{|\epsilon_j| \ge (1-\delta)\|X_j\| t(X_j)\}}.$$

Define $\mathcal{F}$ to be the event that (G.1) holds. Then by Lemma G.1, we know that $\mathbb{P}(\mathcal{F}) \ge 1 - p^{-C'}$. Under the event $\mathcal{F}$ and the condition that $\mathsf{SNR} \to \infty$, we have

$$\begin{aligned}
\xi_{\text{ideal}}(\delta)\mathbf{1}_{\{\mathcal{F}\}} \le\ & 8 \sum_{j\in S^*} \left|\beta_j^*\right|^2 \|X_j\|^2 \mathbf{1}_{\left\{\frac{-z_j^*\epsilon_j}{1-\delta} \ge \|X_j\|\left(\left|\beta_j^*\right| - t(X_j)\right)\right\}} \mathbf{1}_{\left\{\left|\|X_j\|^2 - n\right| \le C\sqrt{n\log p}\right\}} \\
&+ 4 \sum_{j\notin S^*} \lambda^2 \|X_j\|^2 \mathbf{1}_{\left\{\frac{|\epsilon_j|}{1-\delta} \ge \|X_j\| t(X_j)\right\}} \mathbf{1}_{\left\{\left|\|X_j\|^2 - n\right| \le C\sqrt{n\log p}\right\}},
\end{aligned}$$

which implies

$$\begin{aligned}
\mathbb{E}\xi_{\text{ideal}}(\delta)\mathbf{1}_{\{\mathcal{F}\}} \le\ & 16n \sum_{j\in S^*} \left|\beta_j^*\right|^2 \mathbb{P}\left(\frac{\epsilon_j}{1-\delta} \ge \|X_j\| \left(\left|\beta_j^*\right| - t(X_j)\right) \ \&\ \left|\|X_j\|^2 - n\right| \le C\sqrt{n\log p} \right) \\
&+ 16n \sum_{j\notin S^*} \lambda^2 \mathbb{P}\left(\frac{\epsilon_j}{1-\delta} \ge \|X_j\| t(X_j) \ \&\ \left|\|X_j\|^2 - n\right| \le C\sqrt{n\log p} \right).
\end{aligned}$$

We are going to upper bound the above quantity by $\widetilde{\psi}(n,p,s,\lambda,\delta,C)$, defined in (G.16). To this end, we will first show the function $f(y)=y^2\mathbb{P}\left(\frac{\epsilon}{1-\delta}\geq\|\zeta\|\left(y-t(\zeta)\right),\left|\|\zeta\|^2-n\right|\leq C\sqrt{n\log p}\right)$ is a decreasing function of y when $y\geq\lambda$ and $\lambda>0$. Since the function $t(\zeta)$ only depends on $\|\zeta\|$, we can also write $t(\zeta)$ as $t(\|\zeta\|)$ with a slight abuse of notation in (G.18). Define $u_{\min}=\sqrt{n-C\sqrt{n\log p}}$ and $u_{\max}=\sqrt{n+C\sqrt{n\log p}}$. Then, we have

$$\begin{aligned}f(y)&=y^2\mathbb{P}\left(\frac{\epsilon}{1-\delta}\geq\|\zeta\|\left(y-t(\zeta)\right),\left|\|\zeta\|^2-n\right|\leq C\sqrt{n\log p}\right)\\&=y^2\int_{u_{\min}}^{u_{\max}}p(u)G\left(\left(1-\delta\right)u\left(y-t(u)\right)\right)du,\end{aligned}$$

where $p(\cdot)$ is the density of $\|\zeta\|$. According to the same argument used in the proof of Lemma G.3, it can be shown that $\min_{u\in[u_{\min},u_{\max}]}u(\lambda-t(u))\to\infty$. Thus, $u(y-t(u))\geq u(\lambda-t(u))>0$ for $y\geq\lambda$ and $u\in[u_{\min},u_{\max}]$. Moreover, we also have $\left(1-\delta\right)^2u^2y\left(y-t(u)\right)\geq\left(1-\delta\right)^2u^2\lambda\left(\lambda-t(u)\right)>2$ for $y\geq\lambda$ and $u\in[u_{\min},u_{\max}]$. Therefore,

$$\begin{aligned}\frac{2}{\left(1-\delta\right)u\left(y-t(u)\right)}-y\left(1-\delta\right)u&=\frac{2-\left(1-\delta\right)^2u^2y(y-t(u))}{\left(1-\delta\right)u\left(y-t(u)\right)}\\&\leq\frac{2-\left(1-\delta\right)^2u^2\lambda(\lambda-t(u))}{\left(1-\delta\right)u\left(y-t(u)\right)}\\&<0.\end{aligned}$$

This gives

$$\begin{aligned}f'(y)&=\int_{u_{\min}}^{u_{\max}}p(u)\left(2yG\left(\left(1-\delta\right)u\left(y-t(u)\right)\right)-y^2\left(1-\delta\right)ug\left(\left(1-\delta\right)u\left(y-t(u)\right)\right)\right)du\\&\leq\int_{u_{\min}}^{u_{\max}}p(u)y\left(\frac{2}{\left(1-\delta\right)u\left(y-t(u)\right)}-y\left(1-\delta\right)u\right)g\left(\left(1-\delta\right)u\left(y-t(u)\right)\right)du\\&\leq 0,\end{aligned}$$

where we have used (G.17). As a result, $f(y)$ is a decreasing function for all $y\geq\lambda$, which implies

$$\begin{aligned}\mathbb{E}\xi_{\text{ideal}}(\delta)\mathbf{1}_{\{\mathcal{F}\}}&\leq 16n\sum_{j\in S^*}\lambda^2\mathbb{P}\left(\frac{\epsilon_j}{1-\delta}\geq\|X_j\|\left(\lambda-t(X_j)\right)\ \&\ \left|\|X_j\|^2-n\right|\leq C\sqrt{n\log p}\right)\\&\quad+16n\sum_{j\notin S^*}\lambda^2\mathbb{P}\left(\frac{\epsilon_j}{1-\delta}\geq\|X_j\|\,t(X_j)\ \&\ \left|\|X_j\|^2-n\right|\leq C\sqrt{n\log p}\right)\\&=16n\lambda^2\widetilde{\psi}(n,p,s,\lambda,\delta,C).\end{aligned}$$

By applying Markov inequality, we have with probability at least $1-w^{-1}$,

$$\xi_{\text{ideal}}(\delta)\mathbf{1}_{\{\mathcal{F}\}}\leq 16wn\lambda^2\widetilde{\psi}(n,p,s,\lambda,\delta,C),$$

where w is any sequence that goes to infinity. A union bound implies

$$\xi_{\text{ideal}}(\delta)\leq 16wn\lambda^2\widetilde{\psi}(n,p,s,\lambda,\delta,C)\tag{G.29}$$

holds with probability at least $1 - w^{-1} - p^{-C'}$. Taking $\delta = \delta_p = o((\log p)^{-1})$ and $w = \exp(\mathsf{SNR})$, the desired conclusion follows an application of Lemma G.2. Thus, the proof is complete. $\square$

Proof of Proposition 5.1. By Proposition 5.1 of [12], we have

$$\left\|\widetilde{\beta} - \beta^*\right\|^2 \leq \frac{C_1 s \log \frac{ep}{s}}{n},$$

with probability at least $1 - 2^{-C_2 s}$ for some constants $C_1, C_2 > 0$, as long as A is chosen to be sufficiently large. In the rest part of the proof, we assume (G.1) holds. We divide the calculation of $\ell(\widetilde{z}, z^*)$ into three parts. First we have

$$\begin{aligned}\sum_{j=1}^p \lambda^2 \|X_j\|^2 \mathbb{I}\{\widetilde{z}_j \neq 0, z_j^* = 0\} &\leq \sum_{j=1}^p \lambda^2 \|X_j\|^2 \mathbb{I}\left\{|\widetilde{\beta}_j| > \frac{\lambda}{2}, \beta_j^* = 0\right\} \\ &\leq 4\sum_{j=1}^p \|X_j\|^2 (\widetilde{\beta}_j - \beta_j^*)^2 \mathbb{I}\left\{|\widetilde{\beta}_j| > \frac{\lambda}{2}, \beta_j^* = 0\right\} \\ &\leq 8n\sum_{j=1}^p (\widetilde{\beta}_j - \beta_j^*)^2 \mathbb{I}\left\{|\widetilde{\beta}_j| > \frac{\lambda}{2}, \beta_j^* = 0\right\}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\sum_{j=1}^p |\beta_j^*|^2 \|X_j\|^2 \mathbb{I}\{z_j^* \neq 0, \widetilde{z}_j = 0\} &\leq \sum_{j=1}^p |\beta_j^*|^2 \|X_j\|^2 \mathbb{I}\left\{|\beta_j^*| \geq \lambda, |\widetilde{\beta}_j| \leq \frac{\lambda}{2}\right\} \\ &\leq 8n\sum_{j=1}^p (\widetilde{\beta}_j - \beta_j^*)^2 \mathbb{I}\left\{|\beta_j^*| \geq \lambda, |\widetilde{\beta}_j| \leq \frac{\lambda}{2}\right\},\end{aligned}$$

where in the last inequality, since $|\beta_j^*| \geq \lambda$ and $|\widetilde{\beta}_j| \leq \frac{\lambda}{2}$, we have

$$|\beta_j^* - \widetilde{\beta}_j| \geq |\beta_j^*| - |\widetilde{\beta}_j| \geq \frac{|\widetilde{\beta}_j|}{2} + \frac{\lambda}{2} - \frac{\lambda}{2} = \frac{|\widetilde{\beta}_j|}{2}.$$

Finally,

$$\begin{aligned}&4\sum_{j=1}^p |\beta_j^*|^2 \|X_j\|^2 \mathbb{I}\{\widetilde{z}_j z_j^* = -1\} \\ \leq\ &4\sum_{j=1}^p |\beta_j^*|^2 \|X_j\|^2 \mathbb{I}\left\{\beta_j^* \leq -\lambda, \widetilde{\beta}_j > \frac{\lambda}{2}\right\} + 4\sum_{j=1}^p |\beta_j^*|^2 \|X_j\|^2 \mathbb{I}\left\{\beta_j^* \geq \lambda, \widetilde{\beta}_j < -\frac{\lambda}{2}\right\} \\ \leq\ &8n\sum_{j=1}^p (\widehat{\beta}_j - \beta_j^*)^2 \mathbb{I}\left\{\beta_j^* \leq -\lambda, \widetilde{\beta}_j > \frac{\lambda}{2}\right\} \\ &+8n\sum_{j=1}^p (\widetilde{\beta}_j - \beta_j^*)^2 \mathbb{I}\left\{\beta_j^* \geq \lambda, \widehat{\beta}_j < -\frac{\lambda}{2}\right\},\end{aligned}$$

because when $\beta_j^* \leq -\lambda$ and $\widetilde{\beta}_j > \frac{\lambda}{2}$, we have

$$|\beta_j^* - \widetilde{\beta}_j| = -\beta_j^* + \widetilde{\beta}_j \geq |\beta_j^*|,$$

and when $\beta_j^* \geq \lambda$ and $\widetilde{\beta}_j < -\frac{\lambda}{2}$, we have

$$|\beta_j^* - \widetilde{\beta}_j| = \beta_j^* - \widetilde{\beta}_j \geq |\beta_j^*|.$$

Combining all of the above results together, we have

$$\ell(\widetilde{z}, z^*) \leq 8n \left\|\widetilde{\beta} - \beta^*\right\|^2 \leq 8C_1 s \log \frac{ep}{s}.$$

Under the assumption $\limsup s/p < 1/2$ and $\mathsf{SNR} \to \infty$, we have $n\lambda^2 \geq 2 \log \frac{p-s}{s} > C_3 \log \frac{ep}{s}$. Thus, there exists some constant $C_0 > 0$ such that $\ell(\widetilde{z}, z^*) \leq C_0 sn\lambda^2$. A union bound with the probability that the event (G.1) holds leads to the desired result. □

H Proofs in Section 6

In this section, we present the proofs of Theorem 6.1, Lemma 6.1, Lemma 6.2 and Proposition 6.1. The conclusions of Theorem 6.2 and Corollary 6.1 are direct consequences of Theorem 3.1, and thus their proofs are omitted.

