SUPPLEMENT TO “LEAVE-ONE-OUT SINGULAR SUBSPACE PERTURBATION ANALYSIS FOR SPECTRAL CLUSTERING”

BY Anderson Y. Zhang and Harrison H. Zhou

University of Pennsylvania and Yale University

APPENDIX A: PROOF OF THEOREM 2.3

The proof idea is similar to that of Theorem 2.2 but with more involved calculation as r is not necessarily κ. Consider any $i \in [n]$. Define

$$\tilde{\rho}_{-i} := \frac{\hat{\lambda}_{-i,r} - \hat{\lambda}_{-i,r+1}}{\left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T \right) X_i \right\|}.$$

We need to verify $\tilde{\rho}_{-i} > 2$ first in order to apply Theorem 2.1. Recall the definition of P_{-i} in (36) and E_{-i} in (38). Let the SVD of P_{-i} be

$$P_{-i} = \sum_{j=1}^{p\wedge(n-1)} \lambda_{-i,j} u_{-i,j} v_{-i,j}^T,$$

where $\lambda_{-i,1} \geq \lambda_{-i,2} \geq \ldots \geq \lambda_{-i,p\wedge(n-1)}$. Denote $U_{-i,1:r} = (u_{-i,1}, u_{-i,2}, \ldots, u_{-i,r}) \in \mathbb{O}^{p\times r}$. Then by Weyl’s inequality, we have

$$|\hat{\lambda}_{-i,r} - \lambda_{-i,r}|, |\hat{\lambda}_{-i,r+1} - \lambda_{-i,r+1}| \leq \|E_{-i}\| \leq \|E\|. \tag{48}$$

Then the numerator

$$\hat{\lambda}_{-i,r} - \hat{\lambda}_{-i,r+1} \geq \lambda_{-i,r} - \lambda_{-i,r+1} - 2\|E\|. \tag{49}$$

In the following, we are going to connect $\lambda_{-i,r} - \lambda_{-i,r+1}$ with $\lambda_r - \lambda_{r+1}$.

To bridge the gap between $\lambda_{-i,r}, \lambda_{-i,r+1}$ and λ_r, λ_{r+1}, define

$$\tilde{P}_{-i} := (\theta^*_{z_1^*}, \ldots, \theta^*_{z_{i-1}^*}, U_{-i,1:r}U_{-i,1:r}^T\theta^*_{z_i^*}, \theta^*_{z_{i+1}^*}, \ldots, \theta^*_{z_n^*}) \in \mathbb{R}^{p\times n}.$$

Let $\tilde{\lambda}_{-i,1} \geq \tilde{\lambda}_{-i,2} \geq \ldots \geq \tilde{\lambda}_{-i,p\wedge n}$ be its singular values. Note that $U_{-i,1:r}U_{-i,1:r}^T\tilde{P}_{-i}$ is the best rank-r approximation of $\tilde{P}_{-i}$. This is because for any rank-r projection matrix $M \in \mathbb{R}^{p\times p}$ such that $M^2 = M$, we have

$$\begin{aligned}
\left\| \tilde{P}_{-i} - MM^T\tilde{P}_{-i} \right\|_{\mathrm{F}}^2 &= \left\| (I - MM^T)P_{-i} \right\|_{\mathrm{F}}^2 + \left\| (I - MM^T)U_{-i,1:r}U_{-i,1:r}^T\theta^*_{z_i^*} \right\|_{\mathrm{F}}^2 \\
&\geq \left\| (I - U_{-i,1:r}U_{-i,1:r}^T)P_{-i} \right\|_{\mathrm{F}}^2 + 0 \\
&= \left\| \tilde{P}_{-i} - U_{-i,1:r}U_{-i,1:r}^T\tilde{P}_{-i} \right\|_{\mathrm{F}}^2,
\end{aligned}$$

where we use the fact $U_{-i,1:r}U_{-i,1:r}^T P_{-i}$ is the best rank-r approximation of P_{-i}. Hence, $\mathrm{span}(U_{-i,1:r})$ is exactly the leading r left singular space of $\tilde{P}_{-i}$. It immediately implies:

- $\tilde{\lambda}_{-i,j} = \lambda_{-i,j}$ for any $j \geq r+1$, including

$$\tilde{\lambda}_{-i,r+1} = \lambda_{-i,r+1}. \tag{50}$$

- Since $U_{-i,1:r}U_{-i,1:r}^T\tilde{P}_{-i}$ and $U_{-i,1:r}U_{-i,1:r}^T P_{-i}$ only differ by one column where the latter one can be seen as the leave-one-out counterpart of the former one, using the same argument as in (37), we have

$$\lambda_{-i,r}^2 \geq \left(1-\frac{k}{\beta n}\right)\tilde{\lambda}_{-i,r}^2. \tag{51}$$

Then from (49), we have

$$\hat{\lambda}_{-i,r} - \hat{\lambda}_{-i,r+1} \geq \sqrt{1-\frac{k}{\beta n}}\tilde{\lambda}_{-i,r} - \tilde{\lambda}_{-i,r+1} - 2\left\|E\right\|. \tag{52}$$

For the difference between $\tilde{\lambda}_{-i,r}, \tilde{\lambda}_{-i,r+1}$ and λ_r, λ_{r+1}, we use the Weyl's inequality again:

$$\max_{j\in[k]}\left|\tilde{\lambda}_{-i,j}-\lambda_j\right| \leq \left\|P-\tilde{P}_{-i}\right\| = \left\|\theta^*_{z_i^*} - U_{-i,1:r}U_{-i,1:r}^T\theta^*_{z_i^*}\right\|.$$

In the proof of Theorem 2.2, we show $u_{-i,j}\in \text{span}(\{\theta_a^*\}_{a\in[k]})$ for each $j\in[\kappa]$. Then

$$\begin{aligned}\left\|\theta^*_{z_i^*} - U_{-i,1:r}U_{-i,1:r}^T\theta^*_{z_i^*}\right\| &= \left\|\left(u_{-i,r+1},\ldots,u_{-i,\kappa}\right)\left(u_{-i,r+1},\ldots,u_{-i,\kappa}\right)^T\theta^*_{z_i^*}\right\| \\ &= \sqrt{\sum_{a\in[\kappa]:a\geq r+1}\left(u_{-i,a}^T\theta^*_{z_i^*}\right)^2}.\end{aligned}$$

For any $a\in[\kappa]$ such $a\geq r+1$, we have

$$\begin{aligned}\left(u_{-i,a}^T\theta^*_{z_i^*}\right)^2 &\leq \frac{1}{\left|\left\{j\in[n]:z_j^*=z_i^*\right\}\right|-1}\sum_{j\in[n]:j\neq i,z_j^*=z_i^*}\left(u_{-i,a}^T\theta^*_{z_j^*}\right)^2 \leq \frac{1}{\frac{\beta n}{k}-1}(u_{-i,a}^T P_{-i})^2 \\ &\leq \frac{\lambda_{-i,a}^2}{\frac{\beta n}{k}-1} \leq \frac{\lambda_{-i,r+1}^2}{\frac{\beta n}{k}-1}.\end{aligned}$$

Hence, we obtain $\|\theta^*_{z_i^*} - U_{-i,1:r}U_{-i,1:r}^T\theta^*_{z_i^*}\| \leq \sqrt{\kappa}\lambda_{-i,a}/\sqrt{\beta n/k-1}$ and consequently,

$$\max_{j\in[k]}\left|\tilde{\lambda}_{-i,j}-\lambda_j\right| \leq \frac{\sqrt{\kappa}\lambda_{-i,r+1}}{\sqrt{\frac{\beta n}{k}-1}}. \tag{53}$$

Then together with (50), we have $|\lambda_{-i,r+1}-\lambda_{r+1}|\leq \sqrt{\kappa}\lambda_{-i,r+1}/\sqrt{\beta n/k-1}$ and hence

$$\lambda_{-i,r+1} \leq \frac{\lambda_{r+1}}{1-\frac{\sqrt{\kappa}}{\sqrt{\frac{\beta n}{k}-1}}}. \tag{54}$$

Denote $d:=\beta n/k$. With (52), we have

$$\begin{aligned}\hat{\lambda}_{-i,r} - \hat{\lambda}_{-i,r+1} &\geq \sqrt{\frac{d-1}{d}}\left(\lambda_r - \frac{\lambda_{-i,r+1}}{\sqrt{d-1}}\right) - \left(\lambda_{r+1} + \frac{\lambda_{-i,r+1}}{\sqrt{d-1}}\right) - 2\left\|E\right\| \\ &\geq \sqrt{\frac{d-1}{d}}\lambda_r - \lambda_{r+1}\left(1+\left(\frac{1}{\sqrt{d}}+\frac{1}{\sqrt{d-1}}\right)\frac{1}{1-\frac{\sqrt{\kappa}}{\sqrt{d-1}}}\right) - 2\left\|E\right\| \\ &\geq \sqrt{\frac{d-1}{d}}\left(\lambda_r - \lambda_{r+1} - \frac{4}{\sqrt{d}}\lambda_{r+1}\right) - 2\left\|E\right\| \\ &\geq \frac{3}{4}\left(\lambda_r - \lambda_{r+1} - \frac{4}{\sqrt{d}}\lambda_{r+1}\right) - 2\left\|E\right\|, \end{aligned} \tag{55}$$

where in the last two inequalities we use the assumption that $d/k \geq 10$. As a consequence, we have

$$\tilde{\rho}_{-i} \geq \frac{\hat{\lambda}_{-i,r} - \hat{\lambda}_{-i,r+1}}{\left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right) X_i \right\|} \geq \frac{\frac{3}{4}\left(\lambda_r - \lambda_{r+1} - \frac{4}{\sqrt{d}}\lambda_{r+1}\right) - 2\left\|E\right\|}{\left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right) X_i \right\|}.$$

Next, we are going to simplify the denominator of the above display. Using the orthogonality of the singular vectors, we have

$$\begin{aligned}
&\left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right) \theta^*_{z_i^*} \right\| \\
&\leq \left\| \left(I - \hat{U}_{-i,1:\kappa}\hat{U}_{-i,1:\kappa}^T\right) \theta^*_{z_i^*} \right\| + \left\| \left(\hat{u}_{-i,r+1}, \ldots, \hat{u}_{-i,\kappa}\right) \left(\hat{u}_{-i,r+1}, \ldots, \hat{u}_{-i,\kappa}\right)^T \theta^*_{z_i^*} \right\| \\
&= \left\| \left(I - \hat{U}_{-i,1:\kappa}\hat{U}_{-i,1:\kappa}^T\right) \theta^*_{z_i^*} \right\| + \sqrt{\sum_{j=r+1}^{\kappa} \left(\hat{u}_{-i,j}^T \theta^*_{z_i^*}\right)^2} \\
&\leq \frac{3\sqrt{\kappa}\left\|E\right\|}{\sqrt{\frac{\beta n}{k}} - 1} + \sqrt{\sum_{j=r+1}^{\kappa} \left(\frac{\hat{\lambda}_{-i,j}}{\sqrt{\frac{\beta n}{k}} - 1} + \frac{\left\|E\right\|}{\sqrt{\frac{\beta n}{k}} - 1} \right)^2} \\
&\leq \frac{3\sqrt{\kappa}\left\|E\right\|}{\sqrt{\frac{\beta n}{k}} - 1} + \sqrt{\kappa}\left(\frac{\hat{\lambda}_{-i,r+1}}{\sqrt{\frac{\beta n}{k}} - 1} + \frac{\left\|E\right\|}{\sqrt{\frac{\beta n}{k}} - 1} \right),
\end{aligned}$$

where the second to the inequality is due to (41) and (44). By (54) and the Weyl's inequality, we have

$$\hat{\lambda}_{-i,r+1} \leq \lambda_{-i,r+1} + \left\|E\right\| \leq \frac{1}{1 - \frac{\sqrt{\kappa}}{\sqrt{\frac{\beta n}{k}} - 1}} \lambda_{r+1} + \left\|E\right\|.$$

Then, with the assumption $\beta n/k^2 \geq 10$, we have

$$\begin{aligned}
\left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right) \theta^*_{z_i^*} \right\| &\leq \frac{3\sqrt{\kappa}\left\|E\right\|}{\sqrt{\frac{\beta n}{k}} - 1} + \sqrt{\kappa}\left(\frac{\lambda_{r+1}}{\sqrt{\frac{\beta n}{k}} - 1 - \sqrt{\kappa}} + \frac{2\left\|E\right\|}{\sqrt{\frac{\beta n}{k}} - 1} \right) \\
&\leq \frac{\sqrt{k\kappa}}{\sqrt{\beta n}}(6\left\|E\right\| + 2\lambda_{r+1}).
\end{aligned}$$

Hence,

$$\begin{aligned}
\left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right) X_i \right\| &\leq \left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right) \theta^*_{z_i^*} \right\| + \left\| \left(I - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right) \epsilon_i \right\| \\
&\leq \frac{\sqrt{k\kappa}}{\sqrt{\beta n}}(6\left\|E\right\| + 2\lambda_{r+1}) + \left\|E\right\|.
\end{aligned}$$

As a result,

$$\tilde{\rho}_{-i} \geq \frac{\frac{3}{4}\left(\lambda_r - \lambda_{r+1} - \frac{4}{\sqrt{\beta n/k}}\lambda_{r+1}\right) - 2\left\|E\right\|}{\frac{\sqrt{k\kappa}}{\sqrt{\beta n}}(6\left\|E\right\| + 2\lambda_{r+1}) + \left\|E\right\|} \geq \frac{\tilde{\rho}_0}{8} > 2,$$

under the assumption that $\beta n/(k^2) \geq 10$ and (11).

