Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL FOR "PHASE TRANSITIONS FOR HIGH DIMENSIONAL CLUSTERING AND RELATED PROBLEMS"

By Jiashun Jin*, Zheng Tracy Ke†, and Wanjie Wang‡

*Carnegie Mellon University**, *University of Chicago*† *and University of Pennsylvania*‡

This file contains four sections. Section A discusses an extension of ARW where the nonzero coordinates of μ have both positive and negative signs. Section B-D contain supplementary proofs for the main paper [5]: Section B proves Lemmas 2.1–2.4, Section C proves Lemmas 3.1–3.3, and Section D proves the secondary lemmas that are used in Sections B-C.

APPENDIX A: AN EXTENSION OF THE ARW MODEL

We consider an extension of the Asymptotic Rare and Weak (ARW) model in Section 1.2, where Models (1.1)-(1.2) and the calibration (1.8) continue to hold but (1.7) is replaced by a more sophisticated signal configuration:

$$\mu(j) \stackrel{iid}{\sim} (1-\epsilon)\nu_0 + a\cdot\epsilon\cdot\nu_{-\tau} + (1-a)\cdot\epsilon\cdot\nu_{\tau}, \qquad 1 \le j \le p, \tag{A.1}$$

where $0 \le a \le 1/2$ is a constant. This extended model includes the original ARW as a special case with $a = 0$. In this extension, we allow the nonzero coordinates of the feature vector μ to have positive and negative signs. Due to such a change, we need to slightly modify the definition of the (normalized) Hamming distance for signal recovery: $\mathrm{Hamm}_p(\hat{\mu}, \alpha, \beta, \theta) = (p\epsilon_p)^{-1}\sum_{j=1}^p P(\mathrm{sgn}(\mu(j)) \neq \mathrm{sgn}(\hat{\mu}(j)))$. The loss functions for clustering and hypothesis testing remain the same.

When $0 < a < 1/2$, with high probability, the majority of the nonzero coordinates of μ are positive, and the performance of the four methods in Section 1.1 is not affected. Furthermore, the statistical limits and CTUB for all three problems continue to hold. For brevity, we omit the details.

The case of $a = 1/2$ is more delicate. In this case, the two aggregation methods turn out to be ineffective. In light of this, we introduce a variant of the Sparse Aggregation, where we cluster the n subjects by

$$\hat{\ell}_N^{(sa)} = \mathrm{sgn}(X\hat{\mu}_N^{(sa)}). \tag{A.2}$$

Here,

$$\hat{\mu}_N^{(sa)} = \text{argmax}_{\mu \in \{-1,0,1\}^p : \|\mu\|_0 = N} \|X\mu\|_1. \tag{A.3}$$

Also, we use $\hat{\mu}_N^{(sa)}$ to estimate the sign of μ (i.e., for signal recovery), and use the test statistic

$$\hat{T}_N^{(sa)} = N^{-1/2} \|X \hat{\mu}_N^{(sa)}\|_1 \tag{A.4}$$

for hypothesis testing. Note that if we force $\mu(j) \in \{0, 1\}$ in (A.3), then it reduces to the original Sparse Aggregation.

Remark. We have not found a variant of Simple Aggregation that both achieves the statistical limit and is computationally tractable. However, in the less sparse case, the classical PCA turns out to be already optimal. [1]

We now present the statistical limits and CTUB for all three problems. They are different from the ones we present in the main paper [4]. First, we look at the statistical limits.

$$\eta_\theta^{clu}(\beta) = \begin{cases} (1+\theta-2\beta)/4, & \beta < (1-\theta)/2, \\ \theta/2, & (1-\theta)/2 < \beta < (1-\theta), \\ (1-\beta)/2, & \beta > (1-\theta). \end{cases}$$

$$\eta_\theta^{sig}(\beta) = \begin{cases} \theta/2, & \beta < (1-\theta), \\ (1+\theta-\beta)/4, & \beta > (1-\theta). \end{cases}$$

$$\eta_\theta^{hyp}(\beta) = \begin{cases} (1+\theta-2\beta)/4, & \beta < (1-\theta)/2, \\ \theta/2, & (1-\theta)/2 < \beta < (1-\theta), \\ (1+\theta-\beta)/4, & \beta > (1-\theta). \end{cases}$$

Figure 1 (top left panel) displays the statistical limits for three problems. Comparing it with Figure 2 (top left panel), we find that : (a) the black curve (signal recovery) remains the same, (b) the red curve (clustering) remains the same, except for the segment on the left is replaced by $\tau^4 = p/(ns^2)$, (c) for the blue curve (hypothesis testing), the right most segment remains the same, while the other two segments coincide with those of the red curve.