Proof of Theorem 6.1. Recall the definition $\Delta_{\min} = \min_{U \in \mathcal{C}_d} \|(I_d - U)\theta^*\|$. There must exists some $\bar{U} \in \mathcal{C}_d$ such that $\Delta_{\min} = \|(I_d - \bar{U})\theta^*\|$. We define the parameter space

$$\mathcal{Z} = \left\{Z : Z_j = I_d \text{ for all } 1 \leq j \leq p/2 \text{ and } Z_j \in \{I_d, \bar{U}\} \text{ for all } p/2 < j \leq p\right\}.$$

Then, we have

$$\begin{aligned}
& \inf_{\widehat{Z}} \sup_{Z^*} \mathbb{E} \min_{U \in \mathcal{C}_d} \frac{1}{p} \sum_{j=1}^{p} \mathbf{1}_{\{\widehat{Z}_j U \neq Z_j^*\}} \\
\geq \;& \inf_{\widehat{Z}} \sup_{Z^* \in \mathcal{Z}} \frac{1}{p} \sum_{j > p/2} \mathbb{P}(\widehat{Z}_j \neq Z_j^*) \\
\geq \;& \inf_{\widehat{Z}} \frac{1}{p} \sum_{j > p/2} \mathrm{ave}_{Z_{-j}^*} \left(\frac{1}{2} \mathbb{P}_{(Z_j^* = I_d, Z_{-j}^*)}(\widehat{Z}_j \neq I_d) + \frac{1}{2} \mathbb{P}_{(Z_j^* = \bar{U}, Z_{-j}^*)}(\widehat{Z}_j \neq \bar{U}) \right) \\
\geq \;& \frac{1}{p} \sum_{j > p/2} \mathrm{ave}_{Z_{-j}^*} \inf_{\widehat{Z}_j} \left(\frac{1}{2} \mathbb{P}_{(Z_j^* = I_d, Z_{-j}^*)}(\widehat{Z}_j \neq I_d) + \frac{1}{2} \mathbb{P}_{(Z_j^* = \bar{U}, Z_{-j}^*)}(\widehat{Z}_j \neq \bar{U}) \right), \qquad \text{(H.1)}
\end{aligned}$$

where the operator $\mathrm{ave}_{Z_{-j}^*}$ is with respect to the uniform measure of Z_{-j}^* in the space $\mathcal{Z}$. We use the notation Z_{-j}^* for Z^* with Z_j^* excluded. Note that the quantity

$$\inf_{\widehat{Z}_j} \left(\frac{1}{2} \mathbb{P}_{(Z_j^* = I_d, Z_{-j}^*)}(\widehat{Z}_j \neq I_d) + \frac{1}{2} \mathbb{P}_{(Z_j^* = \bar{U}, Z_{-j}^*)}(\widehat{Z}_j \neq \bar{U}) \right)$$

is the optimal testing error for the hypothesis testing problem $H_0 : Y_j \sim \mathcal{N}(\theta^*, I_d)$ vs $H_1 : Y_j \sim \mathcal{N}(\bar{U}\theta^*, I_d)$. By Neyman-Pearson lemma, this quantity equals $\mathbb{P}(\mathcal{N}(0,1) > \Delta_{\min}/2)$. Therefore,

$$\inf_{\widehat{Z}} \sup_{Z^*} \mathbb{E} \min_{U \in \mathcal{C}_d} \frac{1}{p} \sum_{j=1}^{p} \mathbf{1}_{\{\widehat{Z}_j U \neq Z_j^*\}} \geq \frac{1+o(1)}{2} \mathbb{P}(\mathcal{N}(0,1) > \Delta_{\min}/2) = \exp\left(-(1+o(1))\frac{\Delta_{\min}^2}{8}\right),$$

under the condition that $\Delta_{\min} \to \infty$. Moreover, $\inf_{\widehat{Z}} \sup_{Z^*} \mathbb{E} \min_{U \in \mathcal{C}_d} \frac{1}{p} \sum_{j=1}^{p} \mathbf{1}_{\{\widehat{Z}_j \neq U Z_j^*\}} > 0$ for some constant $c > 0$ when $\Delta_{\min} = O(1)$. □

Proof of Lemma 6.1. Let us write $\widehat{\theta}(Z) = \theta(Z) + \bar{\epsilon}(Z)$, where $\theta(Z) = \frac{1}{p}\sum_{j=1}^{p} Z_j^T Z_j^* \theta^*$ and $\bar{\epsilon}(Z) = \frac{1}{p}\sum_{j=1}^{p} Z_j^T \epsilon_j$. We have

$$\begin{aligned} \|\theta(Z) - \theta(Z^*)\| &= \left\| \left(\frac{1}{p} \sum_{j=1}^{p} Z_j^T Z_j^* - I_d \right) \theta^* \right\| \\ &\leq \frac{1}{p} \sum_{j=1}^{p} \|(Z_j^T Z_j^* - I_d)\theta^*\| \\ &\leq \frac{1}{p\Delta_{\min}} \sum_{j=1}^{p} \|(Z_j^T Z_j^* - I_d)\theta^*\|^2 \\ &= \frac{1}{p\Delta_{\min}} \ell(Z, Z^*). \end{aligned} \tag{H.2}$$

For $\|\bar{\epsilon}(Z) - \bar{\epsilon}(Z^*)\| = \left\| \frac{1}{p}\sum_{j=1}^{p} (Z_j - Z_j^*)^T \epsilon_j \right\|$, we notice that

$$\frac{1}{p}\sum_{j=1}^{p} (Z_j - Z_j^*)^T \epsilon_j \sim \mathcal{N}\left(0, \frac{1}{p^2}\sum_{j=1}^{p} (Z_j - Z_j^*)^T (Z_j - Z_j^*)\right).$$

Therefore, by Lemma E.1, we have for each fixed $Z \in \mathcal{C}_d^p$,

$$\mathbb{P}\left(\|\bar{\epsilon}(Z) - \bar{\epsilon}(Z^*)\|^2 \geq \left\| \frac{1}{p^2} \sum_{j=1}^{p} (Z_j - Z_j^*)^T (Z_j - Z_j^*) \right\| (d + 2\sqrt{dx} + 2x) \right) \leq e^{-x}.$$

With a union bound argument, we have

$$\|\bar{\epsilon}(Z) - \bar{\epsilon}(Z^*)\| \lesssim \left\| \frac{1}{p^2} \sum_{j=1}^{p} (Z_j - Z_j^*)^T (Z_j - Z_j^*) \right\|^{1/2} \sqrt{p \log d},$$

uniformly over all $Z \in \mathcal{C}_d^p$ with probability at least $1 - e^{-C'p\log d}$. Since

$$\begin{aligned}\left\|\frac{1}{p^2}\sum_{j=1}^p (Z_j - Z_j^*)^T(Z_j - Z_j^*)\right\| &\leq \frac{1}{p^2}\sum_{j=1}^p \|Z_j - Z_j^*\|^2 \\ &\lesssim \frac{1}{p^2}\sum_{j=1}^p \mathbf{1}_{\{Z_j \neq Z_j^*\}} \\ &\leq \frac{1}{p^2\Delta_{\min}^2}\ell(Z, Z^*),\end{aligned}$$

we have

$$\|\bar{\epsilon}(Z) - \bar{\epsilon}(Z^*)\| \lesssim \sqrt{\frac{\log d}{p\Delta_{\min}^2}}\sqrt{\ell(Z, Z^*)}. \tag{H.3}$$

Combine (H.2) and (H.3), and we obtain

$$\|\widehat{\theta}(Z) - \widehat{\theta}(Z^*)\| \lesssim \frac{1}{p\Delta_{\min}}\ell(Z, Z^*) + \sqrt{\frac{\log d}{p\Delta_{\min}^2}}\sqrt{\ell(Z, Z^*)}. \tag{H.4}$$

With (H.4), we are prepared to derive the bounds (53)-(55). For (53), we have

$$\begin{aligned}&\sum_{j=1}^p \max_{U \neq Z_j^*} \frac{F_j(Z_j^*, U; Z)^2\|\mu_j(B^*, U) - \mu_j(B^*, Z_j^*)\|^2}{\Delta_j(Z_j^*, U)^4\ell(Z, Z^*)} \\ &= \sum_{j=1}^p \max_{U \neq Z_j^*} \frac{\left|\left\langle \epsilon_j, (Z_j^* - U)(\widehat{\theta}(Z^*) - \widehat{\theta}(Z))\right\rangle\right|^2}{\|(Z_j^* - U)\theta^*\|^2\ell(Z, Z^*)} \\ &\leq \frac{1}{\Delta_{\min}^2\ell(Z, Z^*)}\sum_{j=1}^p \max_{U \neq Z_j^*}(\widehat{\theta}(Z^*) - \widehat{\theta}(Z))^T(Z_j^* - U)^T\epsilon_j\epsilon_j^T(Z_j^* - U)(\widehat{\theta}(Z^*) - \widehat{\theta}(Z)) \\ &= \frac{1}{\Delta_{\min}^2\ell(Z, Z^*)} \max_{\tilde{U} \in \mathcal{C}_d^p : \tilde{U}_j \neq Z_j^*, \forall j}(\widehat{\theta}(Z^*) - \widehat{\theta}(Z))^T\left(\sum_{j=1}^p (Z_j^* - \tilde{U}_j)^T\epsilon_j\epsilon_j^T(Z_j^* - \tilde{U}_j)\right)(\widehat{\theta}(Z^*) - \widehat{\theta}(Z)) \\ &\leq \frac{\|\widehat{\theta}(Z^*) - \widehat{\theta}(Z)\|^2}{\Delta_{\min}^2\ell(Z, Z^*)} \max_{\tilde{U} \in \mathcal{C}_d^p : \tilde{U}_j \neq Z_j^*, \forall j}\left\|\sum_{j=1}^p (Z_j^* - \tilde{U}_j)^T\epsilon_j\epsilon_j^T(Z_j^* - \tilde{U}_j)\right\|.\end{aligned}$$

Note that

$$\begin{aligned}&\max_{\tilde{U} \in \mathcal{C}_d^p : \tilde{U}_j \neq Z_j^*, \forall j}\left\|\sum_{j=1}^p (Z_j^* - \tilde{U}_j)^T\epsilon_j\epsilon_j^T(Z_j^* - \tilde{U}_j)\right\|^2 \\ &= \max_{\tilde{U} \in \mathcal{C}_d^p : \tilde{U}_j \neq Z_j^*, \forall j}\left\|\left((Z_1^* - \tilde{U}_1)^T\epsilon_1, \ldots, (Z_p^* - \tilde{U}_p)^T\epsilon_p\right)\right\|^2 \\ &\leq 2\left\|\left(Z_1^T\epsilon_1, \ldots, Z_p^T\epsilon_p\right)\right\|^2 + \max_{\tilde{U} \in \mathcal{C}_d^p : \tilde{U}_j \neq Z_j^*, \forall j}\left\|\left(\tilde{U}_1^T\epsilon_1, \ldots, \tilde{U}_p^T\epsilon_p\right)\right\|^2 \\ &\leq 4\max_{\tilde{U} \in \mathcal{C}_d^p}\left\|\left(\tilde{U}_1^T\epsilon_1, \ldots, \tilde{U}_p^T\epsilon_p\right)\right\|^2.\end{aligned}$$

Note that for any fixed $\tilde{U} \in \mathcal{C}_d^p$, $(\tilde{U}_1^T \epsilon_1, \ldots, \tilde{U}_p^T \epsilon_p)$ each entries independently distributed from a standard normal distribution. By Lemma B.1 of [9], we have $\mathbb{P}(\|(\tilde{U}_1^T \epsilon_1, \ldots, \tilde{U}_p^T \epsilon_p)\| \geq \sqrt{p} + \sqrt{d} + t) \leq \exp\left(-t^2/2\right)$. By a union bound, we have

$$\max_{\tilde{U} \in \mathcal{C}_d^p : \tilde{U}_j \neq Z_j^*, \forall j} \left\| \sum_{j=1}^{p} (Z_j^* - \tilde{U}_j)^T \epsilon_j \epsilon_j^T (Z_j^* - \tilde{U}_j) \right\|^2 \leq 64\,(d + p \log d)$$

with probability at least $1 - \exp\left(-p \log d\right)$. Hence,

$$\begin{aligned} \sum_{j=1}^{p} \max_{U \neq Z_j^*} \frac{F_j(Z_j^*, U; Z)^2 \|\mu_j(B^*, U) - \mu_j(B^*, Z_j^*)\|^2}{\Delta_j(Z_j^*, U)^4 \ell(Z, Z^*)} &\leq 64\,(d + p \log d)\, \frac{\|\widehat{\theta}(Z^*) - \widehat{\theta}(Z)\|^2}{\Delta_{\min}^2 \ell(Z, Z^*)} \\ &\lesssim \frac{(\log d + d/p) \log d}{\Delta_{\min}^4} + \frac{\ell(Z, Z^*)(\log d + d/p)}{p \Delta_{\min}^4}, \end{aligned}$$

where the last inequality is by Lemma E.1 using the fact that $\sum_{j=1}^{p} \|\epsilon_j\|^2 \sim \chi^2_{pd}$ and (H.4). For (54), we have

$$\begin{aligned} & \frac{\tau}{4\Delta_{\min}^2 |T|} \sum_{j \in T} \max_{U \neq Z_j^*} \frac{G_j(Z_j^*, U; Z)^2 \|\mu_j(B^*, U) - \mu_j(B^*, Z_j^*)\|^2}{\Delta_j(Z_j^*, U)^4 \ell(Z, Z^*)} \\ = \; & \frac{\tau}{4\Delta_{\min}^2 |T|} \sum_{j \in T} \max_{U \neq Z_j^*} \frac{\left| \left\langle \widehat{\theta}(Z) - \widehat{\theta}(Z^*), (U^T Z_j^* - I_d)\theta^* \right\rangle \right|^2}{\|(Z_j^* - U)\theta^*\|^2 \ell(Z, Z^*)} \\ \leq \; & \frac{\tau}{4\Delta_{\min}^2} \frac{\|\widehat{\theta}(Z) - \widehat{\theta}(Z^*)\|^2}{\ell(Z, Z^*)} \\ \lesssim \; & \frac{\tau \log d}{p \Delta_{\min}^4} + \frac{\tau \ell(Z, Z^*)}{p^2 \Delta_{\min}^4}, \end{aligned}$$

where the last inequality is by (H.4). Finally, for (55), since $\widehat{\theta}(Z^*) - \theta^* = \frac{1}{p} \sum_{j=1}^{p} Z_j^T \epsilon_j \sim \mathcal{N}(0, p^{-1} I_d)$, we have $\|\widehat{\theta}(Z^*) - \theta^*\|^2 \lesssim \frac{d}{p}$ with probability at least $1 - e^{-C'd}$ by Lemma E.1. Then,

$$\frac{|H_j(Z_j^*, U)|}{\Delta_j(Z_j^*, U)^2} = \frac{\left| \left\langle \widehat{\theta}(Z^*) - \theta^*, (U^T Z_j^* - I_d)\theta^* \right\rangle \right|}{\|(Z_j^* - U)\theta^*\|^2} \leq \frac{\|\widehat{\theta}(Z^*) - \theta^*\|}{\Delta_{\min}} \lesssim \sqrt{\frac{d}{\Delta_{\min}^2 p}},$$

for any $j \in [p]$ and any $U \neq Z_j^*$. The proof is complete. □

Proof of Lemma 6.2. For any $U \in \mathcal{C}_d \backslash \{Z_1^*\}$, we have

$$\begin{aligned} & \mathbb{P}\left(\left\langle \epsilon_1, (Z_1^* - U)\widehat{\theta}(Z^*) \right\rangle \leq -\frac{1-\delta}{2} \|(Z_1^* - U)\theta^*\|^2 \right) \\ \leq \; & \mathbb{P}\left(\langle \epsilon_1, (Z_1^* - U)\theta^* \rangle \leq -\frac{1-\delta-\bar{\delta}}{2} \|(Z_1^* - U)\theta^*\|^2 \right) && \text{(H.5)} \\ & + \mathbb{P}\left(\left\langle \epsilon_1, (Z_1^* - U)(\widehat{\theta}(Z^*) - \theta^*) \right\rangle \leq -\frac{\bar{\delta}}{2} \|(Z_1^* - U)\theta^*\|^2 \right), && \text{(H.6)} \end{aligned}$$

where the sequence $\bar{\delta} = \bar{\delta}_p$ is to be chosen later. For (H.5), a standard Gaussian tail bound gives

$$\mathbb{P}\left(\langle \epsilon_1, (Z_1^* - U)\theta^* \rangle \leq -\frac{1-\delta-\bar{\delta}}{2}\|(Z_1^* - U)\theta^*\|^2\right) \leq \exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|(Z_1^* - U)\theta^*\|^2\right).$$