The remaining part of the proof is to study $\{\hat{u}_{-i,a}^T X_i\}_{a\in[r]}$ and then apply Theorem 2.1. Following the exact argument as in the proof of Theorem 2.2, we have

$$\sqrt{\sum_{a\in r}\left(\frac{\hat{u}_{-i,a}^T X_i}{\hat{\lambda}_{-i,a}}\right)^2} \leq \frac{\sqrt{r}}{\sqrt{\frac{\beta n}{k}}-1} + \frac{1}{\hat{\lambda}_{-i,r}}\frac{\|E\|\sqrt{r}}{\sqrt{\frac{\beta n}{k}}-1} + \frac{1}{\hat{\lambda}_{-i,r}}\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|.$$

Under the assumption that $\beta n/(k^2)\geq 10$ and (11), (55) is lower bounded by $\lambda_r/2$. This also implies $\hat{\lambda}_{-i,r}\geq \lambda_r/2$. Then a direct application of Theorem 2.1 leads to

$$\begin{aligned}\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right\|_{\mathrm{F}} &\leq \frac{4\sqrt{2}}{\tilde{\rho}_{-i}}\left(\frac{\sqrt{r}}{\sqrt{\beta n/k-1}} + \frac{1}{\hat{\lambda}_{-i,r}}\left(\frac{\sqrt{r}\,\|E\|}{\sqrt{\beta n/k-1}} + \left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right)\right)\\ &\leq \frac{128}{\tilde{\rho}_0}\left(\frac{\sqrt{kr}}{\sqrt{\beta n}} + \frac{\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|}{\lambda_r}\right).\end{aligned}$$

APPENDIX B: PROOFS OF RESULTS IN SECTION 3.4

Before presenting the proof of Lemma 3.3, we first show $\hat{r}$ defined in (23) always exists. In addition, since $\hat{r}\in[k]$ is a random variable, we are going to associate it with some deterministic set in $[k]$. Recall $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_{p\wedge n}$ are singular values of the signal matrix P and κ is the its rank. Let its SVD be $P=\sum_{i\in[p\wedge n]}\lambda_i u_i v_i^T$ with $\{u_j\}_{j\in[p\wedge n]}\in\mathbb{R}^p$ being its left singular vectors.

LEMMA B.1. *Under the same conditions as stated in Lemma 3.3, $\hat{r}$ always exists. Furthermore, we have $\hat{r}\in\mathcal{R}$ where*

$$\mathcal{R} := \{a\in[k] : \lambda_a - \lambda_{a+1} \geq (\tilde{\rho}-2)\|E\| \text{ and } \lambda_{a+1}\leq (k\tilde{\rho}+1)\|E\|\}. \tag{56}$$

PROOF. The existence of $\hat{r}$ can be proved by contradiction. If $\hat{r}$ does not exist, it means that $\{a\in[k]:\hat{\lambda}_a-\hat{\lambda}_{a+1}\geq T\}$ is empty, which implies $\hat{\lambda}_1<\hat{\lambda}_{k+1}+kT=\hat{\lambda}_{k+1}+k\tilde{\rho}\|E\|$. By Weyl's inequality, we have $|\hat{\lambda}_a-\lambda_a|\leq\|E\|$ for all singular values of X and P. Then we have $\lambda_1<(k\tilde{\rho}+1)\|E\|$. On the other hand, we have

$$\begin{aligned}\lambda_1^2 &= \max_{w\in\mathbb{R}^p:\|w\|=1}\left\|w^TP\right\|^2 \geq \max_{a,b\in[k]:a\neq b}\max_{w\in\mathbb{R}^p:\|w\|=1}\frac{\beta n}{k}\left(\left\|w^T\theta_a^*\right\|^2+\left\|w^T\theta_b^*\right\|^2\right)\\ &\geq \max_{a,b\in[k]:a\neq b}\max_{w\in\mathbb{R}^p:\|w\|=1}\frac{\beta n}{2k}\left\|w^T\theta_a^*-w^T\theta_b^*\right\|^2 = \frac{\beta n}{2k}\Delta^2,\end{aligned}$$

where the first inequality is due to the mixture model structure in P and the second inequality is due to $2(x_1+x_2)^2\geq(x_1-x_2)^2$ for any two scalars x_1,x_2. Then we have $\lambda_1\geq\sqrt{\beta n/(2k)}\Delta=(\tilde{\psi}_0/\sqrt{2})k^{1.5}\|E\|$ by (25). Since $\tilde{\rho}<\tilde{\psi}_0/64$ is assumed, we have $(k\tilde{\rho}+1)\|E\|<(\tilde{\psi}_0/\sqrt{2})k^{1.5}\|E\|$, which is a contradiction.

To prove the second statement, note that we have $\hat{\lambda}_{\hat{r}}-\hat{\lambda}_{\hat{r}+1}\geq\tilde{\rho}\|E\|$ and $\hat{\lambda}_{\hat{r}+1}\leq k\tilde{\rho}\|E\|$. Since $|\hat{\lambda}_a-\lambda_a|\leq\|E\|$ for all singular values of X and P, we have $\lambda_{\hat{r}}-\lambda_{\hat{r}+1}\geq(\tilde{\rho}-2)\|E\|$ and $\lambda_{\hat{r}+1}\leq(k\tilde{\rho}+1)\|E\|$. Hence, $\hat{r}\in\mathcal{R}$. □

PROOF OF LEMMA 3.3. From Lemma B.1, we know $\hat{r}$ exists and $\hat{r}\in\mathcal{R}$. Consider an arbitrary $r\in\mathcal{R}$ and define $\hat{U}_{1:r}:=(\hat{u}_1,\ldots,\hat{u}_r)\in\mathbb{R}^{p\times r}$. Perform k-means on the columns of $\hat{U}_{1:r}\hat{U}_{1:r}^TX$ and let the output be

$$\left(\check{z}(r),\{\check{\theta}_j(r)\}_{j=1}^k\right) = \mathop{\mathrm{argmin}}_{z\in[k]^n,\{\theta_j\}_{j=1}^k\in\mathbb{R}^p}\sum_{i\in[n]}\left\|\hat{U}_{1:r}\hat{U}_{1:r}^TX-\theta_{z_i}\right\|^2.$$

In the following, we are going to establish statistical properties for $\check{z}(r)$ and eventually obtain a desired upper bound for $\ell(\check{z}(r), z^*)$. Since performing k-means on the columns of $\hat{U}_{1:r}^T X$ is equivalent to k-means on the columns of $\hat{U}_{1:r}\hat{U}_{1:r}^T X$, and since $\hat{r} \in \mathcal{R}$, we have $\tilde{z} = \check{z}(\hat{r})$ and thus the desired upper bound also holds for $\ell(\tilde{z}, z^*)$.

In the rest of the proof we are going to analyze $\check{z}(r)$ for any $r \in \mathcal{R}$. For simplicity, we use the notation $\check{z}, \{\check{\theta}_j\}_{j\in[n]}$ instead of $\check{z}(r), \{\check{\theta}_j(r)\}_{j\in[n]}$. The remaining proof can be decomposed into several parts.

(Preliminary Results for $\check{z}, \{\check{\theta}_j\}_{j\in[n]}$). We are going to use Proposition 3.1 to have some preliminary results. Define $U_{1:r} := (u_1, \ldots, u_r)$ and $U_{(r+1):k} := (u_{r+1}, \ldots, u_k)$. Instead of the decomposition (6), we can write

$$X_i = U_{1:r}U_{1:r}^T\theta^*_{z_i^*} + U_{(r+1):k}U_{(r+1):k}^T\theta^*_{z_i^*} + \epsilon_i = U_{1:r}U_{1:r}^T\theta^*_{z_i^*} + \check{\epsilon}_i,$$

where $\check{\epsilon}_i := U_{(r+1):k}U_{(r+1):k}^T\theta^*_{z_i^*} + \epsilon_i$. In this way, we have a new mixture model with the centers being $\{U_{1:r}U_{1:r}^T\theta^*_a\}_{a\in[k]}$ and the additive noises being $\{\check{\epsilon}_i\}$. Define $\check{E} := (\check{\epsilon}_1, \ldots, \check{\epsilon}_n)$. Then

$$\begin{aligned}\left\|\check{E}\right\| &\leq \|E\| + \left\|\left(U_{(r+1):k}U_{(r+1):k}^T\theta^*_{z_1^*}, \ldots, U_{(r+1):k}U_{(r+1):k}^T\theta^*_{z_n^*}\right)\right\| \\ &= \|E\| + \left\|U_{(r+1):k}U_{(r+1):k}^T P\right\| = \|E\| + \lambda_{r+1} \\ &\leq (k\tilde{\rho} + 2)\|E\|. \qquad (57)\end{aligned}$$

The separation among the new centers is no longer Δ. Define

$$\check{\Delta} := \min_{a,b\in[k]:a\neq b}\left\|U_{1:r}U_{1:r}^T\theta^*_a - U_{1:r}U_{1:r}^T\theta^*_b\right\|.$$

For any $a, b \in [k]$, $U_{1:r}U_{1:r}^T\theta^*_a - U_{1:r}U_{1:r}^T\theta^*_b = (\theta^*_a - \theta^*_b) - U_{(r+1):k}U_{(r+1):k}^T\theta^*_a + U_{(r+1):k}U_{(r+1):k}^T\theta^*_b$. Also,

$$\begin{aligned}\max_{a\in[k]}\left\|U_{(r+1):k}U_{(r+1):k}^T\theta^*_a\right\| &= \max_{a\in[k]}\sqrt{\frac{\sum_{i\in[n]:z_i^*=a}\left\|U_{(r+1):k}U_{(r+1):k}^T\theta^*_a\right\|^2}{|\{i\in[n]:z_i^*=a\}|}} \leq \frac{\left\|U_{(r+1):k}U_{(r+1):k}^T P\right\|_{\mathrm{F}}}{\sqrt{\beta n/k}} \\ &\leq \frac{2\sqrt{k}\lambda_{r+1}}{\sqrt{\beta n/k}} \leq \frac{\sqrt{k}(k\tilde{\rho}+1)\|E\|}{\sqrt{\beta n/k}}. \qquad (58)\end{aligned}$$

Hence, we have

$$\check{\Delta} \geq \min_{a,b\in[k]:a\neq b}\|\theta^*_a - \theta^*_b\| - 2\max_{a\in[k]}\left\|U_{(r+1):k}U_{(r+1):k}^T\theta^*_a\right\| \geq \Delta - \frac{2\sqrt{k}(k\tilde{\rho}+1)\|E\|}{\sqrt{\beta n/k}}. \qquad (59)$$

Then from Proposition 3.1, as long as (which will be verified later)

$$\check{\psi}_0 := \frac{\check{\Delta}}{\beta^{-0.5}kn^{-0.5}\left\|\check{E}\right\|} \geq 16, \qquad (60)$$

we have

$$\ell(\check{z}, z^*) = \frac{1}{n}|i\in[n]: \check{z}_i \neq \phi(z_i^*)| \leq \frac{C_0 k\left\|\check{E}\right\|^2}{n\check{\Delta}^2},$$

and

$$\max_{a\in[k]}\left\|\check{\theta}_{\phi(z)} - U_{1:r}U_{1:r}^T\theta^*_a\right\| \leq C_0\beta^{-0.5}kn^{-0.5}\left\|\check{E}\right\|.$$

where $C_0 = 128$.