Achievability. The statistical limit of clustering is achieved by the classical PCA (the left segment) and the variant (A.2) of Sparse Aggregation (the right two segments). For signal recovery, the right two segments are achieved

[1] Classical PCA for hypothesis testing is to reject the null hypothesis when the leading singular value of X is larger than $\sqrt{p}+\sqrt{n}+\log(p)$; for signal recovery is as the description in Section 3. However, for signal recovery, since we need to estimate not only the support but also the sign of μ, we slightly modify it to $\hat{\mu}_*^{(if)}(j) = \text{sgn}(\hat{y}(j)) \cdot 1\{|\hat{y}(j)| > 2\sqrt{\log(p)}\}$, where $\hat{y} = X\hat{\ell}_*^{(if)}$ and $\hat{\ell}_*^{(if)}$ denotes the class label vector estimated by classical PCA.

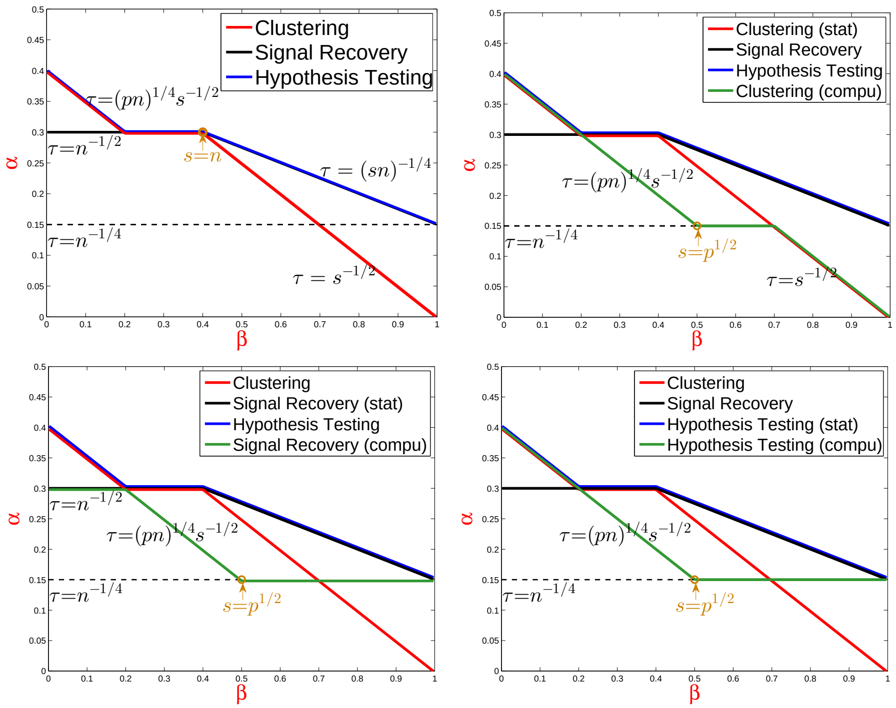

FIG 1. *Top left: statistical limits for clustering (red), signal recovery (black), and hypothesis testing (blue); $s = p\epsilon_p$. Other three panels: CTUB (green) for clustering (top right), signal recovery (bottom left) and hypothesis testing (bottom right), respectively.*

by the modified Sparse Aggregation (A.3), and the left segment is achieved by classical PCA. For hypothesis testing, the left segment is achieved by classical PCA and the right two segments are achieved by the modified Sparse Aggregation (A.4).