Since

$$\left\langle \epsilon_1, (Z_1^* - U)(\widehat{\theta}(Z^*) - \theta^*)\right\rangle = \frac{1}{p}\sum_{j=1}^{p} \epsilon_1^T (Z_1^* - U) Z_j^{*T} \epsilon_j,$$

we can bound (H.6) by

$$\mathbb{P}\left(\frac{1}{p}\epsilon_1^T (I_d - UZ_1^{*T})\epsilon_1 \leq -\frac{\bar{\delta}}{4}\|(Z_1^* - U)\theta^*\|^2\right) \tag{H.7}$$

$$+\mathbb{P}\left(\frac{1}{p}\sum_{j=2}^{p} \epsilon_1^T (Z_1^* - U) Z_j^{*T} \epsilon_j \leq -\frac{\bar{\delta}}{4}\|(Z_1^* - U)\theta^*\|^2\right). \tag{H.8}$$

We first bound (H.7) by

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{p}\epsilon_1^T (I_d - UZ_1^{*T})\epsilon_1 \leq -\frac{\bar{\delta}}{4}\|(Z_1^* - U)\theta^*\|^2\right) \\ \leq\; & \mathbb{P}\left(2\|\epsilon_1\|^2 > \frac{\bar{\delta} p}{4}\|(Z_1^* - U)\theta^*\|^2\right) \\ \leq\; & \exp\left(-C\bar{\delta} p\|(Z_1^* - U)\theta^*\|^2\right), \end{aligned}$$

where we use $\|I_d - UZ_1^{*T}\| \leq 2$, under the condition that $p\Delta_{\min}^2/d \to \infty$ and $\bar{\delta}$ tends to zero at a sufficiently slow rate. Note that conditional on ϵ_1, $\frac{1}{p}\sum_{j=2}^{p} \epsilon_1^T (Z_1^* - U) Z_j^{*T} \epsilon_j \sim \mathcal{N}(0, p^{-2}(p-1)\|(Z_1^* - U)^T \epsilon_1\|^2)$, where the variance is upper bounded by $4p^{-1}\|\epsilon_1\|^2$. We then bound (H.8) by

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{p}\sum_{j=2}^{p} \epsilon_1^T (Z_1^* - U) Z_j^{*T} \epsilon_j \leq -\frac{\bar{\delta}}{4}\|(Z_1^* - U)\theta^*\|^2 \Big| \|\epsilon_1\|^2 < d + 2\sqrt{xd} + 2x\right) \\ & +\mathbb{P}\left(\|\epsilon_1\|^2 \geq d + 2\sqrt{xd} + 2x\right) \\ \leq\; & \mathbb{P}\left(\mathcal{N}(0,1) > \frac{\sqrt{p}\frac{\bar{\delta}}{4}\|(Z_1^* - U)\theta^*\|^2}{2\|\epsilon_1\|} \Big| \|\epsilon_1\|^2 < d + 2\sqrt{xd} + 2x\right) + e^{-x} \\ \leq\; & \exp\left(-\frac{p\bar{\delta}^2\|(Z_1^* - U)\theta^*\|^4}{128\left(d + 2\sqrt{xd} + 2x\right)}\right) + e^{-x}. \end{aligned}$$

Choosing $x = \bar{\delta}\|(Z_1^* - U)\theta^*\|^2\sqrt{p}$, we have

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{p}\sum_{j=2}^{p}\epsilon_1^T(Z_1^*-U)Z_j^{*T}\epsilon_j \leq -\frac{\bar{\delta}}{4}\|(Z_1^*-U)\theta^*\|^2\right) \\ \leq\ & \exp\left(-C\frac{\bar{\delta}^2\|(Z_1^*-U)\theta^*\|^4 p}{d+\bar{\delta}\|(Z_1^*-U)\theta^*\|^2\sqrt{p}}\right) + \exp\left(-C\bar{\delta}\|(Z_1^*-U)\theta^*\|^2\sqrt{p}\right) \\ \leq\ & 2\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|(Z_1^*-U)\theta^*\|^2\right), \end{aligned}$$

under the condition that $\sqrt{p}\Delta_{\min}^2/d \to \infty$ and $\bar{\delta}$ tends to zero at a sufficiently slow rate. Combining the above bounds, we have

$$\mathbb{P}\left(\left\langle \epsilon_1, (Z_1^*-U)\widehat{\theta}(Z^*)\right\rangle \leq -\frac{1-\delta}{2}\|(Z_1^*-U)\theta^*\|^2\right) \leq 4\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|(Z_1^*-U)\theta^*\|^2\right).$$

A similar bound holds for $\mathbb{P}\left(\left\langle \epsilon_j, (Z_j^*-U)\widehat{\theta}(Z^*)\right\rangle \leq -\frac{1-\delta}{2}\|(Z_j^*-U)\theta^*\|^2\right)$ for each $j \in [p]$.

Now we are ready to bound $\xi_{\text{ideal}}(\delta)$. We first bound its expectation. We have

$$\begin{aligned} \mathbb{E}\xi_{\text{ideal}}(\delta) &= \sum_{j=1}^{p}\sum_{U\in\mathcal{C}_d\backslash\{Z_j^*\}}\|(Z_j^*-U)\theta^*\|^2\mathbb{P}\left(\left\langle \epsilon_j, (Z_j^*-U)\widehat{\theta}(Z^*)\right\rangle \leq -\frac{1-\delta}{2}\|(Z_j^*-U)\theta^*\|^2\right) \\ &\leq 4\sum_{j=1}^{p}\sum_{U\in\mathcal{C}_d\backslash\{Z_j^*\}}\|(Z_j^*-U)\theta^*\|^2\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\|(Z_j^*-U)\theta^*\|^2\right) \\ &\leq p\exp\left(-(1+o(1))\frac{\Delta_{\min}^2}{8}\right), \end{aligned}$$

where the last inequality uses the condition that $\Delta_{\min}^2/\log d \to \infty$. The desired conclusion is implied by Markov inequality. □

Proof of Proposition 6.1. We adopt the notation $\widehat{Z}_j = Z_j^{(0)}$ in the proof. For any $j \geq 2$,

$$\|Y_1 - \widehat{Z}_j^T Y_j\|^2 \leq \|Y_1 - Z_1^* Z_j^{*T} Y_j\|^2.$$

After rearrangement, we get

$$\begin{aligned} \|(Z_1^* - \widehat{Z}_j^T Z_j^*)\theta^*\|^2 &\leq 2\left|\left\langle \widehat{Z}_j^T\epsilon_j - \epsilon_1, (Z_1^*-\widehat{Z}_j^T Z_j^*)\theta^*\right\rangle\right| + 2\epsilon_1^T\left(\widehat{Z}_j^T - Z_1^* Z_j^{*T}\right)\epsilon_j \\ &\leq 2\|(Z_1^*-\widehat{Z}_j^T Z_j^*)\theta^*\|\frac{\left|\left\langle \widehat{Z}_j^T\epsilon_j - \epsilon_1, (Z_1^*-\widehat{Z}_j^T Z_j^*)\theta^*\right\rangle\right|}{\|(Z_1^*-\widehat{Z}_j^T Z_j^*)\theta^*\|} + 2\left|\epsilon_1^T\left(\widehat{Z}_j^T - Z_1^* Z_j^{*T}\right)\epsilon_j\right|. \end{aligned}$$

This implies

$$\|(Z_1^* - \widehat{Z}_j^T Z_j^*)\theta^*\| \leq 2\frac{\left|\left\langle \widehat{Z}_j^T\epsilon_j - \epsilon_1, (Z_1^*-\widehat{Z}_j^T Z_j^*)\theta^*\right\rangle\right|}{\|(Z_1^*-\widehat{Z}_j^T Z_j^*)\theta^*\|} + \sqrt{2}\sqrt{\left|\epsilon_1^T\left(\widehat{Z}_j^T - Z_1^* Z_j^{*T}\right)\epsilon_j\right|},$$

and consequently,

$$\|(Z_1^* - \widehat{Z}_j^T Z_j^*)\theta^*\|^2 \leq 8\frac{\left|\left\langle \widehat{Z}_j^T \epsilon_j - \epsilon_1, (Z_1^* - \widehat{Z}_j^T Z_j^*)\theta^* \right\rangle\right|^2}{\|(Z_1^* - \widehat{Z}_j^T Z_j^*)\theta^*\|^2} + 8\left|\epsilon_1^T \left(\widehat{Z}_j^T - Z_1^* Z_j^{*T}\right)\epsilon_j\right|$$

$$\leq 16 \max_{U\in\mathcal{C}_d} \frac{\left|\left\langle U^T \epsilon_j - \epsilon_1, (Z_1^* - U^T Z_j^*)\theta^* \right\rangle\right|^2}{2\|(Z_1^* - U^T Z_j^*)\theta^*\|^2} + 8 \max_{U\in\mathcal{C}_d} \left|\epsilon_1^T \left(U^T - Z_1^* Z_j^{*T}\right)\epsilon_j\right|$$

We are going to calculate its expectation. By a standard union bound argument, we have

$$\mathbb{P}\left(\max_{U\in\mathcal{C}_d} \frac{\left|\left\langle U^T \epsilon_j - \epsilon_1, (Z_1^* - U^T Z_j^*)\theta^* \right\rangle\right|^2}{2\|(Z_1^* - U^T Z_j^*)\theta^*\|^2} > t\right) \leq 2d\exp\left(-\frac{t}{2}\right).$$

Integrating up the tail bound, we obtain

$$\mathbb{E}\max_{U\in\mathcal{C}_d} \frac{\left|\left\langle U^T \epsilon_j - \epsilon_1, (Z_1^* - U^T Z_j^*)\theta^* \right\rangle\right|^2}{2\|(Z_1^* - U^T Z_j^*)\theta^*\|^2} \leq 4\log d$$

Note that $\left|\epsilon_1^T U^T \epsilon_j\right| = \|\epsilon_1\| \left|\|\epsilon_1\|^{-1} \epsilon_1^T U^T \epsilon_j\right|$ where $\|\epsilon_1\|^{-1} \epsilon_1^T U^T \epsilon_j \sim \mathcal{N}(0,1)$ is independent of $\|\epsilon_1\|$, we have

$$\mathbb{E}\max_{U\in\mathcal{C}_d} \left|\epsilon_1^T \left(U^T - Z_1^* Z_j^{*T}\right)\epsilon_j\right| \leq 2\mathbb{E}\max_{U\in\mathcal{C}_d}\left|\epsilon_1^T U^T \epsilon_j\right| \leq \mathbb{E}\|\epsilon_1\|\,\mathbb{E}\max_{U\in\mathcal{C}_d}\frac{\left|\epsilon_1^T U^T \epsilon_j\right|}{\|\epsilon_1\|} \lesssim \sqrt{d\log d},$$

where in the last inequality we integrate up the tail bound of $\mathbb{P}(\max_{U\in\mathcal{C}_d}\left|\epsilon_1^T U^T \epsilon_j\right| / \|\epsilon_1\| > t) \leq 2d\exp\left(-t^2/2\right)$. Hence,

$$\mathbb{E}\|(Z_1^* - \widehat{Z}_j^T Z_j^*)\theta^*\|^2 \lesssim \mathbb{E}\max_{U\in\mathcal{C}_d} \frac{\left|\left\langle U^T \epsilon_j - \epsilon_1, (Z_1^* - U^T Z_j^*)\theta^* \right\rangle\right|^2}{2\|(Z_1^* - U^T Z_j^*)\theta^*\|^2} + \mathbb{E}\max_{U\in\mathcal{C}_d}\left|\epsilon_1^T \left(U^T - Z_1^* Z_j^{*T}\right)\epsilon_j\right| \lesssim \sqrt{d\log d}.$$

Then,

$$\sum_{j=1}^p \mathbb{E}\|(\widehat{Z}_j Z_1^* - Z_j^*)\theta^*\|^2 = \sum_{j=1}^p \mathbb{E}\|(Z_1^* - \widehat{Z}_j^T Z_j^*)\theta^*\|^2 \lesssim p\sqrt{d\log d}.$$

The desired conclusion is implied by Markov inequality. □

I Proofs in Section 7, Appendix B and Appendix C

I.1 Proofs of $\mathbb{Z}_2$ Synchronization

In this section, we present the proofs of Lemma 7.1, Lemma 7.2 and Proposition 7.1. The conclusions of Theorem 7.2 and Corollary 7.1 are direct consequences of Theorem 3.1, and their proofs are omitted. We first need a technical lemma that bound the operator nomr of a Gaussian random matrix. The following lemma is a standard result that can be found in [16].

Lemma I.1. *Consider a matrix $W = (w_{ij}) \in \mathbb{R}^{p\times p}$ with $W_{ij} = W_{ji} \sim \mathcal{N}(0,1)$ independently for all $i < j$ and $W_{ii} = 0$ for all i. There exists some constant $C > 0$, such that*

$$\mathbb{P}\left(\|W\| > C\sqrt{p+x}\right) \leq e^{-x},$$

for any $x > 0$.