(Entrywise Decomposition for $\check{z}$). Next, we are going to have an entrywise decomposition for $\mathbb{I}\{\hat{z}_i \neq \phi(z_i^*)\}$ that is analogous to that of Lemma 3.2. When (60) is satisfied, from Lemma 3.1, we have

$$\mathbb{I}\{\check{z}_i \neq \phi(z_i^*)\} \leq \mathbb{I}\left\{\left(1 - C_0\check{\psi}_0^{-1}\right)\check{\Delta} \leq 2\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\check{\epsilon}_i\right\|\right\}.$$

By the definition of $\check{\epsilon}_i$ and (58), we have

$$\begin{aligned}
\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\check{\epsilon}_i\right\| &\leq \left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\| + \left\|\hat{U}_{1:r}\hat{U}_{1:r}^T U_{(r+1):k}U_{(r+1):k}^T\theta_{z_i^*}^*\right\| \\
&\leq \left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\| + \left\|U_{(r+1):k}U_{(r+1):k}^T\theta_{z_i^*}^*\right\| \\
&\leq \left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\| + \frac{\sqrt{k}(k\tilde{\rho}+1)\left\|E\right\|}{\sqrt{\beta n/k}}.
\end{aligned}$$

Then, we have

$$\begin{aligned}
\mathbb{I}\{\check{z}_i \neq \phi(z_i^*)\} &\leq \mathbb{I}\left\{\left(1 - C_0\check{\psi}_0^{-1}\right)\check{\Delta} \leq 2\left(\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\| + \frac{\sqrt{k}(k\tilde{\rho}+1)\left\|E\right\|}{\sqrt{\beta n/k}}\right)\right\} \\
&= \mathbb{I}\left\{\left(1 - C_0\check{\psi}_0^{-1} - \frac{2\sqrt{k}(k\tilde{\rho}+1)\left\|E\right\|}{\sqrt{\beta n/k}\check{\Delta}}\right)\check{\Delta} \leq 2\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\|\right\}.
\end{aligned}$$

From (56), under the assumption that $\tilde{\rho} > 4$ and $\beta n/k^4 > 400$, we have $\tilde{\rho}_0$ defined as in (11) to satisfy

$$\tilde{\rho}_0 \geq \frac{(\tilde{\rho}-1)\left\|E\right\|}{\max\left\{\left\|E\right\|, \sqrt{\frac{k^2}{\beta n}}(k\tilde{\rho}+1)\left\|E\right\|\right\}} \geq 2.$$

Then Theorem 2.3 can be applied, with which we have

$$\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right\|_{\mathrm{F}} \leq \frac{256\sqrt{rk}}{\sqrt{n\beta}} + \frac{256\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|}{\lambda_r}.$$

Then following the proof of Lemma 3.2, we have

$$\begin{aligned}
&\mathbb{I}\{\check{z}_i \neq \phi(z_i^*)\} \\
&\leq \mathbb{I}\left\{\left(1 - C_0\check{\psi}_0^{-1} - \frac{2\sqrt{k}(k\tilde{\rho}+1)\left\|E\right\|}{\sqrt{\beta n/k}\check{\Delta}}\right)\check{\Delta} \leq 2\left(\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\| + \left\|\hat{U}_{1:r}\hat{U}_{1:r}^T - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right\|_{\mathrm{F}}\left\|E\right\|\right)\right\} \\
&\leq \mathbb{I}\left\{\left(1 - C_0\check{\psi}_0^{-1} - \frac{2\sqrt{k}(k\tilde{\rho}+1)\left\|E\right\|}{\sqrt{\beta n/k}\check{\Delta}}\right)\check{\Delta} \leq 2\left(\frac{256\sqrt{rk}\left\|E\right\|}{\sqrt{n\beta}} + \left(1 + \frac{256\left\|E\right\|}{\lambda_r}\right)\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right)\right\} \\
&\leq \mathbb{I}\left\{\left(1 - C_0\check{\psi}_0^{-1} - \frac{2\sqrt{k}(k\tilde{\rho}+257)\left\|E\right\|}{\sqrt{\beta n/k}\check{\Delta}}\right)\check{\Delta} \leq 2\left(1 + \frac{256\left\|E\right\|}{\lambda_r}\right)\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right\} \\
&\leq \mathbb{I}\left\{\left(1 - C_0\check{\psi}_0^{-1} - \frac{2\sqrt{k}(k\tilde{\rho}+257)\left\|E\right\|}{\sqrt{\beta n/k}\check{\Delta}}\right)\check{\Delta} \leq 2\left(1 + \frac{256}{\tilde{\rho}-2}\right)\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right\},
\end{aligned}$$

where in the last inequality we use $\lambda_r \geq (\tilde{\rho}-2)\left\|E\right\| > 0$ (as long as $\tilde{\rho} > 2$) from (56).

The last step of the proof is to simplify the above display using Δ instead of $\check{\Delta}$. Then, under the assumption that $\tilde{\rho} > 256$, we have $(1+256/(\tilde{\rho}-2))^{-1} \leq (1-512/\tilde{\rho})$. Recall the definition of $\tilde{\psi}_0$ in (25). Under the assumption that $\tilde{\rho} \leq \tilde{\psi}_0/64$, we have

$$\check{\Delta} \geq \Delta\left(1 - \frac{4\beta^{-0.5}k^2n^{-0.5}\tilde{\rho}\,\|E\|}{\Delta}\right) = \Delta\left(1-\frac{4\tilde{\rho}}{\tilde{\psi}_0}\right) \geq \frac{\Delta}{2}, \tag{61}$$

according to (59). Then together with (57), we can verify (60) holds due to

$$\check{\psi}_0 \geq \frac{\Delta/2}{\beta^{-0.5}kn^{-0.5}(k\tilde{\rho}+2)\,\|E\|} \geq \frac{\Delta}{4\beta^{-0.5}k^2n^{-0.5}\tilde{\rho}\,\|E\|} = \frac{\tilde{\psi}_0}{4\tilde{\rho}} \geq 16.$$

Rearranging all the terms with the help of (61), we can simplify $\mathbb{I}\{\check{z}_i \neq \phi(z_i^*)\}$ into

$$\begin{aligned}
&\mathbb{I}\{\check{z}_i \neq \phi(z_i^*)\}\\
&\leq \mathbb{I}\left\{\left(1-4C_0\tilde{\rho}\tilde{\psi}_0 - \frac{4\beta^{-0.5}k^2n^{-0.5}\tilde{\rho}\,\|E\|}{\Delta/2}\right)\left(1-\frac{256}{\tilde{\rho}}\right)\left(1-\frac{4\tilde{\rho}}{\tilde{\psi}_0}\right)\Delta \leq 2\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right\}\\
&\leq \mathbb{I}\left\{\left(1-5C_0\tilde{\rho}\tilde{\psi}_0^{-1} - 256\tilde{\rho}^{-1}\right)\Delta \leq 2\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right\}.
\end{aligned}$$

□

PROOF OF THEOREM 3.2. Recall the definition of $\mathcal{F}$ in (46). Then if $\mathcal{F}$ holds, by appropriate choices of C_1, C_2, we can verify the assumptions needed in Lemma 3.3 hold, which lead to

$$\mathbb{I}\{\tilde{z}_i \neq \phi(z_i^*)\}\mathbb{I}\{\mathcal{F}\} \leq \mathbb{I}\left\{\left(1-C''(\rho_2\psi_2^{-1}+\rho_2^{-1})\right)\Delta \leq 2\left\|\hat{U}_{-i,1:\hat{r}}\hat{U}_{-i,1:\hat{r}}^T\epsilon_i\right\|\right\}\mathbb{I}\{\mathcal{F}\},$$

for some constant $C''>0$. Though $\hat{r}$ is random, the proof of Lemma 3.3 shows that $\hat{r} \in \mathcal{R} \subset [k]$ where $\mathcal{R}$ is defined in (56). Note that for any $r \in [k]$, we can follow the proof of Theorem 3.1 to show

$$\mathbb{E}\mathbb{I}\left\{\left(1-C''(\rho_2\psi_2^{-1}+\rho_2^{-1})\right)\Delta \leq 2\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right\} \leq \exp\left(-(1-C'''(\rho_2\psi_2^{-1}+\rho_2^{-1}))\frac{\Delta^2}{8\sigma^2}\right),$$

for some constant $C'''>0$. Hence, the same upper bound holds for $\mathbb{E}\mathbb{I}\{(1-C''(\rho_2\psi_2^{-1}+\rho_2^{-1}))\Delta \leq 2\|\hat{U}_{-i,1:\hat{r}}\hat{U}_{-i,1:\hat{r}}^T\epsilon_i\|\}$. The rest of the proof follows that of Theorem 3.1 and is omitted here. □

APPENDIX C: PROOF OF THEOREM 3.3

Define $\mathcal{F} = \left\{\|E\| \leq \sqrt{2}(\sqrt{n}+\sqrt{p})\sigma\right\}$. Then by Lemma B.1 of [27], we have $\mathbb{P}(\mathcal{F}) \geq 1-e^{-0.08n}$. Then under the event $\mathcal{F}$, the assumption (26) implies (16) holds, and hence (17) and (18) hold. For simplicity, and without loss of generality, we can let ϕ in (17)-(18) to be the identity, and we get

$$\ell(\hat{z},z^*) = \frac{1}{n}|\{i\in[n]:\hat{z}_i \neq z_i^*\}| \leq \frac{C_0k\left(1+\sqrt{\frac{p}{n}}\right)^2\sigma^2}{\Delta^2},$$

and

$$\max_{a\in[k]}\left\|\hat{\theta}_a - \theta_a^*\right\| \leq C_0\beta^{-0.5}k\left(1+\sqrt{\frac{p}{n}}\right)\sigma,$$

where $C_0>0$ is some constant.

Denote $\hat{P} = \hat{U}_{1:k}\hat{U}_{1:k}^T X$ and let $\hat{P}_{\cdot,i}$ be its ith column so that $\hat{P}_{\cdot,i} = \hat{U}_{1:k}\hat{U}_{1:k}^T X_i$. We define $r \in [k]$ as (with $\lambda_{k+1} := 0$)

$$r = \max\left\{j \in [k] : \lambda_j - \lambda_{j+1} \geq \tau\sqrt{n+p}\sigma\right\}, \tag{62}$$

for a sequence $\tau \to \infty$ to be determined later. We note that if $\Delta/(k^{\frac{3}{2}}\tau\beta^{\frac{1}{2}}(1+p/n)^{\frac{1}{2}}\sigma) \to \infty$, the set $\{j \in [k] : \lambda_j - \lambda_{j+1} \geq \tau\sqrt{n+p}\sigma\}$ is not empty. Otherwise, this would imply $\lambda_1 \leq k\tau\sqrt{n+p}\sigma$ which would contradict with the fact $\lambda_1 \geq \sqrt{\beta n/k}\Delta/(2\sigma)$ (see Proposition A.1 of [27]). By the definition of r in (62), we immediately have

$$\lambda_r - \lambda_{r+1} \geq \tau\sqrt{n+p}\sigma, \tag{63}$$

$$\text{and } \lambda_{r+1} \leq k\tau\sqrt{n+p}\sigma. \tag{64}$$

We split $\hat{U}_{1:k}$ into $(\hat{U}_{1:r}, \hat{U}_{(r+1):k})$ where $\hat{U}_{1:r} := (\hat{u}_1, \ldots, \hat{u}_r)$ and $\hat{U}_{(r+1):k} := (\hat{u}_{r+1}, \ldots, \hat{u}_k)$. We decompose $\hat{P}_{\cdot,i} = \hat{P}_{\cdot,i}^{(1)} + \hat{P}_{\cdot,i}^{(2)}$, where $\hat{P}_{\cdot,i}^{(1)} := \hat{U}_{1:r}\hat{U}_{1:r}^T\hat{P}_{\cdot,i}$ and $\hat{P}_{\cdot,i}^{(2)} := \hat{U}_{(r+1):k}\hat{U}_{(r+1):k}^T\hat{P}_{\cdot,i}$. Similarly, for each $a \in [k]$, we decompose $\hat{\theta}_a = \hat{\theta}_a^{(1)} + \hat{\theta}_a^{(2)}$, where $\hat{\theta}_a^{(1)} := \hat{U}_{1:r}\hat{U}_{1:r}^T\hat{\theta}_a$ and $\hat{\theta}_a^{(2)} := \hat{U}_{(r+1):k}\hat{U}_{(r+1):k}^T\hat{\theta}_a$. Due to the orthogonality of $\{\hat{u}_l\}_{l\in[k]}$, we obtain that for any $i \in [n]$ and any $a \in [k]$ such that $a \neq z_i^*$,

$$\begin{aligned}\mathbb{I}\{\hat{z}_i = a\} &\leq \mathbb{I}\left\{\left\|\hat{P}_{\cdot,i}^{(1)} + \hat{P}_{\cdot,i}^{(2)} - \hat{\theta}_a^{(1)} - \hat{\theta}_a^{(2)}\right\|^2 \leq \left\|\hat{P}_{\cdot,i}^{(1)} + \hat{P}_{\cdot,i}^{(2)} - \hat{\theta}_{z_i^*}^{(1)} - \hat{\theta}_{z_i^*}^{(2)}\right\|^2\right\} \\ &= \mathbb{I}\left\{2\left\langle \hat{P}_{\cdot,i}^{(1)} - \hat{\theta}_{z_i^*}^{(1)}, \hat{\theta}_{z_i^*}^{(1)} - \hat{\theta}_a^{(1)}\right\rangle + \left\|\hat{\theta}_{z_i^*}^{(1)} - \hat{\theta}_a^{(1)}\right\|^2 \leq 2\left\langle \hat{P}_{\cdot,i}^{(2)}, \hat{\theta}_a^{(2)} - \hat{\theta}_{z_i^*}^{(2)}\right\rangle - \left\|\hat{\theta}_a^{(2)}\right\|^2 + \left\|\hat{\theta}_{z_i^*}^{(2)}\right\|^2\right\}\end{aligned}$$