Proof of Lemma 7.1. We can write

$$\widehat{\beta}(z) = \frac{z^T(\lambda^* z^* (z^*)^T + W)z}{p^2} z = \frac{\lambda^* |z^T z^*|^2}{p^2} z + \frac{z^T W z}{p^2} z,$$

where W is a symmetric matrix with $W_{ij} \sim \mathcal{N}(0,1)$ on the off-diagonal and $W_{ii} = 0$ on the diagonal. Then, we have

$$\widehat{\beta}(z^*) = \lambda^* z^* + \frac{z^{*T} W z^*}{p^2} z^* = \beta^* + \frac{z^{*T} W z^*}{p^2} z^*.$$

Using triangle inequality, we have

$$\|\widehat{\beta}(z) - \widehat{\beta}(z^*)\| \leq \left\| \frac{\lambda^* |z^T z^*|^2}{p^2} z - \beta^* \right\| + \left\| \frac{z^T W z}{p^2} z - \frac{z^{*T} W z^*}{p^2} z^* \right\|. \tag{I.1}$$

We will bound the two terms on the right hand side of (I.1) separately. The first term can be bounded as

$$\begin{aligned}
\left\| \frac{\lambda^* |z^T z^*|^2}{p^2} z - \beta^* \right\| &\leq \lambda^* \left| \frac{|z^T z^*|^2}{p^2} - 1 \right| \|z\| + \lambda^* \|z - z^*\| \\
&\leq \lambda^* \frac{\|z - z^*\|^2}{\sqrt{p}} + \lambda^* \|z - z^*\| \qquad \text{(I.2)} \\
&\leq 3\lambda^* \|z - z^*\|,
\end{aligned}$$

where the inequality (I.2) is by

$$|p^2 - |z^T z^*|^2| = (p - |z^T z^*|)(p + |z^T z^*|) \leq 2p(p - |z^T z^*|) \leq p\|z - z^*\|^2. \tag{I.3}$$

For the second term of (I.1), we have

$$\begin{aligned}
&\left\| \frac{z^T W z}{p^2} z - \frac{z^{*T} W z^*}{p^2} z^* \right\| \\
\leq\; & \frac{|\langle W, zz^T - z^*(z^*)^T \rangle|}{p^2} \|z\| + \frac{|z^{*T} W z^*|}{p^2} \|z - z^*\| \\
\leq\; & \sqrt{2}\|W\| \|zz^T - z^* z^{*T}\|_{\mathrm{F}} \frac{\|z\|}{p^2} + \frac{|z^{*T} W z^*|}{p^2} \|z - z^*\| \qquad \text{(I.4)} \\
\leq\; & \frac{3\|W\|}{p} \|z - z^*\| \qquad \text{(I.5)} \\
\lesssim\; & \frac{\|z - z^*\|}{\sqrt{p}}. \qquad \text{(I.6)}
\end{aligned}$$

The inequality (I.4) is by applying SVD to the rank-2 matrix $zz^T - z^*z^{*T} = \sum_{l=1}^2 d_l u_l u_l^T$ so that $|\langle W, zz^T - z^*(z^*)^T \rangle|$ is bounded by

$$\sum_{l=1}^2 |d_l||u_l^T W u_l| \le \|W\|(|d_1| + |d_2|) \le \sqrt{2}\|W\|\sqrt{d_1^2 + d_2^2} = \sqrt{2}\|W\|\|zz^T - z^*(z^*)^T\|_{\mathrm{F}}.$$

To obtain (I.5), we have used $\|zz^T - z^*(z^*)^T\|_{\mathrm{F}}^2 = 2|p^2 - |z^T z^*|^2| \le p\|z - z^*\|^2$ according to (I.3) and $|z^{*T} W z^*| \le p\|W\|$. Finally, (I.6) is by Lemma I.1. Combining the two bounds for the two terms on the right hand side of (I.1), we have

$$\|\widehat{\beta}(z) - \widehat{\beta}(z^*)\| \lesssim \left(\lambda^* + p^{-1/2}\right)\|z - z^*\|. \tag{I.7}$$

Now we are ready to bound (65)-(67). For (65), we have

$$\begin{aligned}
&\sum_{j=1}^p \max_{b \in \{-1,1\} \setminus \{z_j^*\}} \frac{F_j(z_j^*, b; z)^2 \left\|\mu_j(B^*, b) - \mu_j(B^*, z_j^*)\right\|^2}{\Delta_j(z_j^*, b)^4 \ell(z, z^*)} \\
&= \sum_{j=1}^p \frac{F_j(z_j^*, -z_j^*; z)^2 \left\|\mu_j(B^*, -z_j^*) - \mu_j(B^*, z_j^*)\right\|^2}{\Delta_j(z_j^*, -z_j^*)^4 \ell(z, z^*)} \\
&\lesssim \sum_{j=1}^p \frac{\left|\left\langle \epsilon_j, \widehat{\beta}(z) - \widehat{\beta}(z^*) \right\rangle\right|^2}{\|\beta^*\|^2 \ell(z, z^*)} \\
&\le \frac{\|\widehat{\beta}(z) - \widehat{\beta}(z^*)\|^2}{\lambda^{*2} p \ell(z, z^*)} \left\|\sum_{j=1}^p \epsilon_j \epsilon_j^T\right\| \\
&\lesssim \frac{1}{p\lambda^{*2}} + \frac{1}{p^2 \lambda^{*4}},
\end{aligned}$$

where the last inequality is by (64), (I.7) and (E.3). For (66), for any $T \subset [p]$, we have

$$\begin{aligned}
&\frac{\tau}{4\Delta_{\min}^2 |T|} \sum_{j \in T} \max_{b \in \{-1,1\} \setminus \{z_j^*\}} \frac{G_j(z_j^*, b; z)^2 \left\|\mu_j(B^*, b) - \mu_j(B^*, z_j^*)\right\|^2}{\Delta_j(z_j^*, b)^4 \ell(z, z^*)} \\
&= \frac{\tau}{4\Delta_{\min}^2 |T|} \sum_{j \in T} \frac{G_j(z_j^*, -z_j^*; z)^2 \left\|\mu_j(B^*, -z_j^*) - \mu_j(B^*, z_j^*)\right\|^2}{\Delta_j(z_j^*, -z_j^*)^4 \ell(z, z^*)} \\
&\lesssim \frac{\tau}{4\Delta_{\min}^2 |T|} \sum_{j \in T} \frac{\left|\left\langle \beta^*, \widehat{\beta}(z) - \widehat{\beta}(z^*) \right\rangle\right|^2}{\|\beta^*\|^2 \ell(z, z^*)} \\
&\lesssim \frac{\tau \|\widehat{\beta}(z) - \widehat{\beta}(z^*)\|^2}{\|\beta^*\|^2 \ell(z, z^*)} \\
&\lesssim \frac{\tau}{\lambda^{*2} p^2} + \frac{\tau}{\lambda^{*4} p^3},
\end{aligned}$$

where the last inequality is by (I.7). Finally, for (67), since $\|\widehat{\beta}(z^*)-\beta^*\| = \frac{|z^{*T}Wz^*|}{p^2}\|z^*\| \lesssim 1$ by Lemma I.1, we have

$$\max_{b\neq z_j^*}\frac{\left|H_j(z_j^*,b)\right|}{\Delta_j(z_j^*,b)^2} = \frac{\left|H_j(z_j^*,-z_j^*)\right|}{\Delta_j(z_j^*,-z_j^*)^2} = \frac{\left|\left\langle \beta^*,\widehat{\beta}(z^*)-\beta^*\right\rangle\right|}{2\|\beta^*\|^2} \leq \frac{1}{2}\frac{\|\widehat{\beta}(z^*)-\beta^*\|}{\|\beta^*\|} \lesssim \frac{1}{\sqrt{p\lambda^{*2}}},$$

for any $j\in[p]$. The proof is complete. □

Proof of Lemma 7.2. Without loss of generality, assume $z_1^*=1$, and then for $b\neq z_1^*$, we have

$$\begin{aligned}
& \mathbb{P}\left(\left\langle W_1,\widehat{\beta}(z^*)(z_1^*-b)\right\rangle \leq -\frac{1-\delta}{2}4\|\beta^*\|^2\right) \\
= \;& \mathbb{P}\left(\left\langle W_1,\widehat{\beta}(z^*)\right\rangle \leq (1-\delta)p\lambda^{*2}\right) \\
\leq \;& \mathbb{P}\left(\langle W_1,\beta^*\rangle \leq -(1-\delta-\delta')p\lambda^{*2}\right) && \text{(I.8)} \\
& +\mathbb{P}\left(\left\langle W_1,\widehat{\beta}(z^*)-\beta^*\right\rangle \leq -\delta' p\lambda^{*2}\right), && \text{(I.9)}
\end{aligned}$$

for some sequence $\delta'=\delta'_p$ to be determined later. For (I.8), a standard Gaussian tail bound gives

$$\begin{aligned}
& \mathbb{P}\left(\langle W_1,\beta^*\rangle \leq -(1-\delta-\delta')p\lambda^{*2}\right) \\
\leq \;& \mathbb{P}\left(\mathcal{N}(0,p\lambda^{*2}) > (1-\delta-\delta')p\lambda^{*2}\right) \\
\leq \;& \exp\left(-\frac{(1-\delta-\bar{\delta})^2p\lambda^{*2}}{2}\right).
\end{aligned}$$

For (I.9), note that we have

$$\widehat{\beta}(z^*)-\beta^* = \frac{z^{*T}Wz^*}{p^2}z^*,$$

and thus

$$\begin{aligned}
& \mathbb{P}\left(\left\langle W_1,\widehat{\beta}(z^*)-\beta^*\right\rangle \leq -\bar{\delta}p\lambda^{*2}\right) \\
= \;& \mathbb{P}\left(\frac{z^{*T}Wz^*}{p^2}\langle W_1,z^*\rangle \leq -\bar{\delta}p\lambda^{*2}\right) \\
\leq \;& \mathbb{P}\left(\|W\||\langle W_1,z^*\rangle| \geq \bar{\delta}p^2\lambda^{*2}\right) \\
\leq \;& \mathbb{P}\left(\|W\||\langle W_1,z^*\rangle| \geq \bar{\delta}p^2\lambda^{*2}, \|W\|\leq C\sqrt{p+x}\right) + \mathbb{P}(\|W\|>C\sqrt{p+x}) \\
\leq \;& \mathbb{P}\left(C\sqrt{p+x}|\langle W_1,z^*\rangle| \geq \bar{\delta}p^2\lambda^{*2}\right) + e^{-x} \\
\leq \;& 2\exp\left(-C'\frac{\bar{\delta}^2p^3\lambda^{*4}}{p+x}\right) + e^{-x}.
\end{aligned}$$

Take $x=\bar{\delta}p^{3/2}\lambda^{*2}$, and we have

$$\begin{aligned}
& \mathbb{P}\left(\left\langle W_1,\widehat{\beta}(z^*)-\beta^*\right\rangle \leq -\bar{\delta}p\lambda^{*2}\right) \\
\leq \;& 2\exp\left(-C'\bar{\delta}^2p^2\lambda^{*4}\right) + 3\exp\left(-C'\bar{\delta}p^{3/2}\lambda^{*2}\right) \\
\leq \;& 5\exp\left(-\frac{(1-\delta-\bar{\delta})^2p\lambda^{*2}}{2}\right),
\end{aligned}$$

where the last inequality uses the condition that $p\lambda^{*2} \to \infty$ and the fact that $\bar{\delta}$ tends to zero at a sufficiently slow rate. Combining the above bounds, we obtain

$$\mathbb{P}\left(\left\langle W_1, \widehat{\beta}(z^*)(z_1^* - b)\right\rangle \leq -\frac{1-\delta}{2}4\|\beta^*\|^2\right) \leq 6\exp\left(-\frac{(1-\delta-\bar{\delta})^2 p\lambda^{*2}}{2}\right).$$

A similar bound holds for $\mathbb{P}\left(\left\langle W_j, \widehat{\beta}(z^*)(z_j^* - b)\right\rangle \leq -\frac{1-\delta}{2}4\|\beta^*\|^2\right)$ for each $j \in [p]$. This implies

$$\begin{aligned}
\mathbb{E}\xi_{\text{ideal}}(\delta) &= 4p\lambda^{*2}\sum_{j=1}^{p}\mathbb{P}\left(2\left\langle W_j, \widehat{\beta}(z^*)z_j^*\right\rangle \leq -\frac{1-\delta}{2}4\|\beta^*\|^2\right) \\
&\leq 24p^2\lambda^{*2}\exp\left(-\frac{(1-\delta-\bar{\delta})^2 p\lambda^{*2}}{2}\right) \\
&\leq p\exp\left(-(1+o(1))\frac{p\lambda^{*2}}{2}\right),
\end{aligned}$$

where the last inequality uses $p\lambda^{*2} \to \infty$. The desired conclusion is implied by Markov inequality. □

Proof of Proposition 7.1. We can write $z^{(0)} = \mathrm{argmin}_{z\in\{-1,1\}^p}\|z - \sqrt{p}\widetilde{u}\|^2$. This implies

$$\|z^{(0)} - z^*\|^2 \leq 2\|\sqrt{p}\widetilde{u} - z^*\|^2 + 2\|z^{(0)} - \sqrt{p}\widetilde{u}\|^2 \leq 4\|\sqrt{p}\widetilde{u} - z^*\|^2.$$

Similarly, we also have $\|z^{(0)} + z^*\|^2 \leq 4\|\sqrt{p}\widetilde{u} + z^*\|^2$. Thus,

$$\|z^{(0)} - z^*\|^2 \wedge \|z^{(0)} + z^*\|^2 \leq 4\left(\|\sqrt{p}\widetilde{u} - z^*\|^2 \wedge \|\sqrt{p}\widetilde{u} + z^*\|^2\right). \tag{I.10}$$

Since $\widetilde{u}$ is the leading eigenvector of Y and $z^*/\sqrt{p}$ is the leading eigenvector of $\lambda^* z^* z^{*T}$, by Davis-Kahan theorem, we have

$$\left\|\widetilde{u} - \frac{z^*}{\sqrt{p}}\right\| \wedge \left\|\widetilde{u} + \frac{z^*}{\sqrt{p}}\right\| \lesssim \frac{\|W\|}{p|\lambda^*|} \lesssim \frac{1}{\sqrt{p\lambda^{*2}}},$$

where the last inequality above is by Lemma I.1. Finally, by (I.10), we have

$$\ell(z^{(0)}, z^*) \wedge \ell(z^{(0)}, -z^*) \lesssim p,$$

with high probability. □

I.2 Proofs of $\mathbb{Z}/k\mathbb{Z}$ Synchronization

In this section, we present the proofs of Theorem B.1, Lemma B.1, Lemma B.2 and Proposition B.1. The conclusion of Theorem B.2 and Corollary B.1 are direct consequences of Theorem 3.1, and thus their proofs are omitted. We first present a technical lemma.

Lemma I.2. *Consider i.i.d. random variables $w_{ij} \sim \mathcal{N}(0,1)$ for $1 \leq i \neq j \leq p$ and some $z^* \in (\mathbb{Z}/k\mathbb{Z})^p$. For any constant $C' > 0$, there exists some constant $C > 0$ only depending on C' such that*

$$\max_{z \in (\mathbb{Z}/k\mathbb{Z})^p} \left| \frac{\sum_{i \neq j} w_{ij} \left[(z_i \circ z_j^{-1}) - (z_i^* \circ z_j^{*-1}) \right]}{\sqrt{\sum_{i \neq j} \left[(z_i \circ z_j^{-1}) - (z_i^* \circ z_j^{*-1}) \right]^2}} \right| \leq C\sqrt{p \log k}, \tag{I.11}$$

with probability at least $1 - e^{-C'p\log k}$.

Proof. The result is implied by a standard Gaussian tail bound and a union bound argument. □

Proof of Theorem B.1. Given the similarity between $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}_2$, the proof for $k = 2$ is very similar to the proof of Theorem 7.1 and thus is omitted here. We only need to consider the case $k \geq 3$. Let $\rho = o(1)$ tends to zero at a sufficiently slow rate. Consider the parameter space

$$\mathcal{Z} = \{z : z_j = 2 \text{ for all } 1 \leq j \leq (1-\rho)p \text{ and } z_j \in \{0,1\} \text{ for all } (1-\rho)p < j \leq p\}.$$

Then, by the same argument that leads to (H.1), we have

$$\begin{aligned} &\inf_{\widehat{z}} \sup_{z^*} \mathbb{E} \min_{a \in \mathbb{Z}/k\mathbb{Z}} \left(\frac{1}{p} \sum_{j=1}^p \mathbf{1}_{\{\widehat{z}_j \neq z_j^* \circ a^{-1}\}} \right) \\ \geq\ & \frac{1}{p} \sum_{j > (1-\rho)p} \text{ave}_{z^*_{-j}} \inf_{\widehat{z}_j} \left(\frac{1}{2} \mathbb{P}_{(z_j^*=1, z^*_{-j})}(\widehat{z}_j \neq 1) + \frac{1}{2} \mathbb{P}_{(z_j^*=0, z^*_{-j})}(\widehat{z}_j \neq 0) \right). \end{aligned}$$

Note that the quantity

$$\inf_{\widehat{z}_j} \left(\frac{1}{2} \mathbb{P}_{(z_j^*=1, z^*_{-j})}(\widehat{z}_j \neq 1) + \frac{1}{2} \mathbb{P}_{(z_j^*=0, z^*_{-j})}(\widehat{z}_j \neq 0) \right)$$

is the optimal error for the hypothesis testing problem,

$$\begin{aligned} H_0: &\quad Y_{ij} \sim \mathcal{N}(\lambda^*[(z_i^* - 1) \bmod k], 1) \text{ and } Y_{ji} \sim \mathcal{N}(\lambda^*[(1 - z_i^*) \bmod k], 1) \text{ for } i \in [p]\backslash\{j\}, \\ H_1: &\quad Y_{ij} \sim \mathcal{N}(\lambda^*[(z_i^* - 0) \bmod k], 1) \text{ and } Y_{ji} \sim \mathcal{N}(\lambda^*[(0 - z_i^*) \bmod k], 1) \text{ for } i \in [p]\backslash\{j\}. \end{aligned}$$

Note that

$$\lambda^{*-1} |[(z_i^* - 1) \bmod k] - [(z_i^* - 0) \bmod k]| \begin{cases} = 1, & \text{if } i \leq (1-\rho)p \\ \leq k-1, & \text{o.w.} \end{cases}$$

and same result holds for $|[(1 - z_i^*) \bmod k] - [(0 - z_i^*) \bmod k]|$. We have

$$\sum_{j \neq i} \left(|\mathbb{E}_{H_0} Y_{ij} - \mathbb{E}_{H_1} Y_{ij}| + |\mathbb{E}_{H_0} Y_{ji} - \mathbb{E}_{H_1} Y_{ji}| \right) \leq 2\lambda^* \left((1-\rho)p + (k-1)\rho p \right) \leq 2\lambda^*(1+\rho k)p.$$

By Neyman-Pearson lemma, the testing error can be lower bounded by $\mathbb{P}(\mathcal{N}(0,1) \geq \lambda^*(1+\rho k)p/\sqrt{2(p-1)}$. Therefore, we have

$$\begin{aligned}\inf_{\widehat{z}} \sup_{z^*} \mathbb{E} \min_{a\in\mathbb{Z}/k\mathbb{Z}} \left(\frac{1}{p}\sum_{j=1}^{p} \mathbf{1}_{\{\widehat{z}_j \neq z_j^* \circ a^{-1}\}}\right) &\geq \rho\mathbb{P}(\mathcal{N}(0,1) \geq \lambda^*(1+\rho k)p/\sqrt{2(p-1)}) \\ &= \exp\left(-(1+o(1))\frac{p\lambda^{*2}}{4}\right),\end{aligned}$$

under the condition that $p\lambda^{*2} \to \infty$ and ρ tends to zero at a sufficiently slow rate. When $p\lambda^{*2} = O(1)$, we can take $\rho = 1/2$ instead and obtain a constant lower bound. □

Now we are ready to state the proofs of Lemma B.1 and Lemma B.2. Note that under the setting of $\mathbb{Z}/k\mathbb{Z}$ synchronization, the error terms are

$$\begin{aligned}F_j(a,b,z) &= \left\langle \epsilon_j, \widehat{\lambda}(z^*)\left(z^*\circ a^{-1} - z^*\circ b^{-1}\right) - \widehat{\lambda}(z)\left(z\circ a^{-1} - z\circ b^{-1}\right)\right\rangle, \\ G_j(a,b,z) &= \frac{1}{2}\left(\|\lambda^*(z^*\circ a^{-1}) - \widehat{\lambda}(z)(z\circ a^{-1})\|^2 - \|\lambda^*(z^*\circ a^{-1}) - \widehat{\lambda}(z^*)(z^*\circ a^{-1})\|^2\right) \\ &\quad -\frac{1}{2}\left(\|\lambda^*(z^*\circ a^{-1}) - \widehat{\lambda}(z)(z\circ b^{-1})\|^2 - \|\lambda^*(z^*\circ a^{-1}) - \widehat{\lambda}(z^*)(z^*\circ b^{-1})\|^2\right), \\ H_j(a,b) &= \frac{1}{2}\|\lambda^*(z^*\circ a^{-1}) - \widehat{\lambda}(z^*)(z^*\circ a^{-1})\|^2 \\ &\quad -\frac{1}{2}\left(\|\lambda^*(z^*\circ a^{-1}) - \widehat{\lambda}(z^*)(z^*\circ b^{-1})\|^2 - \|\lambda^*(z^*\circ a^{-1}) - \lambda^*(z^*\circ b^{-1})\|^2\right),\end{aligned}$$

where we have used the notation $\epsilon_j = W_j$ for the jth column of the error matrix, and $z\circ a^{-1}$ stands for the vector $\{z_i \circ a^{-1}\}_{i\in[p]}$.

Proof of Lemma B.1. Throughout the proof, we define $h(z,z^*) = \sum_{j\in[p]} \mathbf{1}_{\{z_j \neq z_j^*\}}, \forall z$. Then we have

$$\sum_{i\neq j}\left[(z_i^*\circ z_j^{*-1}) - (z_i\circ z_j^{-1})\right]^2 \leq k^2\sum_{i,j}\left(\mathbf{1}_{\{z_i^*\neq z_i\}} + \mathbf{1}_{\{z_j^*\neq z_j\}}\right) \leq 2pk^2 h(z,z^*). \tag{I.12}$$

In addition, (B.11) is equivalent to $h(z,z^*) \leq \frac{1}{p\lambda^{*2}}\ell(z,z^*)$.

Under the assumption that $\max_{a\in\mathbb{Z}/k\mathbb{Z}}\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}} \leq (1-\alpha)p$ where $\alpha > 0$ is a constant, we have

$$\sum_{i\neq j}(z_i^*\circ z_j^{*-1})^2 \geq \sum_{i\neq j}\mathbf{1}_{\{z_i^*\neq z_j^*\}} \geq \alpha p^2. \tag{I.13}$$

Similarly, since for any z such that $h(z,z^*) \leq \alpha p/2$ we have $\max_{a\in\mathbb{Z}/k\mathbb{Z}}\sum_{j=1}^{p}\mathbf{1}_{\{z_j=a\}} \leq (1-\alpha/2)p$, we also obtain

$$\sum_{i\neq j}(z_i\circ z_j^{-1})^2 \geq \sum_{i\neq j}\mathbf{1}_{\{z_i\neq z_j\}} \geq \alpha p^2/2. \tag{I.14}$$

Now we will derive a bound for $|\widehat{\lambda}(z) - \widehat{\lambda}(z^*)|$, where $\widehat{\lambda}(z) = \frac{\sum_{i \neq j} Y_{ij}(z_i \circ z_j^{-1})}{\sum_{i \neq j}(z_i \circ z_j^{-1})^2}$. We can write $\widehat{\lambda}(z)$ as $\widehat{\lambda}(z) = \lambda(z) + \bar{w}(z)$, where

$$\lambda(z) = \lambda^* \frac{\sum_{i \neq j}(z_i^* \circ z_j^{*-1})(z_i \circ z_j^{-1})}{\sum_{i \neq j}(z_i \circ z_j^{-1})^2},$$

and

$$\bar{w}(z) = \frac{\sum_{i \neq j} w_{ij}(z_i \circ z_j^{-1})}{\sum_{i \neq j}(z_i \circ z_j^{-1})^2}.$$

Then,

$$|\widehat{\lambda}(z) - \widehat{\lambda}(z^*)| \leq |\lambda(z) - \lambda(z^*)| + |\bar{w}(z) - \bar{w}(z^*)|, \tag{I.15}$$

and we will bound the two terms on the right hand side of the above inequality separately. For the first term of (I.15), we have

$$\begin{aligned}
|\lambda(z) - \lambda(z^*)| &= |\lambda^*| \frac{\left| \sum_{i \neq j} \left[(z_i^* \circ z_j^{*-1}) - (z_i \circ z_j^{-1}) \right] (z_i \circ z_j^{-1}) \right|}{\sum_{i \neq j}(z_i \circ z_j^{-1})^2} \\
&\leq |\lambda^*| \frac{\sqrt{\sum_{i \neq j} \left[(z_i^* \circ z_j^{*-1}) - (z_i \circ z_j^{-1}) \right]^2}}{\sqrt{\sum_{i \neq j}(z_i \circ z_j^{-1})^2}} \\
&\leq \frac{\sqrt{2}|\lambda^*|}{\sqrt{\alpha} p} \sqrt{\sum_{i \neq j} \left[(z_i^* \circ z_j^{*-1}) - (z_i \circ z_j^{-1}) \right]^2} \\
&\lesssim \frac{|\lambda^*| k}{p} \sqrt{p h(z, z^*)}, \qquad (\text{I.16})
\end{aligned}$$

where we have used (I.12) and (I.14). For the second term of (I.15), we have

$$\begin{aligned}
|\bar{w}(z) - \bar{w}(z^*)| \leq{}& \left| \frac{\sum_{i \neq j} w_{ij} \left[(z_i \circ z_j^{-1}) - (z_i^* \circ z_j^{*-1}) \right]}{\sum_{i \neq j}(z_i \circ z_j^{-1})^2} \right| \qquad (\text{I.17}) \\
&+ \left| \frac{1}{\sum_{i \neq j}(z_i \circ z_j^{-1})^2} - \frac{1}{\sum_{i \neq j}(z_i^* \circ z_j^{*-1})^2} \right| \left| \sum_{i \neq j} w_{ij}(z_i^* \circ z_j^{*-1}) \right|. \qquad (\text{I.18})
\end{aligned}$$

We can bound (I.17) by

$$\begin{aligned}
&\frac{\sqrt{\sum_{i \neq j} \left[(z_i \circ z_j^{-1}) - (z_i^* \circ z_j^{*-1}) \right]^2}}{\sum_{i \neq j}(z_i \circ z_j^{-1})^2} \left| \frac{\sum_{i \neq j} w_{ij} \left[(z_i \circ z_j^{-1}) - (z_i^* \circ z_j^{*-1}) \right]}{\sqrt{\sum_{i \neq j} \left[(z_i \circ z_j^{-1}) - (z_i^* \circ z_j^{*-1}) \right]^2}} \right| \\
\lesssim{}& \frac{\sqrt{k^2 (\log k) h(z, z^*)}}{p},
\end{aligned}$$

where we have used (I.12), (I.14) and Lemma I.2. For (I.18), we have

$$\begin{aligned}
& \left|\frac{1}{\sum_{i\neq j}(z_i\circ z_j^{-1})^2}-\frac{1}{\sum_{i\neq j}(z_i^*\circ z_j^{*-1})^2}\right| \\
= & \left|\frac{\sum_{i\neq j}\left[(z_i\circ z_j^{-1})^2-(z_i^*\circ z_j^{*-1})^2\right]}{\left(\sum_{i\neq j}(z_i\circ z_j^{-1})^2\right)\left(\sum_{i\neq j}(z_i^*\circ z_j^{*-1})^2\right)}\right| \\
\lesssim & \frac{k^2h(z,z^*)}{p^3},
\end{aligned}$$

and thus

$$\left|\frac{1}{\sum_{i\neq j}(z_i\circ z_j^{-1})^2}-\frac{1}{\sum_{i\neq j}(z_i^*\circ z_j^{*-1})^2}\right|\left|\sum_{i\neq j}w_{ij}(z_i^*\circ z_j^{*-1})\right|\lesssim\frac{k^3\sqrt{\log p}h(z,z^*)}{p^2},$$

with probability at least $1-p^{-C'}$ by $\sum_{i\neq j}(z_i^*\circ z_j^{*-1})^2\lesssim k^2p^2$ and a standard Gaussian tail bound. Therefore, we can bound (I.17) by $\frac{\sqrt{k^2(\log k)h(z,z^*)}}{p}+\frac{k^3\sqrt{\log p}h(z,z^*)}{p^2}$ up to a constant. Together with (I.16), we have

$$|\widehat{\lambda}(z)-\widehat{\lambda}(z^*)|\lesssim\frac{|\lambda^*|k}{p}\sqrt{ph(z,z^*)}+\frac{\sqrt{k^2(\log k)h(z,z^*)}}{p}+\frac{k^3\sqrt{\log p}h(z,z^*)}{p^2}. \tag{I.19}$$

Moreover, we also have $\widehat{\lambda}(z^*)-\lambda^*=\bar{w}(z^*)\sim\mathcal{N}(0,1/\sum_{i\neq j}(z_i^*\circ z_j^{*-1})^2)$ where the variance is smaller than $1/\alpha p^2$ according to (I.14). Thus,

$$|\widehat{\lambda}(z^*)-\lambda^*|\lesssim\frac{1}{\sqrt{p}}, \tag{I.20}$$

with probability $1-e^{-C'p}$. This implies $\left|\widehat{\lambda}(z^*)\right|\lesssim|\lambda^*|+\frac{1}{\sqrt{p}}$ and

$$\|\lambda^*z^*\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^2\lesssim k^2. \tag{I.21}$$

Using (I.19) and (I.20), we have

$$\begin{aligned}
& \|\widehat{\lambda}(z)z\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\| \\
\leq & |\widehat{\lambda}(z)-\widehat{\lambda}(z^*)|\|z\circ a^{-1}\|+|\widehat{\lambda}(z^*)|\|z\circ a^{-1}-z^*\circ a^{-1}\| \\
\lesssim & \left(|\lambda^*|k\sqrt{k}\sqrt{h(z,z^*)}+\frac{\sqrt{k^3\log kh(z,z^*)}}{\sqrt{p}}+\frac{k^3\sqrt{k}h(z,z^*)}{p\sqrt{p}}\right)+\left(|\lambda^*|+\frac{1}{\sqrt{p}}\right)\sqrt{kh(z,z^*)} \\
\lesssim & |\lambda^*|k\sqrt{k}\sqrt{h(z,z^*)}, \qquad\qquad (I.22)
\end{aligned}$$

under the condition that $\frac{p\lambda^{*2}}{k^4}\to\infty$.