We denote $\tau'' = o(1)$ to be another sequence which we will specify later. Then the above display can be decomposed and upper bounded by

$$\begin{aligned}\mathbb{I}\{\hat{z}_i = a\} \leq& \mathbb{I}\left\{\left\|\hat{\theta}_{z_i^*}^{(1)} - \hat{\theta}_a^{(1)}\right\| - \frac{\tau''\Delta^2 + \left\|\hat{\theta}_{z_i^*}^{(2)}\right\|^2}{\left\|\hat{\theta}_{z_i^*}^{(1)} - \hat{\theta}_a^{(1)}\right\|} \leq 2\left\|\hat{P}_{\cdot,i}^{(1)} - \hat{\theta}_{z_i^*}^{(1)}\right\|\right\} \\ &+ \mathbb{I}\left\{\tau''\Delta^2 \leq 2\left\langle \hat{P}_{\cdot,i}^{(2)}, \hat{\theta}_a^{(2)} - \hat{\theta}_{z_i^*}^{(2)}\right\rangle\right\} =: A_{i,a} + B_{i,a}.\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}\ell(\hat{z}, z^*) &\leq \frac{1}{n}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}\mathbb{I}\{\hat{z}_i = a\} \\ &\leq \mathbb{P}\left(\mathcal{F}^\complement\right) + \frac{1}{n}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}A_{i,a}\mathbb{I}\{\mathcal{F}\} + \frac{1}{n}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}B_{i,a}\mathbb{I}\{\mathcal{F}\}. \end{aligned} \tag{65}$$

We are going to establish upper bounds first for $n^{-1}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}B_{i,a}\mathbb{I}\{\mathcal{F}\}$ and then for $n^{-1}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}A_{i,a}\mathbb{I}\{\mathcal{F}\}$.

(Analysis on $n^{-1}\sum_{i\in[n]}\sum_{a\neq z_i^*}\mathbb{E}B_{i,a}\mathbb{I}\{\mathcal{F}\}$). For $\sum_{i\in[n]}\sum_{a\neq z_i^*}\mathbb{E}B_{i,a}\mathbb{I}\{\mathcal{F}\}$, we can di-

rectly use upper bounds established in Section 4.4.3 of [27][1]. It proves that for any $i \in [n]$,

$$\sum_{a\in[k]:a\neq z_i^*} B_{i,a}\mathbb{I}\left\{\mathcal{F}\cap\mathcal{T}\right\} \leq 2\exp\left(-\frac{1}{2}\left(c_4\frac{\tau''\Delta}{k^{\frac{7}{2}}\tau^2\beta^{-\frac{1}{2}}(1+\frac{p}{n})\sigma}\sqrt{\frac{n-k}{3n}}\right)^2\frac{\Delta^2}{\sigma^2}\right),$$

where $c_4 > 0$ is some constant, and $\mathcal{T}$ is some high-probability event in the sense that

$$\mathbb{P}\left(\mathcal{T}\right) \geq 1 - nk\exp\left(-\frac{(n-k)}{9}\right).$$

Hence,

$$\frac{1}{n}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}B_{i,a}\mathbb{I}\left\{\mathcal{F}\right\} \leq \frac{1}{n}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}B_{i,a}\mathbb{I}\left\{\mathcal{F}\cap\mathcal{T}\right\} + \mathbb{P}\left(\mathcal{T}^{\complement}\right)$$
$$\leq 2\exp\left(-\frac{1}{2}\left(c_4\frac{\tau''\Delta}{k^{\frac{7}{2}}\tau^2\beta^{-\frac{1}{2}}(1+\frac{p}{n})\sigma}\sqrt{\frac{n-k}{3n}}\right)^2\frac{\Delta^2}{\sigma^2}\right) + nk\exp\left(-\frac{(n-k)}{9}\right).$$

(Analysis on $n^{-1}\sum_{i\in[n]}\sum_{a\neq z_i^*}\mathbb{E}A_{i,a}\mathbb{I}\left\{\mathcal{F}\right\}$). We first follow some algebra as in Section 4.4.2 of [27] to simplify $A_{i,a}\mathbb{I}\left\{\mathcal{F}\right\}$. For any $i\in[n]$ and $a\neq z_i^*$, it proves

$$A_{i,a}\mathbb{I}\left\{\mathcal{F}\right\} \leq \mathbb{I}\left\{\left(1 - c_1\tau'' - \frac{c_1k^2\tau\beta^{-\frac{1}{2}}\sqrt{1+\frac{p}{n}}\sigma}{\Delta}\right)\Delta \leq 2\left\|\hat{P}^{(1)}_{\cdot,i} - \hat{\theta}^{(1)}_{z_i^*}\right\|\right\}\mathbb{I}\left\{\mathcal{F}\right\}, \tag{66}$$

for some constant $c_1 > 0$. Still working on the event $\mathcal{F}$, it also proves

$$\left\|\hat{P}^{(1)}_{\cdot,i} - \hat{\theta}^{(1)}_{z_i^*}\right\| \leq \left\|\hat{P}^{(1)}_{\cdot,i} - \hat{U}_{1:r}\hat{U}_{1:r}^T\theta^*_{z_i^*}\right\| + 8\sqrt{2}\sqrt{\beta^{-1}k^2\left(1+\frac{p}{n}\right)}\sigma. \tag{67}$$

Our following analysis on $A_{i,a}\mathbb{I}\left\{\mathcal{F}\right\}$ is different from the rest proof in Section 4.4.2 of [27]. Note that $\hat{P}^{(1)}_{\cdot,i} - \hat{U}_{1:r}\hat{U}_{1:r}^T\theta^*_{z_i^*} = \hat{U}_{1:r}\hat{U}_{1:r}^TX_i - \hat{U}_{1:r}\hat{U}_{1:r}^T\theta^*_{z_i^*} = \hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i$. Then (66) and (67) give

$$A_{i,a}\mathbb{I}\left\{\mathcal{F}\right\} \leq \mathbb{I}\left\{\left(1 - c_2\tau'' - \frac{c_2k^2\tau\beta^{-\frac{1}{2}}\left(1+\sqrt{\frac{p}{n}}\right)\sigma}{\Delta}\right)\Delta \leq 2\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\|\right\}\mathbb{I}\left\{\mathcal{F}\right\}, \tag{68}$$

where we use $\tau\to\infty$ and the fact that $1+\sqrt{p/n}$, $\sqrt{1+p/n}$ are of the same order.

Recall the definition of X_{-i} in (8) and $\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T$ is the leave-one-out counterpart of $\hat{U}_{1:r}\hat{U}_{1:r}^T$. For (68), we can decompose $\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\|$ into

$$\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\| \leq \left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\| + \left\|\hat{U}_{1:r}\hat{U}_{1:r}^T - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right\|_{\mathrm{F}}\left\|\epsilon_i\right\|.$$

To upper bound $\|\hat{U}_{1:r}\hat{U}_{1:r}^T - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\|_{\mathrm{F}}$, we are going to use Theorem 2.3. Since (63)-(64) hold, under the assumption $\beta n/k^4 \geq 100$, we have

$$\frac{\lambda_r - \lambda_{r+1}}{\max\left\{\|E\|, \sqrt{\frac{k^2}{n\beta}}\lambda_{r+1}\right\}} \geq \frac{\tau}{2}.$$

[1]The model in [27] assumes $\{\epsilon_j\}\overset{iid}{\sim}\mathcal{N}(0,I)$ while in this paper we assume $\{\epsilon_j\}\overset{iid}{\sim}\mathcal{N}(0,\sigma^2I)$. To directly use results from [27], we can re-scale our data to have $X_j' = X_j/\sigma$ for all $j\in[n]$. Then $\{X_j'\}$ has $\mathcal{N}(0,I)$ noise and the separation between their centers becomes Δ/σ. Then all the results from [27] can be used here with Δ replaced by Δ/σ.

Applying Theorem 2.3, we have

$$\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T - \hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\right\|_{\mathrm{F}} \le \frac{256\sqrt{rk}}{\sqrt{n\beta}} + \frac{256\left\|\hat{U}_{-i,1:r}\hat{U}_{-i.1:r}^T\epsilon_i\right\|}{\lambda_r}.$$

Hence,

$$\begin{aligned}
\left\|\hat{U}_{1:r}\hat{U}_{1:r}^T\epsilon_i\right\| &\le \left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\| + \left(\frac{256\sqrt{rk}}{\sqrt{n\beta}} + \frac{256\left\|\hat{U}_{-i,1:r}\hat{U}_{-i.1:r}^T\epsilon_i\right\|}{\lambda_r}\right)\|E\| \\
&= \frac{256k\|E\|}{\sqrt{n\beta}} + \left(1 + \frac{256\|E\|}{\lambda_r}\right)\left\|\hat{U}_{-i,1:r}\hat{U}_{-i.1:r}^T\epsilon_i\right\| \\
&\le \frac{256\sqrt{2}k(\sqrt{n}+\sqrt{p})\sigma}{\sqrt{n\beta}} + \left(1 + \frac{256\sqrt{2}(\sqrt{n}+\sqrt{p})\sigma}{\tau\sqrt{n+p}\sigma}\right)\left\|\hat{U}_{-i,1:r}\hat{U}_{-i.1:r}^T\epsilon_i\right\| \\
&\le 512k\beta^{-0.5}\left(1+\sqrt{\frac{p}{n}}\right)\sigma + \left(1+512\tau^{-1}\right)\left\|\hat{U}_{-i,1:r}\hat{U}_{-i.1:r}^T\epsilon_i\right\|,
\end{aligned}$$

where in the second to the last inequality, we use (63) for λ_r and the event $\mathcal{F}$ for $\|E\|$. Then (68) leads to

$$\begin{aligned}
A_{i,a}\mathbb{I}\{\mathcal{F}\} &\le \mathbb{I}\left\{\left(1 - c_3\tau'' - \frac{c_3k^2\tau\beta^{-\frac{1}{2}}\left(1+\sqrt{\frac{p}{n}}\right)\sigma}{\Delta}\right)\Delta \le 2\left(1+512\tau^{-1}\right)\left\|\hat{U}_{-i,1:r}\hat{U}_{-i.1:r}^T\epsilon_i\right\|\right\}\mathbb{I}\{\mathcal{F}\} \\
&\le \mathbb{I}\left\{\left(1 - c_4\left(\frac{k^2\tau\beta^{-\frac{1}{2}}\left(1+\sqrt{\frac{p}{n}}\right)\sigma}{\Delta} + \tau^{-1}\right)\right)\Delta \le 2\left\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\right\|\right\},
\end{aligned}$$

where $c_3, c_4 > 0$ are some constants. As long as $1 - c_4(k^2\tau\beta^{-0.5}(1+\sqrt{p/n})\sigma/\Delta + \tau^{-1}) > 1/2$, we can use Lemma E.2 to calculate the tail probability of $\|\hat{U}_{-i,1:r}\hat{U}_{-i,1:r}^T\epsilon_i\|$. Following the proof of Theorem 3.1, we have

$$\mathbb{E}A_{i,a}\mathbb{I}\{\mathcal{F}\} \le \exp\left(-\left(1 - c_5\left(\frac{k^2\tau\beta^{-\frac{1}{2}}\left(1+\sqrt{\frac{p}{n}}\right)\sigma}{\Delta} + \tau^{-1}\right)\right)\frac{\Delta^2}{8\sigma^2}\right),$$

for some constant $c_5 > 0$. Then we have,

$$n^{-1}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}A_{i,a}\mathbb{I}\{\mathcal{F}\} \le k\exp\left(-\left(1 - c_5\left(\frac{k^2\tau\beta^{-\frac{1}{2}}\left(1+\sqrt{\frac{p}{n}}\right)\sigma}{\Delta} + \tau^{-1}\right)\right)\frac{\Delta^2}{8\sigma^2}\right).$$