Now we are ready to bound (B.6)-(B.8). For (B.6), we have

$$\begin{aligned}
&\sum_{j=1}^{p}\max_{b\in\mathbb{Z}/k\mathbb{Z}\backslash\{z_j^*\}}\frac{F_j(z_j^*,b;z)^2\left\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\right\|^2}{\Delta_j(z_j^*,b)^4\ell(z,z^*)}\\
=&\sum_{j=1}^{p}\max_{b\in\mathbb{Z}/k\mathbb{Z}\backslash\{z_j^*\}}\frac{F_j(z_j^*,b;z)^2}{\Delta_j(z_j^*,b)^2\ell(z,z^*)}\\
\leq&\sum_{a\in\mathbb{Z}/k\mathbb{Z}}\sum_{b\in\mathbb{Z}/k\mathbb{Z}\backslash\{a\}}\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}}\frac{F_j(a,b;z)^2}{(p\lambda^{*2})^2h(z,z^*)}\\
\leq&\frac{\max_{a\in\mathbb{Z}/k\mathbb{Z}}\|\widehat{\lambda}(z)z\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^2}{(p\lambda^{*2})^2h(z,z^*)}\sum_{a\in\mathbb{Z}/k\mathbb{Z}}\left\|\sum_{j=1}^{p}\mathbf{1}_{\{z_j^*=a\}}\epsilon_j\epsilon_j^T\right\|\\
\lesssim&\frac{k^4}{p\lambda^{*2}},
\end{aligned}$$

by (I.22) and (E.3). For (B.7), we have

$$\begin{aligned}
&\frac{\tau}{4\Delta_{\min}^2|T|}\sum_{j\in T}\max_{b\in\mathbb{Z}/k\mathbb{Z}\backslash\{z_j^*\}}\frac{G_j(z_j^*,b;z)^2\left\|\mu_j(B^*,b)-\mu_j(B^*,z_j^*)\right\|^2}{\Delta_j(z_j^*,b)^2\ell(z,z^*)}\\
=&\frac{\tau}{4\Delta_{\min}^2|T|}\sum_{j\in T}\max_{b\in\mathbb{Z}/k\mathbb{Z}\backslash\{z_j^*\}}\frac{G_j(z_j^*,b;z)^2}{\Delta_j(z_j^*,b)^2\ell(z,z^*)}\\
\lesssim&\frac{\tau}{(p\lambda^{*2})^3h(z,z^*)}\max_{a\in\mathbb{Z}/k\mathbb{Z}}\|\widehat{\lambda}(z)z\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^4\\
&+\frac{\tau}{(p\lambda^{*2})^3h(z,z^*)}\left(\max_{a\in\mathbb{Z}/k\mathbb{Z}}\|\widehat{\lambda}(z)z\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^2\right)\left(\max_{a\in\mathbb{Z}/k\mathbb{Z}}\|\lambda^*z^*\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^2\right)\\
&+\frac{\tau}{(p\lambda^{*2})^2h(z,z^*)}\max_{a\in\mathbb{Z}/k\mathbb{Z}}\|\widehat{\lambda}(z)z\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^2\\
\lesssim&\frac{\tau k^6}{p^2\lambda^{*2}},
\end{aligned}$$

by (I.22) and (I.21). Finally, for (B.8), we have

$$\begin{aligned}
\frac{|H_j(z_j^*,a)|}{\Delta_j(z_j^*,a)^2}\lesssim&\frac{\max_{a\in\mathbb{Z}/k\mathbb{Z}}\|\lambda^*z^*\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^2}{p\lambda^{*2}}\\
&+\sqrt{\frac{\max_{a\in\mathbb{Z}/k\mathbb{Z}}\|\lambda^*z^*\circ a^{-1}-\widehat{\lambda}(z^*)z^*\circ a^{-1}\|^2}{p\lambda^{*2}}}\\
\lesssim&\frac{k^2}{p\lambda^{*2}}+\sqrt{\frac{k^2}{p\lambda^{*2}}},
\end{aligned}$$

by (I.21). The proof is complete. □

Proof of Lemma B.2. For any z_1^* and $b \neq z_1^*$, we have

$$\begin{aligned}
&\mathbb{P}\left(\sum_{j=1}^{p} w_{j1}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})\widehat{\lambda}(z^*) \leq -\frac{1-\delta}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right) \\
\leq\ &\mathbb{P}\left(\sum_{j=1}^{p} w_{j1}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})\lambda^* \leq -\frac{1-\delta-\bar{\delta}}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right) \\
&+\mathbb{P}\left(\sum_{j=1}^{p} w_{j1}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})(\widehat{\lambda}(z^*) - \lambda^*) \leq -\frac{\bar{\delta}}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right),
\end{aligned}$$

where the sequence $\bar{\delta} = \bar{\delta}_p$ is to be determined later. For the first term, by a standard Gaussian tail bound, we have

$$\begin{aligned}
&\mathbb{P}\left(\sum_{j=1}^{p} w_{j1}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})\lambda^* \leq -\frac{1-\delta-\bar{\delta}}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right) \\
\leq\ &\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right).
\end{aligned}$$

For the second term, as we have established in the proof of Lemma B.1, $\widehat{\lambda}(z^*) - \lambda^* \sim \mathcal{N}(0, 1/\sum_{i\neq j}(z_i^* \circ z_j^{*-1})^2)$ where the variance is smaller than $1/\alpha p^2$. Then by a standard Gaussian tail bound, we have

$$\begin{aligned}
&\mathbb{P}\left(\sum_{j=1}^{p} w_{j1}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})(\widehat{\lambda}(z^*) - \lambda^*) \leq -\frac{\bar{\delta}}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right) \\
\leq\ &\mathbb{P}\left(\frac{C\sqrt{x}}{p}\left|\sum_{j=1}^{p} w_{j1}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})\right| \geq \frac{\bar{\delta}}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right) \\
&+\mathbb{P}\left(|\widehat{\lambda}(z^*) - \lambda^*| > \frac{C\sqrt{x}}{p}\right) \\
\leq\ &2\exp\left(-C'\frac{\bar{\delta}^2 p^2 \lambda^{*4}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2}{x}\right) + e^{-x}.
\end{aligned}$$

Take $x = \bar{\delta}p|\lambda^*|^2\sqrt{\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2}$, and we obtain the bound

$$\begin{aligned}
&3\exp\left(-C'\bar{\delta}p|\lambda^*|^2\sqrt{\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2}\right) \\
\leq\ &3\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right),
\end{aligned}$$

where the above inequality uses the condition that $p/k^2 \to \infty$ and the fact that $\bar{\delta}$ tends to zero at a sufficiently slow rate. Combining the above bounds, we obtain

$$\mathbb{P}\left(\sum_{j=1}^{p} w_{j1}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})\widehat{\lambda}(z^*) \leq -\frac{1-\delta}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right)$$

$$\leq \quad 4\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_1^{*-1} - z_j^* \circ b^{-1})^2\right).$$

A similar bound holds for

$$\mathbb{P}\left(\sum_{j=1}^{p} w_{jl}(z_j^* \circ z_l^{*-1} - z_j^* \circ b^{-1})\widehat{\lambda}(z^*) \leq -\frac{1-\delta}{2}\lambda^{*2}\sum_{j=1}^{p}(z_j^* \circ z_l^{*-1} - z_j^* \circ b^{-1})^2\right)$$

for each $l \in [p]$. Since $\max_l \max_{b \neq z_l^*} \sum_{j=1}^{p}(z_j^* \circ z_l^{*-1} - z_j^* \circ b^{-1})^2 \geq \lambda^{*2}p$, this implies

$$\mathbb{E}\xi_{\text{ideal}}(\delta) \leq p\exp\left(-\frac{1+o(1)}{8}\lambda^{*2}p\right),$$

under the condition that $p\lambda^{*2} \to \infty$. The conclusion is implied by Markov inequality. □

Proof of Proposition B.1. Let $P = \mathbb{E}Y \in \mathbb{R}^{p\times p}$. Then, $P_{ij} = \lambda^*(z_i^* \circ z_j^{*-1})$ and thus P has k different columns. We denote the k different columns by $\theta_0, \cdots, \theta_{k-1} \in \mathbb{R}^p$. Namely, θ_l is the vector $\lambda^*(z^* \circ l^{-1})$. The same analysis in the proof of Proposition 4.1 leads to the bound

$$\sum_{j=1}^{p} \|\theta_{\pi(\bar{z}_j)} - \theta_{z_j^*}\|^2 \lesssim (M+1)kp, \tag{I.23}$$

for some permutation π acting on the set $\{0, 1, \cdots, k-1\}$ under the condition that $\min_{a\in\mathbb{Z}/k\mathbb{Z}} \sum_{j=1}^{p} \mathbf{1}_{\{z_j^*=a\}} \geq \frac{\alpha p}{k}$ for some constant $\alpha > 0$. Since $\min_{a\neq b} \|\theta_a - \theta_b\|^2 \gtrsim p\lambda^{*2}$, we have

$$\sum_{j=1}^{p} \mathbf{1}_{\{\pi(\bar{z}_j)\neq z_j^*\}} \lesssim \frac{(M+1)k}{\lambda^{*2}}. \tag{I.24}$$

Therefore, when $\frac{(M+1)k^2}{p\lambda^{*2}} = o(1)$, we must have

$$\min_{a\in\mathbb{Z}/k\mathbb{Z}} \sum_{j=1}^{p} \mathbf{1}_{\{\bar{z}_j=a\}} \geq \frac{\alpha p}{2k}. \tag{I.25}$$

For any $a, b \in \mathbb{Z}/k\mathbb{Z}$, recall the notation $\bar{\mathcal{Z}}_{ab} = \{(i,j) : \bar{z}_i = a, \bar{z}_j = b\}$. We will bound the difference between $\bar{Y}_l = \frac{1}{|\bar{\mathcal{Z}}_{l0}|}\sum_{(i,j)\in\bar{\mathcal{Z}}_{l0}} Y_{ij}$ and $\lambda^*(\pi(l)\circ\pi(0)^{-1})$. By triangle inequality, we have

$$|\bar{Y}_l - \lambda^*(\pi(l)\circ\pi(0)^{-1})| \leq \left|\frac{1}{|\bar{\mathcal{Z}}_{l0}|}\sum_{(i,j)\in\bar{\mathcal{Z}}_{l0}}(\mathbb{E}Y_{ij} - \lambda^*(\pi(l)\circ\pi(0)^{-1}))\right| + \left|\frac{1}{|\bar{\mathcal{Z}}_{l0}|}\sum_{(i,j)\in\bar{\mathcal{Z}}_{l0}} w_{ij}\right|. \tag{I.26}$$

For the second term of (I.26), we have

$$\left|\frac{1}{|\bar{\mathcal{Z}}_{l0}|}\sum_{(i,j)\in\bar{\mathcal{Z}}_{l0}} w_{ij}\right| \lesssim \frac{k}{p}\max_{\mathcal{Z}_{l0}}\left|\frac{1}{\sqrt{|\mathcal{Z}_{l0}|}}\sum_{(i,j)\in\mathcal{Z}_{l0}} w_{ij}\right|,$$

where the above inequality is by (I.25), and the maximization is over all k^p possible clustering configurations. A simple union bound argument implies $\max_{\mathcal{Z}_{l0}}\left|\frac{1}{\sqrt{|\mathcal{Z}_{l0}|}}\sum_{(i,j)\in\mathcal{Z}_{l0}} w_{ij}\right| \lesssim \sqrt{p\log k}$ with probability at least $1-e^{-C'p\log k}$. Therefore,

$$\max_l\left|\frac{1}{|\bar{\mathcal{Z}}_{l0}|}\sum_{(i,j)\in\bar{\mathcal{Z}}_{l0}} w_{ij}\right| \lesssim \frac{k\sqrt{\log k}}{\sqrt{p}},$$

with high probability. For the first term of (I.26), we have

$$\begin{aligned}
&\left|\frac{1}{|\bar{\mathcal{Z}}_{l0}|}\sum_{(i,j)\in\bar{\mathcal{Z}}_{l0}}(\mathbb{E}Y_{ij}-\lambda^*(\pi(l)\circ\pi(1)^{-1}))\right| \\
\leq\ & |\lambda^*|k\frac{1}{|\bar{\mathcal{Z}}_{l0}|}\sum_{(i,j)\in\bar{\mathcal{Z}}_{l0}}\left(\mathbf{1}_{\{z_i^*\neq\pi(l)\}}+\mathbf{1}_{\{z_j^*\neq\pi(1)\}}\right) \\
\leq\ & k|\lambda^*|\left(\frac{\sum_{i=1}^p\mathbf{1}_{\{\bar{z}_i=l,z_i^*\neq\pi(l)\}}}{\sum_{i=1}^p\mathbf{1}_{\{\bar{z}_i=l\}}}+\frac{\sum_{j=1}^p\mathbf{1}_{\{\bar{z}_j=1,z_j^*\neq\pi(1)\}}}{\sum_{j=1}^p\mathbf{1}_{\{\bar{z}_j=1\}}}\right) \\
\lesssim\ & \frac{k^2|\lambda^*|}{p}\frac{(M+1)k}{\lambda^{*2}},
\end{aligned}$$

where the last inequality is by (I.24) and (I.25). Combining the two bounds, we can then bound (I.26) by

$$\max_l|\bar{Y}_l-\lambda^*(\pi(l)\circ\pi(0)^{-1})|\lesssim\frac{k^3(M+1)}{p|\lambda^*|}+\frac{k\sqrt{\log k}}{\sqrt{p}}=o(|\lambda^*|),$$

under the condition that $\frac{(M+1)k^3}{p\lambda^{*2}}=o(1)$. We thus have

$$\max_l||\bar{Y}_l|-|\lambda^*|(\pi(l)\circ\pi(0)^{-1})|=o(|\lambda^*|),$$

with high probability. Since the difference among $|\lambda^*|(\pi(l)\circ\pi(0)^{-1})$ with different l's is at least $|\lambda^*|$, the order of $\{|\bar{Y}_l|\}$ perfectly recovers the order of $\{|\lambda^*|(\pi(l)\circ\pi(0)^{-1})\}$. Since $|\bar{Y}_l|=|\bar{Y}|_{(\widehat{\pi}(l))}$, we have

$$\begin{aligned}
\ell(z^{(0)},z^*\circ\pi(0)^{-1}) &= \sum_{j=1}^p\|\theta_{z_j^{(0)}}-\theta_{z_j^*\circ\pi(0)^{-1}}\|^2 \\
&= \sum_{j=1}^p\|\theta_{\pi(\bar{z}_j)}-\theta_{z_j^*}\|^2 \\
&\lesssim (M+1)kp,
\end{aligned}$$

where the last inequality is by (I.23). The proof is complete. □

I.3 Proofs of Permutation Synchronization

In this section, we present the proofs of Theorem C.1, Lemma C.1, Lemma C.2 and Proposition C.1. The conclusions of Theorem C.2 and Corollary C.1 are direct consequences of Theorem 3.1, and thus their proofs are omitted. We first state a technical lemma.

Lemma I.3. *Consider the error matrix $W = Y - \mathbb{E}Y \in \mathbb{R}^{pd\times pd}$ in the problem of permutation synchronization. There exists some constant $C > 0$, such that*

$$\mathbb{P}\left(\|W\| > C\sqrt{pd + x}\right) \leq e^{-x},$$

for any $x > 0$.

Proof. This lemma is an extension of Lemma I.1, and can be proved using the same technique in [16]. We omit the details. □

Proof of Theorem C.1. Let $\bar{U}$ to be a matrix obtained by switching the first and the second rows of I_d. In other words, we have $\|\bar{U} - I_d\|_{\rm F}^2 = 4$. Consider the parameter space

$$\mathcal{Z} = \left\{Z : Z_j = I_d \text{ for all } 1 \leq j \leq p/2 \text{ and } Z_j \in \{I_d, \bar{U}\} \text{ for all } p/2 < j \leq p\right\}.$$

Then, by the same argument that leads to (H.1), we have

$$\begin{aligned} & \inf_{\widehat{Z}} \sup_{Z^*} \mathbb{E} \min_{U} \frac{1}{p} \sum_{j=1}^{p} \mathbf{1}_{\{\widehat{Z}_j \neq U Z_j^*\}} \\ \geq\ & \frac{1}{p} \sum_{j>p/2} \text{ave}_{Z^*_{-j}} \inf_{\widehat{Z}_j} \left(\frac{1}{2}\mathbb{P}_{(Z_j^* = I_d, Z^*_{-j})}(\widehat{Z}_j \neq I_d) + \frac{1}{2}\mathbb{P}_{(Z_j^*=\bar{U}, Z^*_{-j})}(\widehat{Z}_j \neq \bar{U})\right), \end{aligned}$$

Note that the quantity

$$\inf_{\widehat{Z}_j} \left(\frac{1}{2}\mathbb{P}_{(Z_j^* = I_d, Z^*_{-j})}(\widehat{Z}_j \neq I_d) + \frac{1}{2}\mathbb{P}_{(Z_j^*=\bar{U}, Z^*_{-j})}(\widehat{Z}_j \neq \bar{U})\right)$$

is the optimal testing error for a hypothesis testing problem where under H_0 we have $\mathbb{E}Y_{ij} = \lambda^* Z_i^*$ for all $i \in [p]\backslash\{j\}$ and under H_1 we have $\mathbb{E}Y_{ij} = \lambda^* Z_i^* \bar{U}^T$ for all $i \in [p]\backslash\{j\}$. The distributions are both Gaussian under the two hypotheses. By Neyman-Pearson lemma, the optimal testing error is $\mathbb{P}(\mathcal{N}(0,1) > \lambda^*\sqrt{p-1})$. Therefore,

$$\inf_{\widehat{Z}} \sup_{Z^*} \mathbb{E} \min_{U} \frac{1}{p} \sum_{j=1}^{p} \mathbf{1}_{\{\widehat{Z}_j \neq U Z_j^*\}} \geq \frac{1+o(1)}{2}\mathbb{P}(\mathcal{N}(0,1) > \lambda^*\sqrt{p-1}) = \exp\left(-\frac{1+o(1)}{2}p\lambda^{*2}\right),$$

under the condition that $p\lambda^{*2} \to \infty$. When $p\lambda^{*2} = O(1)$, we obtain a constant lower bound. □

Proof of Lemma C.1. We can write

$$\widehat{B}(z) = \frac{\langle \lambda^* Z^* Z^{*T} + W, ZZ^T\rangle}{p^2 d^2} Z = \lambda^* \frac{\langle Z^* Z^{*T}, ZZ^T\rangle}{p^2 d^2} Z + \frac{\langle W, ZZ^T\rangle}{p^2 d^2} Z,$$

where $W = Y - \lambda^* Z^* Z^{*T} \in \mathbb{R}^{pd\times pd}$ is a Gaussian error matrix. Then, we have $\widehat{B}(Z^*) = B^* + \frac{\langle W, Z^* Z^{*T}\rangle}{p^2 d^2} Z^*$. By triangle inequality, we have

$$\|\widehat{B}(Z) - \widehat{B}(Z^*)\|_{\mathrm{F}} \le \|\lambda^* \frac{\langle Z^* Z^{*T}, ZZ^T\rangle}{p^2 d^2} Z - B^*\|_{\mathrm{F}} + \|\frac{\langle W, ZZ^T\rangle}{p^2 d^2} Z - \frac{\langle W, Z^* Z^{*T}\rangle}{p^2 d^2} Z^*\|_{\mathrm{F}}. \quad \text{(I.27)}$$

We will bound the two terms on the right hand side of (I.27) separately. The first term can be bounded as

$$\begin{aligned}
\|\lambda^* \frac{\langle Z^* Z^{*T}, ZZ^T\rangle}{p^2 d^2} Z - B^*\|_{\mathrm{F}} &\le \lambda^* \|Z\|_{\mathrm{F}} \left| \frac{\langle Z^* Z^{*T}, ZZ^T\rangle}{p^2 d^2} - 1 \right| + \lambda^* \|Z - Z^*\|_{\mathrm{F}} \\
&\le \frac{2\lambda^*}{p^{1/2} d^{3/2}} \|Z - Z^*\|_{\mathrm{F}}^2 + \lambda^* \|Z - Z^*\|_{\mathrm{F}} \qquad \text{(I.28)} \\
&\lesssim \lambda^* \|Z - Z^*\|_{\mathrm{F}},
\end{aligned}$$

where the inequality (I.28) is due to

$$\begin{aligned}
|p^2 d^2 - \langle Z^* Z^{*T}, ZZ^T\rangle| &= \frac{1}{2} \|ZZ^T - Z^* Z^{*T}\|_{\mathrm{F}}^2 \\
&\le \|Z^*(Z - Z^*)^T\|_{\mathrm{F}}^2 + \|(Z - Z^*)Z^T\|_{\mathrm{F}}^2 \\
&\le (\|Z\|^2 + \|Z^*\|^2)\|Z - Z^*\|_{\mathrm{F}}^2 \\
&\le 2p\|Z - Z^*\|_{\mathrm{F}}^2.
\end{aligned}$$

For the second term of (I.27), we have

$$\begin{aligned}
&\|\frac{\langle W, ZZ^T\rangle}{p^2 d^2} Z - \frac{\langle W, Z^* Z^{*T}\rangle}{p^2 d^2} Z^*\|_{\mathrm{F}} \\
\le\ & \frac{|\langle W, ZZ^T - Z^* Z^{*T}\rangle|}{p^2 d^2} \|Z\|_{\mathrm{F}} + \|Z - Z^*\|_{\mathrm{F}} \frac{|\langle W, Z^* Z^{*T}\rangle|}{p^2 d^2} \\
\le\ & \frac{\sqrt{2d}\sqrt{pd}\|W\| \|ZZ^T - Z^* Z^{*T}\|_{\mathrm{F}}}{p^2 d^2} + \|Z - Z^*\|_{\mathrm{F}} \frac{\|W\|\sqrt{d}\|Z^* Z^{*T}\|_{\mathrm{F}}}{p^2 d^2} \qquad \text{(I.29)} \\
\lesssim\ & \frac{\|W\| \|Z - Z^*\|_{\mathrm{F}}}{pd} \\
\lesssim\ & \frac{\|Z - Z^*\|_{\mathrm{F}}}{\sqrt{pd}}. \qquad \text{(I.30)}
\end{aligned}$$

The inequality (I.29) is by applying SVD to the rank-$(2d)$ matrix $ZZ^T - Z^* Z^{*T} = \sum_{l=1}^{2d} d_l u_l u_l^T$ so that $|\langle W, ZZ^T - Z^* Z^{*T}\rangle|$ is bounded by

$$\sum_{l=1}^{2d} |d_l| |u_l^T W u_l| \le \|W\| \sum_{l=1}^{2d} |d_l| \le \sqrt{2d}\|W\| \sqrt{\sum_{l=1}^{2d} d_l^2} = \sqrt{2d}\|W\| \|ZZ^T - Z^* Z^{*T}\|_{\mathrm{F}}.$$

Similarly, we also have $|\langle W, Z^* Z^{*T}\rangle| \le \sqrt{d}\|W\| \|Z^* Z^{*T}\|_{\mathrm{F}}$. The last inequality (I.30) is by Lemma I.3. Combining the above bounds, we have

$$\|\widehat{B}(Z) - \widehat{B}(Z^*)\|_{\mathrm{F}} \le \left(\lambda^* + \frac{1}{\sqrt{pd}}\right) \|Z - Z^*\|_{\mathrm{F}}. \qquad \text{(I.31)}$$

Now we are ready to bound (C.3)-(C.5). For (C.3), we have

$$\begin{aligned}
&\sum_{j=1}^{p}\max_{U\neq Z_j^*}\frac{F_j(Z_j^*,U;Z)^2\left\|\mu_j(B^*,U)-\mu_j(B^*,Z_j^*)\right\|^2}{\Delta_j(Z_j^*,U)^4\ell(Z,Z^*)}\\
&=\sum_{j=1}^{p}\max_{U\neq Z_j^*}\frac{\left|\left\langle \epsilon_j,(\widehat{B}(Z^*)-\widehat{B}(Z))(Z_j^*-U)^T\right\rangle\right|^2}{\|B^*(Z_j^*-U)^T\|_{\mathrm{F}}^2\ell(Z,Z^*)}\\
&\lesssim\sum_{j=1}^{p}\max_{U\neq Z_j^*}\frac{\left|\left\langle Z_j^*-U,\epsilon_j^T(\widehat{B}(Z^*)-\widehat{B}(Z))\right\rangle\right|^2}{p\lambda^{*2}\|Z_j^*-U\|_{\mathrm{F}}^2\ell(Z,Z^*)}\\
&\lesssim\sum_{j=1}^{p}\max_{U\neq Z_j^*}\frac{\|Z_j^*-U\|_{\mathrm{F}}^2\|\epsilon_j^T(\widehat{B}(Z^*)-\widehat{B}(Z))\|_{\mathrm{F}}^2}{p\lambda^{*2}\|Z_j^*-U\|_{\mathrm{F}}^2\ell(Z,Z^*)}\\
&=\sum_{j=1}^{p}\frac{\|\epsilon_j^T(\widehat{B}(Z^*)-\widehat{B}(Z))\|_{\mathrm{F}}^2}{p\lambda^{*2}\ell(Z,Z^*)}\\
&=\frac{\left\langle(\widehat{B}(Z^*)-\widehat{B}(Z))(\widehat{B}(Z^*)-\widehat{B}(Z))^T,\sum_{j=1}^p\epsilon_j\epsilon_j^T\right\rangle}{p\lambda^{*2}\ell(Z,Z^*)}\\
&\leq\frac{\left\|\sum_{j=1}^p\epsilon_j\epsilon_j^T\right\|\left\|(\widehat{B}(Z^*)-\widehat{B}(Z))(\widehat{B}(Z^*)-\widehat{B}(Z))^T\right\|_*}{p\lambda^{*2}\ell(Z,Z^*)},
\end{aligned}$$

where $\|\cdot\|_*$ is the matrix nuclear norm. Note that $\sum_{j=1}^p\epsilon_j\epsilon_j^T=WW^T$ and $\|(\widehat{B}(Z^*)-\widehat{B}(Z))(\widehat{B}(Z^*)-\widehat{B}(Z))^T\|_*=\|\widehat{B}(Z^*)-\widehat{B}(Z)\|_{\mathrm{F}}^2$. We have

$$\begin{aligned}
\sum_{j=1}^{p}\max_{U\neq Z_j^*}\frac{F_j(Z_j^*,U;Z)^2\left\|\mu_j(B^*,U)-\mu_j(B^*,Z_j^*)\right\|^2}{\Delta_j(Z_j^*,U)^4\ell(Z,Z^*)}&\lesssim\frac{\|\widehat{B}(Z^*)-\widehat{B}(Z)\|_{\mathrm{F}}^2\,\|W\|^2}{p\lambda^{*2}\ell(Z,Z^*)}\\
&\lesssim\frac{d}{p\lambda^{*2}}+\frac{1}{p^2\lambda^{*4}},
\end{aligned}$$

where we have used Lemma I.3 and (I.31). For (C.4), we have

$$\begin{aligned}
&\frac{\tau}{\Delta_{\min}^2|T|}\sum_{j\in T}\max_{U\neq Z_j}\frac{G_j(Z_j^*,U;Z)^2\left\|\mu_j(B^*,U)-\mu_j(B^*,Z_j^*)\right\|^2}{\Delta_j(Z_j^*,U)^4\ell(Z,Z^*)}\\
=\;&\frac{\tau}{\Delta_{\min}^2|T|}\sum_{j\in T}\max_{U\neq Z_j}\frac{\left|\left\langle\widehat{B}(Z^*)-\widehat{B}(Z),B^*(I_d-Z_j^{*T}U)\right\rangle\right|^2}{\|B^*(Z_j^*-U)^T\|_{\mathrm{F}}^2\ell(Z,Z^*)}\\
\lesssim\;&\frac{\tau}{p\lambda^{*2}}\frac{\|\widehat{B}(Z)-\widehat{B}(Z^*)\|_{\mathrm{F}}^2\|B^*\|_{\mathrm{F}}^2}{p\lambda^{*2}\ell(Z,Z^*)}\\
\lesssim\;&\frac{\tau\left(\lambda^2+\frac{1}{pd}\right)}{(p\lambda^{*2})^2}.
\end{aligned}$$