(Obtaining the Final Result.) From (65) and the above upper bounds on $n^{-1}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}B_{i,a}\mathbb{I}\{\mathcal{F}\}$ and $n^{-1}\sum_{i\in[n]}\sum_{a\in[k]:a\neq z_i^*}\mathbb{E}A_{i,a}\mathbb{I}\{\mathcal{F}\}$, we have

$$\begin{aligned}
\mathbb{E}\ell(\hat{z},z^*) &\le e^{-0.08n} + 2\exp\left(-\frac{1}{2}\left(c_4\frac{\tau''\Delta}{k^{\frac{7}{2}}\tau^2\beta^{-\frac{1}{2}}(1+\frac{p}{n})\sigma}\sqrt{\frac{n-k}{3n}}\right)^2\frac{\Delta^2}{\sigma^2}\right) + nk\exp\left(-\frac{(n-k)}{9}\right) \\
&\quad + k\exp\left(-\left(1 - c_5\left(\frac{k^2\tau\beta^{-\frac{1}{2}}\left(1+\sqrt{\frac{p}{n}}\right)\sigma}{\Delta} + \tau^{-1}\right)\right)\frac{\Delta^2}{8\sigma^2}\right).
\end{aligned}$$

Since we assume $\beta n/k^4 \geq 100$, we have $(n-k)/n > 0.99$. Hence, under the assumption that $\Delta/(k^{3.5}\beta^{-0.5}(1+\frac{p}{n})\sigma) \to \infty$, we can take τ, τ'' to be

$$\tau = \tau''^{-1} := \left(\frac{\Delta}{k^{3.5}\beta^{-0.5}\left(1+\frac{p}{n}\right)\sigma}\right)^{0.25}$$

such that $\tau \to \infty$ and $\tau'' = o(1)$. Then for some constant $c_6 > 0$, we have

$$\begin{aligned}\mathbb{E}\ell(\hat{z}, z^*) &\leq e^{-0.08n} + 2\exp\left(-\frac{c_4^2}{12}\left(\frac{\Delta}{k^{3.5}\beta^{-0.5}\left(1+\frac{p}{n}\right)\sigma}\right)^{0.5}\frac{\Delta^2}{\sigma^2}\right) + nke^{-0.1n} \\ &\quad + k\exp\left(-\left(1-2c_5\left(\frac{\Delta}{k^{3.5}\beta^{-0.5}\left(1+\frac{p}{n}\right)\sigma}\right)^{-0.25}\right)\frac{\Delta^2}{8\sigma^2}\right) \\ &\leq \exp\left(-\left(1-c_6\left(\frac{\Delta}{k^{3.5}\beta^{-0.5}\left(1+\frac{p}{n}\right)\sigma}\right)^{-0.25}\right)\frac{\Delta^2}{8\sigma^2}\right) + 2e^{-0.08n}.\end{aligned}$$

APPENDIX D: PROOFS OF RESULTS IN SECTION 3.6

D.1. Proof of Theorem 3.4. The proof of Theorem 3.4 relies on the following entrywise decomposition that is analogous to Lemma 3.2 but in an opposite direction. Note the the singular vectors $\hat{u}_1$, and $\{\hat{u}_{1,-i}\}_{i\in[n]}$ are all identifiable up to sign. Without loss of generality, we assume $\langle \hat{u}_1, u_1\rangle \geq 0$ and $\langle \hat{u}_{1,-i}, u_1\rangle \geq 0$ for all $i \in [n]$.

LEMMA D.1. *Consider the model (28). Let $\phi \in \Phi$ be the permutation such that $\ell(\check{z}, z^*) = \frac{1}{n}|\{i \in [n] : \check{z}_i \neq \phi(z_i^*)\}|$. Then there exists a constants $C, C_1 > 0$ such that if*

$$\frac{\Delta}{\beta^{-0.5}n^{-0.5}\|E\|} \geq C, \tag{69}$$

then for any $i \in [n]$,

$$\mathbb{I}\{\check{z}_i \neq \phi(z_i^*)\} \geq \mathbb{I}\left\{\left(1+\frac{C_1\beta^{-0.5}n^{-0.5}\|E\|}{\Delta}\right)\Delta \leq -2(\hat{u}_{1,-i}^T\epsilon_i)\mathit{sign}(u_1^T\theta_{\phi(z_i^*)})\right\}. \tag{70}$$

PROOF. The proof mainly follows the proofs of Lemma 3.1 and Lemma 3.2 with some modifications such as adding a negative term instead of a positive term in order to obtain a lower bound.

We first write $\check{z}$ equivalently as

$$\left(\check{z}, \{\check{\theta}_j\}_{j=1}^2\right) = \underset{z\in[2]^n, \{\theta_j\}_{j=1}^2\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i\in[n]} \left\|\hat{u}_1\hat{u}_1^T X_i - \theta_{z_i}\right\|^2,$$

where $\check{\theta}_a = \hat{u}_1\check{c}_a$ for each $a \in [2]$. Note that $k = 2$. From Proposition 3.1, we have

$$\frac{1}{n}|\{i \in [n] : \check{z}_i \neq \phi(z_i^*)\}| \leq \frac{C_0 k\|E\|^2}{n\Delta^2},$$

and

$$\max_{a\in[2]}\left\|\check{\theta}_{\phi(a)} - \theta_a^*\right\| \leq C_0\beta^{-0.5}kn^{-0.5}\|E\|, \tag{71}$$

for some permutation $\phi : [2] \to [2]$ and some constant $C_0 > 0$. Without loss of generality, assume $\phi = \text{Id}$.

Recall that $\theta_1^* = -\theta_2^* = \delta \mathbb{1}_p$, $u_1 = 1/\sqrt{p}\mathbb{1}_p$, $\lambda_1 = \delta\sqrt{np} = \frac{\Delta\sqrt{n}}{2}$, and $|u_1^T(\theta_{z_i^*}^* - (-\theta_{z_i^*}^*))| = 2\delta\sqrt{p} = \Delta$. By Davis-Kahan Theorem, we have

$$\min_{s\in\pm 1} \|\hat{u}_1 - su_1\| \leq \frac{\|E\|}{\lambda_1} = \frac{2\|E\|}{\sqrt{n}\Delta} \leq 1/16,$$

where the last inequality is due to the assumption (16). Since we assume $\langle \hat{u}_1, u_1\rangle \geq 0$, we have $\|\hat{u}_1 - su_1\| = \min_{s\in\pm 1}\|\hat{u}_1 - su_1\|$.

Consider any $i \in [n]$ and any $a \in [2]$ such that $a \neq z_i^*$. Note that for any scalars x, y, w, if $|x-y| \leq |x-w|$, we have equivalently $\text{sign}(w-y)(y+w)/2 \geq \text{sign}(w-y)x$. Since $(y+w)/2 = (y-w)/2 + w$, a sufficient condition is $|w-y|/2 + |w| \leq (-\text{sign}(w-y))x$. Hence, we have

$$\begin{aligned}
&\mathbb{I}\left\{\left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_a\right\| \leq \left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_{z_i^*}\right\|\right\} \\
&= \mathbb{I}\left\{\left|\hat{u}_1^T X_i - \hat{u}_1^T\check{\theta}_a\right| \leq \left|\hat{u}_1^T X_i - \hat{u}_1^T\check{\theta}_{z_i^*}\right|\right\} \\
&= \mathbb{I}\left\{\left|\hat{u}_1^T\epsilon_i - \hat{u}_1^T\left(\check{\theta}_a - \theta_{z_i^*}^*\right)\right| \leq \left|\hat{u}_1^T\epsilon_i - \hat{u}_1^T\left(\check{\theta}_{z_i^*} - \theta_{z_i^*}^*\right)\right|\right\} \\
&\geq \mathbb{I}\left\{\frac{1}{2}\left|\hat{u}_1^T(\check{\theta}_{z_i^*} - \check{\theta}_a)\right| + \left|\hat{u}_1^T\left(\check{\theta}_{z_i^*} - \theta_{z_i^*}^*\right)\right| \leq -(\hat{u}_1^T\epsilon_i)\text{sign}(\hat{u}_1^T(\check{\theta}_{z_i^*} - \check{\theta}_a))\right\} \\
&\geq \mathbb{I}\left\{\left\|\check{\theta}_{z_i^*} - \check{\theta}_a\right\| + 2\left\|\check{\theta}_{z_i^*} - \theta_{z_i^*}^*\right\| \leq -2(\hat{u}_1^T\epsilon_i)\text{sign}(\hat{u}_1^T(\check{\theta}_{z_i^*} - \check{\theta}_a))\right\}.
\end{aligned}$$

We are going to show $\text{sign}(\hat{u}_1^T(\check{\theta}_{z_i^*} - \check{\theta}_a)) = \text{sign}(u_1^T(\theta_{z_i^*}^* - \theta_a^*))$. By (71), we have

$$\begin{aligned}
\left\langle\check{\theta}_{z_i^*} - \check{\theta}_a, \theta_{z_i^*}^* - \theta_a^*\right\rangle &= \left\|\theta_{z_i^*}^* - \theta_a^*\right\|^2 + \left\langle\check{\theta}_{z_i^*} - \theta_{z_i^*}^*, \theta_{z_i^*}^* - \theta_a^*\right\rangle + \left\langle\check{\theta}_a - \theta_a^*, \theta_{z_i^*}^* - \theta_a^*\right\rangle \\
&\geq \Delta^2\left(1 - \frac{2C_0 k\beta^{-0.5}n^{-0.5}\|E\|}{\Delta}\right) \\
&> 0,
\end{aligned}$$

where the last inequality holds as long as $\Delta > 2C_0\beta^{-0.5}kn^{-0.5}\|E\|$. Due to the fact $\theta_{z_i^*}^* - \theta_a^* \in \text{span}(u_1)$, $\check{\theta}_{z_i^*} - \check{\theta}_a^* \in \text{span}(\hat{u}_1)$, and $\langle\hat{u}_1, u_1\rangle \geq 0$, if $u_1, \theta_{z_i^*}^* - \theta_a^*$ are in the same direction, then $\hat{u}_1, \check{\theta}_{z_i^*} - \check{\theta}_a^*$ must also be in the same direction, and vice versa. Hence, we have $\text{sign}(\hat{u}_1^T(\check{\theta}_{z_i^*} - \check{\theta}_a)) = \text{sign}(u_1^T(\theta_{z_i^*}^* - \theta_a^*))$. Thus,

$$\begin{aligned}
&\mathbb{I}\left\{\left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_a\right\| \leq \left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_{z_i^*}\right\|\right\} \\
&\geq \mathbb{I}\left\{\left\|\check{\theta}_{z_i^*} - \check{\theta}_a\right\| + 2\left\|\check{\theta}_{z_i^*} - \theta_{z_i^*}^*\right\| \leq -2(\hat{u}_1^T\epsilon_i)\text{sign}(u_1^T(\theta_{z_i^*}^* - \theta_a^*))\right\}.
\end{aligned}$$

Following the same analysis as in the proof of Lemma 3.1, we can get the following result that is analogous to (45):

$$\begin{aligned}
&\mathbb{I}\left\{\left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_a\right\| \leq \left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_{z_i^*}\right\|\right\} \\
&\geq \mathbb{I}\left\{\left(1 + \frac{4C_0\beta^{-0.5}kn^{-0.5}\|E\|}{\Delta}\right)\Delta \leq -2(\hat{u}_1^T\epsilon_i)\text{sign}(u_1^T(\theta_{z_i^*}^* - \theta_a^*))\right\}.
\end{aligned}$$

Next, we are going to decompose $\hat{u}_1^T\epsilon_i$ following the proof of Lemma 3.2. Denote $\hat{u}_{1,-i}$ be the leave-one-out counterpart of $\hat{u}_1$, i.e., $\hat{u}_{1,-i}$ is the leading left singular vector of X_{-i}.