Finally, for (C.5), we have

$$\|\widehat{B}(Z^*) - B^*\|_{\mathrm{F}} = \left|\frac{\langle W, ZZ^T\rangle}{p^2d^2}\right| \|Z\|_{\mathrm{F}} \leq \frac{\sqrt{d}\|W\|\|Z^*Z^{*T}\|_{\mathrm{F}}}{p^2d^2}\|Z^*\|_{\mathrm{F}} \lesssim 1,$$

by Lemma I.3, and thus for all $j \in [p]$ and $U \neq Z_j^*$,

$$\frac{\left|H_j(Z_j^*, U)\right|}{\Delta_j(Z_j^*, U)^2} = \frac{\left|\left\langle B^* - \widehat{B}(Z^*), B^*(I_d - Z_j^{*T}U)\right\rangle\right|}{\|B^*(Z_j^* - U)^T\|_{\mathrm{F}}^2} \lesssim \|\widehat{B}(Z^*) - B^*\|_{\mathrm{F}}\sqrt{\frac{d}{p\lambda^{*2}}} \lesssim \sqrt{\frac{d}{p\lambda^{*2}}}.$$

The proof is complete. □

Proof of Lemma C.2. We use the notation $\widehat{\lambda}(Z) = \frac{\langle Y, ZZ^T\rangle}{p^2d^2}$. Then, for $U \neq Z_1^*$, we have

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\lambda}(Z^*)\sum_{j=1}^p \left\langle W_{j1}, Z_j^*(Z_1^* - U)^T\right\rangle \leq -\frac{1-\delta}{2}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right) \\
\leq\ &\mathbb{P}\left(\lambda^*\sum_{j=1}^p \left\langle W_{j1}, Z_j^*(Z_1^* - U)^T\right\rangle \leq -\frac{1-\delta-\bar{\delta}}{2}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right) \\
&+\mathbb{P}\left((\widehat{\lambda}(Z^*) - \lambda^*)\sum_{j=1}^p \left\langle W_{j1}, Z_j^*(Z_1^* - U)^T\right\rangle \leq -\frac{\bar{\delta}}{2}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right).
\end{aligned}$$

The first term can be bounded by a standard Gaussian tail bound,

$$\begin{aligned}
&\mathbb{P}\left(\lambda^*\sum_{j=1}^p \left\langle W_{j1}, Z_j^*(Z_1^* - U)^T\right\rangle \leq -\frac{1-\delta-\bar{\delta}}{2}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right) \\
\leq\ &\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right).
\end{aligned}$$

For the second term, note that we have $\widehat{\lambda}(Z^*) - \lambda^* = (p^2d^2)^{-1}\left\langle W, Z^*Z^{*T}\right\rangle$ which is normally distribution. We have

$$\begin{aligned}
&\mathbb{P}\left((\widehat{\lambda}(Z^*) - \lambda^*)\sum_{j=1}^p \left\langle W_{j1}, Z_j^*(Z_1^* - U)^T\right\rangle \leq -\frac{\bar{\delta}}{2}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right) \\
\leq\ &\mathbb{P}\left(\frac{C\sqrt{pd+x}}{pd}\left|\sum_{j=1}^p \left\langle W_{j1}, Z_j^*(Z_1^* - U)^T\right\rangle\right| \geq \frac{\bar{\delta}}{2}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right) \\
&+\mathbb{P}\left(|\widehat{\lambda}(Z^*) - \lambda^*| > \frac{C\sqrt{pd+x}}{pd}\right) \\
\leq\ &2\exp\left(-C'\frac{p^2d^2\bar{\delta}^2\lambda^{*4}p\|Z_1 - U\|_{\mathrm{F}}^2}{pd+x}\right) + e^{-x}.
\end{aligned}$$

Take $x = pd\bar{\delta}\lambda^{*2}\sqrt{p}\|Z_1 - U\|_{\mathrm{F}}$, and we obtain the bound

$$\begin{aligned} & 2\exp\left(-C_1 p^2 d\bar{\delta}^2\lambda^{*4}\|Z_1 - U\|_{\mathrm{F}}^2\right) + \exp\left(-pd\bar{\delta}\lambda^{*2}\sqrt{p}\|Z_1 - U\|_{\mathrm{F}}\right) \\ \leq\ & 3\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right), \end{aligned}$$

under the condition that $p\lambda^{*2} \to \infty$ and $\bar{\delta}$ tends to zero at a sufficiently slow rate. Combining the above bounds, we have

$$\begin{aligned} & \mathbb{P}\left(\widehat{\lambda}(Z^*)\sum_{j=1}^{p}\left\langle W_{j1}, Z_j^*(Z_1^* - U)^T\right\rangle \leq -\frac{1-\delta}{2}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right) \\ \leq\ & 4\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\lambda^{*2}p\|Z_1^* - U\|_{\mathrm{F}}^2\right). \end{aligned}$$

A similar bound holds for $\mathbb{P}\left(\widehat{\lambda}(Z^*)\sum_{j=1}^{p}\left\langle W_{jl}, Z_j^*(Z_l^* - U)^T\right\rangle \leq -\frac{1-\delta}{2}\lambda^{*2}p\|Z_l^* - U\|_{\mathrm{F}}^2\right)$ for each $l \in [p]$.

Now we are ready to bound $\xi_{\mathrm{ideal}}(\xi)$. We have

$$\begin{aligned} \mathbb{E}\xi_{\mathrm{ideal}}(\xi) &= \sum_{l=1}^{p}\sum_{U\in\mathcal{P}_d}\lambda^{*2}p\|Z_l^* - U\|_{\mathrm{F}}^2\mathbb{P}\left(\widehat{\lambda}(Z^*)\sum_{j=1}^{p}\left\langle W_{jl}, Z_j^*(Z_l^* - U)^T\right\rangle \leq -\frac{1-\delta}{2}\lambda^{*2}p\|Z_l^* - U\|_{\mathrm{F}}^2\right) \\ &\leq 4\sum_{l=1}^{p}\sum_{U\in\mathcal{P}_d}\lambda^{*2}p\|Z_l^* - U\|_{\mathrm{F}}^2\exp\left(-\frac{(1-\delta-\bar{\delta})^2}{8}\lambda^{*2}p\|Z_l^* - U\|_{\mathrm{F}}^2\right) \\ &= p\exp\left(-\frac{1+o(1)}{2}p\lambda^{*2}\right), \end{aligned}$$

under the condition that $\frac{p\lambda^{*2}}{d\log d} \to \infty$. The desired conclusion is implied by Markov inequality. □

Proof of Proposition C.1. It is direct to check that the matrix U^* such that $U^{*T} = (Z_1^*, \cdots, Z_p^*)/\sqrt{p}$ satisfies $U^* \in \mathcal{O}(pd, d)$ and collects the eigenvectors of $\mathbb{E}Y$. By Davis-Kahan theorem, there exists some $\mathcal{O}(d, d)$, such that

$$\|\widehat{U} - U^*O\| \lesssim \frac{\|W\|}{p\lambda^*} \lesssim \sqrt{\frac{d}{p\lambda^{*2}}},$$

where the last inequality is by Lemma I.3. According to the definition of $Z_j^{(0)}$, we have

$$\begin{aligned} \|Z_j^{(0)} - Z_j^*O\|_{\mathrm{F}} &\leq \|Z_j^{(0)} - \sqrt{p}\widehat{U}_j\|_{\mathrm{F}} + \|\sqrt{p}\widehat{U}_j - \sqrt{p}U_j^*O\|_{\mathrm{F}} \\ &\leq 2\|\sqrt{p}\widehat{U}_j - \sqrt{p}U_j^*O\|_{\mathrm{F}}. \end{aligned}$$

Therefore,

$$\sum_{j=1}^{p}\|Z_j^{(0)} - Z_j^*O\|_{\mathrm{F}}^2 \leq 4\sum_{j=1}^{p}\|\sqrt{p}\widehat{U}_j - \sqrt{p}U_j^*O\|_{\mathrm{F}}^2 = 4p\|\widehat{U} - U^*O\|_{\mathrm{F}}^2 \lesssim pd\|\widehat{U} - U^*O\|^2 \lesssim \frac{d^2}{\lambda^{*2}}.$$

Then,

$$\begin{aligned}
\sum_{i=1}^{p}\sum_{j=1}^{p}\|Z_i^{(0)} - Z_i^*Z_j^{*T}Z_j^{(0)}\|_{\mathrm{F}}^2 &= \sum_{i=1}^{p}\sum_{j=1}^{p}\|Z_i^{(0)}Z_j^{(0)T} - Z_i^*Z_j^{*T}\|_{\mathrm{F}}^2 \\
&= \|Z^{(0)}Z^{(0)T} - Z^*Z^{*T}\|_{\mathrm{F}}^2 \\
&\leq 2\|Z^{(0)}(Z^{(0)} - Z^*O)^T\|_{\mathrm{F}}^2 + 2\|(Z^{(0)} - Z^*O)O^TZ^{*T}\|_{\mathrm{F}}^2 \\
&\leq 4p\|Z^{(0)} - Z^*O\|_{\mathrm{F}}^2 \\
&= 4p\sum_{j=1}^{p}\|Z_j^{(0)} - Z_j^*O\|_{\mathrm{F}}^2 \\
&\lesssim \frac{pd^2}{\lambda^{*2}}.
\end{aligned}$$

Let $\bar{j} = \mathrm{argmin}_{j\in[p]}\sum_{i=1}^{p}\|Z_i^{(0)} - Z_i^*Z_j^{*T}Z_j^{(0)}\|_{\mathrm{F}}^2$. Then, we have

$$\sum_{i=1}^{p}\|Z_i^{(0)} - Z_i^*Z_{\bar{j}}^{*T}Z_{\bar{j}}^{(0)}\|_{\mathrm{F}}^2 \leq \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}\|Z_i^{(0)} - Z_i^*Z_j^{*T}Z_j^{(0)}\|_{\mathrm{F}}^2 \lesssim \frac{d^2}{\lambda^{*2}}.$$

The proof is complete. □

References

[1] Braverman, M. and Mossel, E. [2008]. Noisy sorting without resampling, *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics, pp. 268–276.

[2] Candes, E. J. and Plan, Y. [2011]. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, *IEEE Transactions on Information Theory* **57**(4): 2342–2359.

[3] Chen, Y. and Candès, E. J. [2018]. The projected power method: An efficient algorithm for joint alignment from pairwise differences, *Communications on Pure and Applied Mathematics* **71**(8): 1648–1714.

[4] Chen, Y., Guibas, L. J. and Huang, Q.-X. [2014]. Near-optimal joint object matching via convex relaxation, *arXiv preprint arXiv:1402.1473* .

[5] Gao, C. [2017]. Phase transitions in approximate ranking, *arXiv preprint arXiv:1711.11189* .

[6] Kuhn, H. W. [1955]. The hungarian method for the assignment problem, *Naval research logistics quarterly* **2**(1-2): 83–97.

[7] Laurent, B. and Massart, P. [2000]. Adaptive estimation of a quadratic functional by model selection, *The Annals of Statistics* **28**(5): 1302–1338.

[8] Ling, S. [2020]. Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods, *arXiv preprint arXiv:2008.05341* .

[9] Löffler, M., Zhang, A. Y. and Zhou, H. H. [2019]. Optimality of spectral clustering for gaussian mixture model, *arXiv preprint arXiv:1911.00538* .

[10] Lu, Y. and Zhou, H. H. [2016]. Statistical and computational guarantees of Lloyd's algorithm and its variants, *arXiv preprint arXiv:1612.02099* .

[11] Mao, C., Weed, J. and Rigollet, P. [2017]. Minimax rates and efficient algorithms for noisy sorting, *arXiv preprint arXiv:1710.10388* .

[12] Ndaoud, M. and Tsybakov, A. B. [2018]. Optimal variable selection and adaptive noisy compressed sensing, *arXiv preprint arXiv:1809.03145* .

[13] Pachauri, D., Kondor, R. and Singh, V. [2013]. Solving the multi-way matching problem by permutation synchronization, *Advances in neural information processing systems*, pp. 1860–1868.

[14] Shah, N. B. and Wainwright, M. J. [2017]. Simple, robust and optimal ranking from pairwise comparisons, *The Journal of Machine Learning Research* **18**(1): 7246–7283.

[15] Si, H., Vikalo, H. and Vishwanath, S. [2014]. Haplotype assembly: An information theoretic view, *2014 IEEE Information Theory Workshop (ITW 2014)*, IEEE, pp. 182–186.

[16] Vershynin, R. [2010]. Introduction to the non-asymptotic analysis of random matrices, *arXiv preprint arXiv:1011.3027* .

[17] Yan, J., Cho, M., Zha, H., Yang, X. and Chu, S. M. [2015]. Multi-graph matching via affinity optimization with graduated consistency regularization, *IEEE transactions on pattern analysis and machine intelligence* **38**(6): 1228–1242.

[18] Zhang, T., Chen, Y., Bao, S., Alley, M. T., Pauly, J. M., Hargreaves, B. A. and Vasanawala, S. S. [2017]. Resolving phase ambiguity in dual-echo dixon imaging using a projected power method, *Magnetic resonance in medicine* **77**(5): 2066–2076.

[19] Zhou, X., Zhu, M. and Daniilidis, K. [2015]. Multi-image matching via fast alternating minimization, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4032–4040.