Since we assume $\langle \hat{u}_{1,-i}, u_1\rangle \geq 0$, we have $\|\hat{u}_{1,-i} - u_1\| \leq 2\|E\|/(\sqrt{n-1}\Delta)$. As a result, we have $\|\hat{u}_{1,-i} - \hat{u}_1\| \leq 4\|E\|/(\sqrt{n-1}\Delta)$ which leads to

$$\langle \hat{u}_{1,-i}, \hat{u}_1\rangle \geq 1 - 4\|E\|/(\sqrt{n-1}\Delta) > 0. \tag{72}$$

We have the following decomposition:

$$\begin{aligned}
&(\hat{u}_1^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\\
&= \left\langle \hat{u}_1, \hat{u}_1\hat{u}_1^T\epsilon_i\right\rangle \text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\\
&= \left\langle \hat{u}_1, (\hat{u}_{1,-i}\hat{u}_{1,-i}^T)\epsilon_i\right\rangle \text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a)) + \left\langle \hat{u}_1, (\hat{u}_1\hat{u}_1^T - \hat{u}_{1,-i}\hat{u}_{1,-i}^T)\epsilon_i\right\rangle \text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\\
&= \langle \hat{u}_1, \hat{u}_{1,-i}\rangle (\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a)) + \left\langle \hat{u}_1, (\hat{u}_1\hat{u}_1^T - \hat{u}_{1,-i}\hat{u}_{1,-i}^T)\epsilon_i\right\rangle \text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\\
&\leq \langle \hat{u}_1, \hat{u}_{1,-i}\rangle (\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a)) + \left\|\hat{u}_1\hat{u}_1^T - \hat{u}_{1,-i}\hat{u}_{1,-i}^T\right\| \|\epsilon_i\|.
\end{aligned}$$

Note that $\lambda_1/\|E\| = \Delta\sqrt{n}/(2\|E\|)$ is greater than 16 under the assumption (69) holds for a large constant C. From Theorem 2.2 we have

$$\left\|\hat{u}_1\hat{u}_1^T - \hat{u}_{1,-i}\hat{u}_{1,-i}^T\right\| \leq \frac{128}{\lambda_1/\|E\|}\left(\frac{k}{\sqrt{\beta n}} + \frac{\left\|\hat{u}_{1,-i}\hat{u}_{1,-i}^T\epsilon_i\right\|}{\lambda_1}\right).$$

Then,

$$\begin{aligned}
&(\hat{u}_1^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\\
&\leq \langle \hat{u}_1, \hat{u}_{1,-i}\rangle (\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a)) + \left(\frac{128k}{\sqrt{n\beta}(\lambda_1/\|E\|)} + \frac{128\left\|\hat{u}_{1,-i}\hat{u}_{1,-i}^T\epsilon_i\right\|}{\lambda_1^2/\|E\|}\right)\|E\|\\
&= \langle \hat{u}_1, \hat{u}_{1,-i}\rangle (\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a)) + \frac{256n^{-0.5}k\beta^{-0.5}\|E\|^2}{\Delta} + \frac{512\left|\hat{u}_{1,-i}^T\epsilon_i\right| n^{-1}\|E\|^2}{\Delta^2}.
\end{aligned}$$

So far we have obtained

$$\begin{aligned}
&\mathbb{I}\left\{\left\|\hat{u}_1\hat{u}_1^TX_i - \check{\theta}_a\right\| \leq \left\|\hat{u}_1\hat{u}_1^TX_i - \check{\theta}_{z_i^*}\right\|\right\}\\
&\geq \mathbb{I}\Bigg\{\left(1 + \frac{4C_0\beta^{-0.5}kn^{-0.5}\|E\|}{\Delta}\right)\Delta \leq -2\langle \hat{u}_1, \hat{u}_{1,-i}\rangle (\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\\
&\quad - \frac{256n^{-0.5}k\beta^{-0.5}\|E\|^2}{\Delta} - \frac{512\left|\hat{u}_{1,-i}^T\epsilon_i\right| n^{-1}\|E\|^2}{\Delta^2}\Bigg\}\\
&= \mathbb{I}\Bigg\{\left(1 + \frac{4C_0\beta^{-0.5}kn^{-0.5}\|E\|}{\Delta} + \frac{256n^{-0.5}k\beta^{-0.5}\|E\|^2}{\Delta^2}\right)\Delta\\
&\quad \leq -2\langle \hat{u}_1, \hat{u}_{1,-i}\rangle (\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a)) - \frac{512\left|\hat{u}_{1,-i}^T\epsilon_i\right| n^{-1}\|E\|^2}{\Delta^2}\Bigg\}.
\end{aligned}$$

From (72) we have

$$\begin{aligned}
\langle \hat{u}_{1,-i}, \hat{u}_1\rangle - \frac{512n^{-1}\|E\|^2}{\Delta^2} &\geq 1 - 4\frac{\|E\|(n-1)^{-0.5}}{\Delta} - \frac{512n^{-1}\|E\|^2}{\Delta^2}\\
&\geq 1 - \frac{16n^{-0.5}\|E\|}{\Delta} \geq \frac{1}{2},
\end{aligned}$$

assuming $\frac{\Delta}{n^{-0.5}\|E\|} \geq 64$. For any $x, y, z, w \in \mathbb{R}$ such that $x \geq 0$, $1 \geq z \geq 0$, and $z|y| > w \geq 0$, we have $\mathbb{I}\{x \leq zy - w\} \geq \mathbb{I}\{x \leq (z - w/|y|)\, y\}$. We then have,

$$\begin{aligned}
&\mathbb{I}\left\{\left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_a\right\| \leq \left\|\hat{u}_1\hat{u}_1^T X_i - \check{\theta}_{z_i^*}\right\|\right\} \\
&\geq \mathbb{I}\Bigg(\left(1 + \frac{4C_0\beta^{-0.5}kn^{-0.5}\|E\|}{\Delta} + \frac{256n^{-0.5}k\beta^{-0.5}\|E\|^2}{\Delta^2}\right)\Delta \\
&\qquad \leq -2\left(1 - \frac{16n^{-0.5}\|E\|}{\Delta}\right)(\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\Bigg) \\
&\geq \mathbb{I}\left\{\left(1 + \frac{C_1\beta^{-0.5}n^{-0.5}\|E\|}{\Delta}\right)\Delta \leq -2(\hat{u}_{1,-i}^T\epsilon_i)\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a))\right\}.
\end{aligned}$$

Since $\theta^*_a = -\theta^*_{z_i^*}$, we have $\text{sign}(u_1^T(\theta^*_{z_i^*} - \theta^*_a)) = \text{sign}(u_1^T\theta^*_{z_i^*})$. The proof is complete. □

PROOF OF THEOREM 3.4. Recall that $\lambda_1 = \Delta\sqrt{n}/2$. Same as the proof of Theorem 3.1, we work on the high-probability event (46).

For the upper bound, from Lemma 3.2, there exists some $\phi \in \Phi$ such that for any $i \in [n]$,

$$\mathbb{I}\{\hat{z}_i \neq \phi(z_i^*)\} \leq \mathbb{I}\left\{\left(1 - C_1\psi_3^{-1}\right)\Delta \leq 2\left\|\hat{u}_{1,-i}\hat{u}_{-i}^T\epsilon_i\right\|\right\} = \mathbb{I}\left\{\left(1 - C_1\psi_3^{-1}\right)\Delta \leq 2\left|\hat{u}_{1,-i}^T\epsilon_i\right|\right\},$$

for some $C_1 > 0$, where the last inequality is due to that ψ_3 is large. By Davis-Kahan Theorem, we know there exists some $s_i \in \{-1, 1\}$ such that $\|\hat{u}_{1,-i} - s_iu_1\| \leq 2\|E\|/(\sqrt{n-1}\Delta) \leq 4\psi_3^{-1}$. Since $\langle \hat{u}_{1,-i}, u_1\rangle \geq 0$ is assumed, we have $s_i = 1$ for all $i \in [n]$. Then

$$\begin{aligned}
\mathbb{I}\{\hat{z}_i \neq \phi(z_i^*)\} &\leq \mathbb{I}\left\{\left(1 - C_1\psi_3^{-1}\right)\Delta \leq 2\left|u_1^T\epsilon_i\right| + 2\left|(\hat{u}_{1,-i} - s_iu_1)^T\epsilon_i\right|\right\} \\
&\leq \mathbb{I}\left\{\left(1 - (C_1 + C_2)\psi_3^{-1}\right)\Delta \leq 2\left|u_1^T\epsilon_i\right|\right\} + \mathbb{I}\left\{C_2\psi_3^{-1}\Delta \leq 2\left|(\hat{u}_{1,-i} - s_iu_1)^T\epsilon_i\right|\right\},
\end{aligned}$$

where $C_2 > 0$ is a constant whose value will be determined later. Due to the independence of $\hat{u}_{1,-i} - s_iu_1$ and ϵ_i, we have $(\hat{u}_{1,-i} - s_iu_1)^T\epsilon_i \sim \text{SG}(16\psi_3^{-2}\sigma^2)$ and then

$$\mathbb{E}\mathbb{I}\left\{C_2\Delta \leq 2\left|(\hat{u}_{1,-i} - s_iu_1)^T\epsilon_i\right|\right\} \leq 2\exp\left(-\frac{C_2^2\Delta^2}{128\sigma^2}\right).$$

On the other hand, $u_1^T\epsilon_i = p^{-\frac{1}{2}}\sum_{j=1}^p \epsilon_{i,j}$ where $\{\epsilon_{i,j}\}_{j\in[p]}$ are i.i.d. with variance $\bar{\sigma}^2$, which can be approximated by a normal distribution. Since the distribution F is sub-Gaussian, its moment generating function exists. Then we can use the following KMT quantile inequality (see Proposition [KMT] of [30]). Let $Y \overset{d}{=} \bar{\sigma}^{-1}p^{-\frac{1}{2}}\sum_{j=1}^p \epsilon_{i,j}$. There exist some constants $D, \eta > 0$ and $Z \sim \mathcal{N}(0,1)$, such that whenever $|Y| \leq \eta\sqrt{p}$, we have

$$|Y - Z| \leq \frac{DY^2}{\sqrt{p}} + \frac{D}{\sqrt{p}}.$$

Then,

$$\begin{aligned}
&\mathbb{E}\mathbb{I}\left\{\left(1 - (C_1 + C_2)\psi_3^{-1}\right)\Delta \leq 2\left|u_1^T\epsilon_i\right|\right\} \\
&= \mathbb{E}\mathbb{I}\left\{\left(1 - (C_1 + C_2)\psi_3^{-1}\right)\frac{\Delta}{\bar{\sigma}} \leq 2|Y|\right\} \\
&\leq \mathbb{E}\mathbb{I}\left\{\left(1 - (C_1 + C_2)\psi_3^{-1}\right)\frac{\Delta}{\bar{\sigma}} \leq 2|Z| + \frac{2DY^2}{\sqrt{p}} + \frac{2D}{\sqrt{p}}\right\} + \mathbb{E}\mathbb{I}\{|Y| > \eta\sqrt{p}\} \\
&\leq \mathbb{E}\mathbb{I}\left\{\left(1 - (C_1 + C_2 + C_3 + 2D)\psi_3^{-1}\right)\frac{\Delta}{\bar{\sigma}} \leq 2|Z|\right\} + \mathbb{E}\mathbb{I}\left\{\frac{2DY^2}{\sqrt{p}} \geq C_3\right\} + \mathbb{E}\mathbb{I}\{|Y| > \eta\sqrt{p}\},
\end{aligned}$$

where $C_3 > 0$ is a constant. Using the fact that $Y \sim \mathrm{SG}(1)$ with zero mean, we have

$$\mathbb{E}\mathbb{I}\left\{\left(1-(C_1+C_2)\psi_3^{-1}\right)\Delta \leq 2\left|u_1^T\epsilon_i\right|\right\}$$
$$\leq 2\exp\left(-\frac{\left(1-(C_1+C_2+C_3+2D)\psi_3^{-1}\right)^2\Delta^2}{8\bar{\sigma}^2}\right)+2\exp\left(-\frac{C_3\sqrt{p}}{4D}\right)+2\exp\left(-\frac{\eta^2 p}{2}\right).$$

Then we have

$$\mathbb{E}\ell(\check{z},z^*)$$
$$\leq \frac{1}{n}\sum_{i=1}^n \mathbb{E}\mathbb{I}\left\{\left(1-(C_1+C_2)\psi_3^{-1}\right)\Delta \leq 2\left|u_1^T\epsilon_i\right|\right\}+\frac{1}{n}\sum_{i=1}^n \mathbb{E}\mathbb{I}\left\{C_2\Delta\leq 2\left|\left(\hat{u}_{1,-i}-s_iu_1\right)^T\epsilon_i\right|\right\}+e^{-0.5n}$$
$$\leq 2\exp\left(-\frac{\left(1-(C_1+C_2+C_3+2D)\psi_3^{-1}\right)^2\Delta^2}{8\bar{\sigma}^2}\right)$$
$$+2\exp\left(-\frac{C_2^2\Delta^2}{128\sigma^2}\right)+2\exp\left(-\frac{C_3\sqrt{p}}{4D}\right)+2\exp\left(-\frac{\eta^2 p}{2}\right)+e^{-0.5n},$$

where $e^{-0.5n}$ is the probability that (46) does not hold. Since $\sigma \leq C\bar{\sigma}$, when C_2 is chosen to satisfy $C_2^2/(128C^2) \geq 16$, we have

$$\mathbb{E}\ell(\check{z},z^*) \leq 2\exp\left(-\frac{\left(1-C''\psi_3^{-1}\right)^2\Delta^2}{8\bar{\sigma}^2}\right)+\exp\left(-C''\sqrt{p}\right)+e^{-0.5n},$$

for some constant $C'' > 0$.

For the lower bound, from (70) we know

$$\mathbb{I}\left\{\check{z}_i\neq\phi(z_i^*)\right\}\geq \mathbb{I}\left\{\left(1+C_4\psi_3^{-1}\right)\Delta\leq -2(\hat{u}_{1,-i}^T\epsilon_i)\mathrm{sign}(u_1^T(\theta_{\phi(z_i^*)}-\theta_{3-\phi(z_i^*)}))\right\},$$

for some constant $C_4 > 0$ assuming ψ_3 is large. Using the same argument as in the upper bound, we are going to decompose $\hat{u}_{1,-i}^T\epsilon_i$ into $u_1^T\epsilon_i$ and $(\hat{u}_{1,-i}-y_1)^T\epsilon_i$. Hence,

$$\begin{aligned}\mathbb{I}\left\{\check{z}_i\neq\phi(z_i^*)\right\} &\geq \mathbb{I}\left\{\left(1+C_4\psi_3^{-1}\right)\Delta\leq -2(u_1^T\epsilon_i)\mathrm{sign}(u_1^T(\theta_{\phi(z_i^*)}-\theta_{3-\phi(z_i^*)}))-2\left|(\hat{u}_{1,-i}-s_iu_1)^T\epsilon_i\right|\right\}\\ &\geq \mathbb{I}\left\{\left(1+(C_4+C_5)\,\psi_3^{-1}\right)\Delta\leq -2(u_1^T\epsilon_i)\mathrm{sign}(u_1^T(\theta_{\phi(z_i^*)}-\theta_{3-\phi(z_i^*)}))\right\}\\ &\quad -\mathbb{I}\left\{C_5\psi_3^{-1}\Delta\leq 2\left|(\hat{u}_{1,-i}-s_iu_1)^T\epsilon_i\right|\right\},\end{aligned}$$

for some constant $C_5 > 0$ whose value to be chosen. Let

$$Y' \overset{d}{=} \bar{\sigma}^{-1}(u_1^T\epsilon_i)\mathrm{sign}(u_1^T(\theta_{\phi(z_i^*)}-\theta_{3-\phi(z_i^*)}))=\mathrm{sign}(u_1^T(\theta_{\phi(z_i^*)}-\theta_{3-\phi(z_i^*)}))\bar{\sigma}^{-1}p^{-\frac{1}{2}}\sum_{j=1}^p\epsilon_{i,j}.$$

Then using the same argument above, there exists some $Z' \sim \mathcal{N}(0,1)$ such that whenever $Y' \leq \eta'\sqrt{p}$, we have $|Y'-Z'| \leq \frac{D'Y'^2}{\sqrt{p}}+\frac{D'}{\sqrt{p}}$ where $D',\eta' > 0$ are constants. Then

$$\mathbb{E}\mathbb{I}\left\{\left(1+(C_4+C_5)\,\psi_3^{-1}\right)\Delta\leq -2(u_1^T\epsilon_i)\mathrm{sign}(u_1^T(\theta_{\phi(z_i^*)}-\theta_{3-\phi(z_i^*)}))\right\}$$
$$=\mathbb{E}\mathbb{I}\left\{\left(1+(C_4+C_5)\,\psi_3^{-1}\right)\frac{\Delta}{\bar{\sigma}}\leq -2Y'\right\}$$
$$\geq \mathbb{E}\mathbb{I}\left\{\left(1+(C_4+C_5)\,\psi_3^{-1}\right)\frac{\Delta}{\bar{\sigma}}\leq -2Z'-\frac{2DY'^2}{\sqrt{p}}-\frac{2d}{\sqrt{p}}\right\}\mathbb{I}\left\{Y'\leq\eta'\sqrt{p}\right\}$$
$$\geq \mathbb{E}\mathbb{I}\left\{\left(1+(C_4+C_5+2D+C_6)\,\psi_3^{-1}\right)\frac{\Delta}{\bar{\sigma}}\leq -2Z'\right\}-\mathbb{E}\mathbb{I}\left\{\frac{2DY'^2}{\sqrt{p}}\geq C_6\right\}-\mathbb{E}\mathbb{I}\left\{Y'>\eta'\sqrt{p}\right\},$$

where $C_6 > 0$ is a constant. Then following the proof of the upper bound, and by a proper choice of C_5, we have

$$\mathbb{E}\ell(\check{z}, z^*) \geq 2\exp\left(-\frac{\left(1 + C'''\psi_3^{-1}\right)^2 \Delta^2}{8\bar{\sigma}^2}\right) - \exp\left(-C'''\sqrt{p}\right) - e^{-0.5n},$$

for some constant $C''' > 0$. $\square$

D.2. Proofs of Lemma 3.4 and Theorem 3.5.

PROOF OF LEMMA 3.4. For the upper bound, we consider the following likelihood ratio test. For any $x \in \mathbb{R}^p$, define the two log-likelihood functions as

$$l_1(x) = \sum_{j=1}^p \log f(x_j - \delta), \text{ and } l_2(x) = \sum_{j=1}^p \log f(x_j + \delta).$$

Then for each $i \in [n]$, define the likelihood ratio test as

$$\hat{z}_i^{\text{LRT}} = \begin{cases} 1, \text{ if } l_1(X_i) \geq l_2(X_i), \\ 2, \text{ otherwise.} \end{cases}$$

Then for any $i \in [n]$ such that $z_i^* = 1$, we have

$$\mathbb{E}\mathbb{I}\left\{\hat{z}_i^{\text{LRT}} = 2\right\} = \mathbb{P}\left(l_2(X_i) > l_1(X_i)\right) = \mathbb{P}\left(\sum_{j=1}^p \log\frac{f(2\delta + \epsilon_{i,j})}{f(\epsilon_{i,j})} > 0\right) = \mathbb{P}\left(\sum_{j=1}^p \log\frac{f_{\frac{\Delta}{\sqrt{p}}}(\epsilon_{i,j})}{f_0(\epsilon_{i,j})} > 0\right),$$

where we use the fact $2\delta = \frac{\Delta}{\sqrt{p}}$. Since Δ is a constant, by local asymptotic normality (c.f., Chapter 7, [41]), we have

$$\sum_{j=1}^p \log\frac{f_{\frac{\Delta}{\sqrt{p}}}(\epsilon_{i,j})}{f_0(\epsilon_{i,j})} \xrightarrow{d} \mathcal{N}\left(-\frac{\mathcal{I}\Delta^2}{2}, \mathcal{I}\Delta^2\right).$$

Then, $\lim_{p\to\infty} \mathbb{E}\mathbb{I}\left\{\hat{z}_i^{\text{LRT}} = 2\right\} \leq C_1 \exp\left(-\mathcal{I}\Delta^2/8\right)$ for some constant $C_1 > 0$. We have the same upper bound if $z_i^* = 2$ instead. Hence,

$$\lim_{p\to\infty} \inf_z \sup_{z^* \in [2]^n} \mathbb{E}\ell(z, z^*) \leq \lim_{p\to\infty} \sup_{z^*\in[2]^n} \mathbb{E}\ell(\hat{z}^{\text{LRT}}, z^*) \leq \exp\left(-\frac{\mathcal{I}\Delta^2}{8}\right).$$

For the lower bound, instead of allowing $z^* \in [2]^n$, we consider a slightly smaller parameter space. Define $\mathcal{Z} = \{z \in [2]^n : z_i = 1, \forall 1 \leq i \leq n/3, z_i = 2, \forall n/3 + 1 \leq i \leq 2n/3\}$. Then for any $z, z' \in \mathcal{Z}$ we have $\ell(z, z') = n^{-1}\sum_{i=1}^n \mathbb{I}\{z_i \neq z_i'\} \leq 1/3$ due to the fact $n^{-1}\sum_{i=1}^n \mathbb{I}\{\phi(z_i) \neq z_i'\} \geq 1/3$ if $\phi \neq \text{Id}$. Hence,

$$\begin{aligned}\inf_z \sup_{z^*\in[2]^n} \mathbb{E}\ell(z, z^*) &\geq \inf_z \sup_{z^*\in\mathcal{Z}} \mathbb{E}\ell(z, z^*) \geq n^{-1} \inf_z \sup_{z^*\in\mathcal{Z}} \mathbb{E}\sum_{i\in[n]} \mathbb{I}\{z_i \neq z_i^*\} \\ &\geq n^{-1} \sum_{i > 2n/3} \inf_{z_i} \sup_{z_i^*\in[2]} \mathbb{E}\mathbb{I}\{z_i \neq z_i^*\} = \frac{1}{3} \inf_{z_n} \sup_{z_n^*\in[2]} \mathbb{E}\mathbb{I}\{z_n \neq z_n^*\},\end{aligned}$$

where it is reduced into a testing problem on whether X_n has mean θ_1^* or θ_2^*. According to the Neyman-Pearson Lemma, the optimal procedure is the likelihood ratio test $\hat{z}_n^{\text{LRT}}$ defined above. By the same argument, we have

$$\lim_{p\to} \inf_z \sup_{z^*\in[2]^n} \mathbb{E}\ell(z, z^*) \geq \frac{1}{3} \lim_{p\to} \inf_{z_n} \sup_{z_n^*\in[2]} \mathbb{E}\mathbb{I}\{z_n \neq z_n^*\} \geq C_2 \exp\left(-\frac{\mathcal{I}\Delta^2}{8}\right),$$

for some constant $C_2 > 0$. $\square$

PROOF OF THEOREM 3.5. First, we have the following connection between the Fisher information $\mathcal{I}$ and the variance $\bar{\sigma}^2$:

$$\mathcal{I}\bar{\sigma}^2 = \left(\int \left(\frac{f'}{f}\right)^2 f \mathrm{d}x\right)\left(\int x^2 f \mathrm{d}x\right) \geq \left(\int \frac{f'}{f} x f \mathrm{d}x\right)^2 = \left(\int x f' \mathrm{d}x\right)^2 = 1,$$

where we use Cauchy-Schwarz inequality and the integral by part $\int x f' \mathrm{d}x = \int x f \mathrm{d}x - \int f \mathrm{d}x = 0 - 1 = -1$. The equation holds if and only if $f'/f \propto x$, which is equivalent to F being normally distributed. □

APPENDIX E: AUXILIARY LEMMAS AND PROPOSITIONS AND THEIR PROOFS

PROPOSITION E.1. *For Y and $\hat{Y}$ defined in (1), we have (2) holds assuming $\sigma_r - \sigma_{r+1} > 2\left\|(I - U_r U_r^T) y_n\right\|$.*

PROOF. Recall the augmented matrix Y' is defined as $(Y, U_r U_r^T y_n)$. Note that $U_r U_r^T Y$ is the best rank-r approximation of Y. Since

$$\left\|\left(I - U_r U_r^T\right) Y'\right\|_{\mathrm{F}} = \left\|\left(\left(I - U_r U_r^T\right) Y, 0\right)\right\|_{\mathrm{F}} = \left\|\left(I - U_r U_r^T\right) Y\right\|_{\mathrm{F}},$$

we have $U_r U_r^T Y'$ also being the best rank-r approximation of Y'. This proves that span(U_r) and $U_r U_r^T$ are also the leading r left singular subspace and projection matrix of Y'. Then $\hat{U}_r \hat{U}_r^T - U_r U_r^T$ is about the perturbation between $\hat{Y}$ and Y'.

Let σ_r', σ_{r+1}' be the rth and $(r+1)$th largest singular values of Y', respectively. By Wedin's Thereom (see Section 2.3 of [9]), if $\sigma_r' - \hat{\sigma}_{r+1} > 0$, then we have

$$\|\sin\Theta(\hat{U}_r, U_r)\|_{\mathrm{F}} \leq \frac{\left\|\hat{Y} - Y'\right\|_{\mathrm{F}}}{\sigma_r' - \hat{\sigma}_{r+1}} = \frac{\left\|(I - U_r U_r^T) y_n\right\|}{\sigma_r' - \hat{\sigma}_{r+1}}. \tag{73}$$

Regarding the values of σ_r' and σ_{r+1}', first we have $\sigma_r' \geq \sigma_r$. This is because

$$\sigma_r' = \inf_{x \in \text{span}(U_r)} \left\|x^T Y'\right\| = \inf_{x \in \text{span}(U_r)} \left\|\left(x^T Y, x^T y_n\right)\right\| \geq \inf_{x \in \text{span}(U_r)} \left\|x^T Y\right\| \geq \sigma_r.$$

In addition, we have $\sigma_{r+1}' = \sigma_{r+1}$, due to the fact that $(I - U_r U_r^T) Y' = ((I - U_r U_r^T) Y, 0)$. By Weyl's inequality, we have

$$\left|\hat{\sigma}_{r+1} - \sigma_{r+1}'\right| \leq \left\|Y - Y'\right\| = \left\|(I - U_r U_r^T) y_n\right\|.$$

Hence, if $\sigma_r - \sigma_{r+1} > 2\left\|(I - U_r U_r^T) y_n\right\|$ is further assumed, we have

$$\sigma_r' - \hat{\sigma}_{r+1} \geq \sigma_r - \sigma_{r+1} - \left\|(I - U_r U_r^T) y_n\right\| \geq \frac{1}{2}\left(\sigma_r - \sigma_{r+1}\right). \tag{74}$$

With (73), (74), and the fact $\|\hat{U}_r \hat{U}_r^T - U_r U_r^T\|_{\mathrm{F}} = \sqrt{2}\|\sin\Theta(\hat{U}_r, U_r)\|_{\mathrm{F}}$ (see Lemma 1 of [9]), the proof is complete. □

LEMMA E.1. *Let $E = (\epsilon_1, \ldots, \epsilon_n) \in \mathbb{R}^{p \times n}$ be a random matrix with each column $\epsilon_i \sim SG_p(\sigma^2), \forall i \in [n]$ independently. Then*

$$\mathbb{P}\left(\|E\| \geq 4t\sigma(\sqrt{n} + \sqrt{p})\right) \leq \exp\left(-\frac{(t^2 - 3)n}{2}\right),$$

for any $t \geq 2$.

PROOF. We follow a standard ϵ-net argument. Let $\mathcal{U}$ and $\mathcal{V}$ be a 1/4 covering set of the unit sphere in $\mathbb{R}^p$ and in $\mathbb{R}^n$, respectively. That is, for any $u \in \mathbb{R}^p$ such that $\|u\| = 1$, there exists a $u' \in \mathcal{U}$ such that $\|u'\| = 1$ and $\|u - u'\| \leq 1/4$. Similarly, for any $v \in \mathbb{R}^n$ such that $\|v\| = 1$, there exists a $v' \in \mathcal{V}$ such that $\|v'\| = 1$ and $\|v - v'\| \leq 1/4$. Then

$$\begin{aligned}\left|u^T E v\right| &= \left|u^{'T} E v' + u^{'T} E(v - v') + (u - u')^T E v' + (u - u')^T E(v - v')\right| \\ &\leq \left|u^{'T} E v'\right| + \left|u^{'T} E(v - v')\right| + \left|(u - u')^T E v'\right| + \left|(u - u')^T E(v - v')\right|.\end{aligned}$$

Maximizing over u, v on both sides, we have

$$\|E\| = \max_{u \in \mathbb{R}^p, v \in \mathbb{R}^n : \|u\| = \|v\| = 1} \left|u^T E v\right| \leq \max_{u' \in \mathcal{U}, v' \in \mathcal{V}} \left|u^{'T} E v'\right| + \frac{1}{4}\|E\| + \frac{1}{4}\|E\| + \frac{1}{16}\|E\|.$$

Hence,

$$\|E\| \leq 4 \max_{u' \in \mathcal{U}, v' \in \mathcal{V}} \left|u^{'T} E v'\right|.$$

For any $u' \in \mathcal{U}, v' \in \mathcal{V}$, we have each $u'^T \epsilon_i$ being an independent SG(σ^2) and then $u'^T E v' \sim$ SG(σ^2). Note $|U| \leq 9^p \leq e^{3p}$ and similarly $|V| \leq e^{3n}$. Then by the tail probability of sub-Gaussian random variable and by the union bound, we have

$$\begin{aligned}\mathbb{P}\left(\|E\| \leq 4t\sigma(\sqrt{n} + \sqrt{p})\right) &\leq \mathbb{P}\left(\max_{u' \in \mathcal{U}, v' \in \mathcal{V}} \left|u'^T E v'\right| \leq t\sigma(\sqrt{n} + \sqrt{p})\right) \\ &\leq |U|\,|V| \exp\left(-\frac{t^2\left(\sqrt{n} + \sqrt{p}\right)^2}{2}\right) \\ &\leq \exp\left(-\frac{(t^2 - 3)n}{2}\right),\end{aligned}$$

for any $t \geq 2$. ☐

LEMMA E.2. *Let $X \sim SG_d(\sigma^2)$. Consider any $k \leq d$. For any matrix $U = (u_1, \ldots, u_k) \in \mathbb{R}^{d \times k}$ that is independent of X and is with orthogonal columns $\{u_i\}_{i \in [k]}$. We have*

$$\mathbb{P}\left(\left\|UU^T X\right\|^2 \geq \sigma^2(k + 2\sqrt{kt} + 2t)\right) \leq e^{-t}.$$

PROOF. Note that $\mathrm{tr}(UU^T) = \mathrm{tr}((UU^T)^2) = k$ and $\left\|UU^T\right\| = 1$. This is a direct consequence of Theorem 1 in [18] for concentration of quadratic forms of sub-Gaussian random vectors. ☐

PROOF OF PROPOSITION 3.1. Define $\hat{P} = \sum_{i \in [r]} \hat{\lambda}_i \hat{u}_i \hat{v}_i^T$. Due to the fact that $\hat{P}$ is the best rank-r approximation of X in spectral norm and P is rank-κ, under the assumption that $\kappa \leq r$, we have that

$$\left\|\hat{P} - X\right\| \leq \|P - X\| = \|E\|.$$

Since $r \leq k$ is assumed, the rank of $\hat{P} - P$ his at most $2k$, and we have

$$\left\|\hat{P} - P\right\|_{\mathrm{F}} \leq \sqrt{2k}\left\|\hat{P} - P\right\| \leq \sqrt{2k}\left(\left\|\hat{P} - X\right\| + \|P - X\|\right) \leq 2\sqrt{2k}\,\|E\| \tag{75}$$

Now, denote $\hat{\Theta} := (\hat{\theta}_{\hat{z}_1}, \hat{\theta}_{\hat{z}_2}, \ldots, \hat{\theta}_{\hat{z}_n})$. Since $\hat{\Theta}$ is the solution to the k-means objective (15), we have that

$$\left\|\hat{\Theta} - \hat{P}\right\|_{\mathrm{F}} \leq \left\|P - \hat{P}\right\|_{\mathrm{F}}.$$

Hence, by the triangle inequality, we obtain that

$$\left\|\hat{\Theta} - P\right\|_{\mathrm{F}} \leq 2\left\|\hat{P} - P\right\|_{\mathrm{F}} \leq 4\sqrt{2k}\,\|E\|.$$

Now, define the set S as

$$S = \left\{i \in [n] : \left\|\hat{\theta}_{\hat{z}_i} - \theta^*_{z_i^*}\right\| > \frac{\Delta}{2}\right\}.$$

Since $\left\{\hat{\theta}_{\hat{z}_i} - \theta^*_{z_i^*}\right\}_{i\in[n]}$ are exactly the columns of $\hat{\Theta} - P$, we have that

$$|S| \leq \frac{\left\|\hat{\Theta} - P\right\|_{\mathrm{F}}^2}{(\Delta/2)^2} \leq \frac{128k\,\|E\|^2}{\Delta^2}.$$

Under the assumption (16) we have

$$\frac{\beta\Delta^2 n}{k^2\,\|E\|^2} \geq 256,$$

which implies

$$|S| \leq \frac{\beta n}{2k}.$$

We now show that all the data points in S^C are correctly clustered. We define

$$C_j = \left\{i \in [n] : z_i^* = j, i \in S^C\right\},\ j \in [k].$$

The following holds:

- For each $j \in [k]$, C_j cannot be empty, as $|C_j| \geq |\{i : z_i^* = j\}| - |S| > 0$.
- For each pair $j, l \in [k], j \neq l$, there cannot exist some $i \in C_j, i' \in C_l$ such that $\hat{z}_i = \hat{z}_{i'}$. Otherwise $\hat{\theta}_{\hat{z}_i} = \hat{\theta}_{\hat{z}_{i'}}$ which would imply

$$\begin{aligned}\left\|\theta^*_j - \theta^*_l\right\| &= \left\|\theta^*_{z_i^*} - \theta^*_{z_{i'}^*}\right\| \\ &\leq \left\|\theta^*_{z_i^*} - \hat{\theta}_{\hat{z}_i}\right\| + \left\|\hat{\theta}_{\hat{z}_i} - \hat{\theta}_{\hat{z}_{i'}}\right\| + \left\|\hat{\theta}_{\hat{z}_{i'}} - \theta^*_{z_{i'}^*}\right\| < \Delta,\end{aligned}$$

  contradicting with the definition of Δ.

Since $\hat{z}_i$ can only take values in $[k]$, we conclude that the sets $\{\hat{z}_i : i \in C_j\}$ are disjoint for all $j \in [k]$. That is, there exists a permutation $\phi \in \Phi$, such that

$$\hat{z}_i = \phi(j),\ i \in C_j,\ j \in [k].$$

This implies that $\sum_{i\in S^C} \mathbb{I}\{\hat{z}_i \neq \phi(z_i^*)\} = 0$. Hence, we obtain that

$$|\{i \in [n] : \hat{z}_i \neq \phi(z_i^*)\}| \leq |S| \leq \frac{128k\,\|E\|^2}{\Delta^2}.$$

Since $|S| \leq \frac{\beta n}{2k}$ (which means $\ell(\hat{z}, z^*) \leq \frac{\beta n}{2k}$ from the above display), for any $\psi \in \Phi$ such that $\psi \neq \phi$, we have $|\{i \in [n] : \hat{z}_i \neq \psi(z_i^*)\}| \geq 2\beta n/k - |S| \geq \beta n/k$. As a result, we have

$$\ell(\hat{z}, z^*) = \frac{1}{n}|\{i \in [n] : \hat{z}_i \neq \phi(z_i^*)\}| \leq \frac{128k\,\|E\|^2}{n\Delta^2}.$$

Moreover, for each $a \in [k]$, we have

$$\left\|\hat{\theta}_{\phi(a)} - \theta_a^*\right\|^2 \leq \frac{\left\|\hat{\Theta} - P\right\|_{\mathrm{F}}^2}{|\{i \in [n] : \hat{z}_i = \phi(a), z_i^* = a\}|} \leq \frac{\left\|\hat{\Theta} - P\right\|_{\mathrm{F}}^2}{\frac{\beta n}{k} - |S|} \leq \frac{64k^2 \|E\|^2}{\beta n}$$

□