Next, we present a CTUB for each of the three problems:

$$\tilde{\eta}_\theta^{clu}(\beta) = \begin{cases} (1+\theta-2\beta)/4, & \beta < 1/2, \\ \theta/4, & 1/2 < \beta < 1-\theta/2, \\ (1-\beta)/2, & \beta > (1-\theta). \end{cases}$$

$$\tilde{\eta}_\theta^{sig}(\beta) = \begin{cases} \theta/2, & \beta < (1-\theta)/2, \\ (1+\theta-2\beta)/4, & (1-\theta)/2 < \beta < 1/2, \\ \theta/4, & \beta > 1/2. \end{cases}$$

$$\tilde{\eta}_\theta^{hyp}(\beta) = \begin{cases} (1+\theta-2\beta)/4, & \beta < 1/2, \\ \theta/4, & \beta > 1/2. \end{cases}$$

See Figure 1 (top right and the two bottom panels).

Methods associated with CTUB. The CTUB for clustering is associated with the methods of classical PCA (left segment) and IF-PCA (right two segments). The CTUB for signal recovery is associated with the methods of classical PCA (left segment) and IF-PCA (right segment). The CTUB for hypothesis testing is associated with the methods of classical PCA (left segment) and IF-PCA (right segment).

Remark. We now make a connection to the recent literature on the Gaussian mixture learning (e.g.[1, 3]). In our framework, we calibrate with (ϵ, τ). In the latter, we calibrate with $\|\mu\|$ and $\|\mu\|_0$. For brevity, we only discuss the problem of hypothesis testing. The statistical limits for hypothesis testing can be (roughly) re-stated as follows:

- $\sqrt{np} \ll s \ll p$: $s\tau^2 = \sqrt{p/n}$.
- $n \ll s \ll \sqrt{np}$: $\tau = n^{-1/2}$.
- $s \ll n$: $\tau = (sn)^{-1/4}$.

Note that the first item corresponds to the *non-sparse* cases in the Gaussian mixture learning literature, where $\|\mu\|^2 = s\tau^2 = \sqrt{p/n}$; the results match with those in, e.g., [1, 3]. The second one is part of the *sparse* case in the Gaussian mixture learning literature, where $1 \ll \|\mu\|^2 = p/n \ll \sqrt{p/n}$ and $n \ll \|\mu\|_0 \ll \sqrt{np}$. The last one is also part of the sparse case, where $\|\mu\|^2 = \sqrt{s/n}$ and $s \ll n$.

APPENDIX B: PROOF OF LEMMAS IN SECTION 2

In this sectoin, we prove the post-selection random matrix theory results in Section 2, specifically Lemmas 2.1–2.4.

B.1. Preliminary lemmas for Section 2.

Lemma B.1 states the well-known Bernstein inequality [6]. Lemma B.2 is a result from classical Random Matrix Theory [7, Page 21]. Lemma B.3 states some properties about columns of the matrix $Z^{(q)}$; it is proved in Section D.1.

Lemma B.1 *Let $X_1, \cdots, X_N$ be independent random variables with $E[X_k] = 0$ and $\mathrm{var}(X_k) \leq v_k$, for $1 \leq k \leq N$. Suppose $E(|X_k|^m) \leq v_k m! c^{m-2}/2$ for all $m \geq 2$, where $c > 0$ is a constant. Then for all $\lambda > 0$,*

$$P\big(\big|\sum_{k=1}^{N} X_k\big| \geq \lambda\sqrt{N}\big) \leq \exp\left(-\frac{\lambda^2/2}{\sum_{k=1}^{n} v_k/N + c\lambda/\sqrt{N}}\right).$$

Lemma B.2 *Let A be an $N \times n$ matrix whose entries are independent standard normal random variables. Then for every $x \geq 0$, with probability at least*

$1 - 2\exp(-x^2/2)$,

$$\sqrt{N} - \sqrt{n} - x \le s_{\min}(A) \le s_{\max}(A) \le \sqrt{N} + \sqrt{n} + x,$$

where $s_{\min}(A)$ and $s_{\max}(A)$ are the respective minimum and maximum singular values of A.

Fix $q > 0$. With $e_1 = (1, 0, \cdots, 0)'$ and $z \sim N(0, I_p)$, we introduce a few notations:

$$\begin{aligned}
\pi_0^{(q)} &= P(\|z\|^2 > n + 2\sqrt{qn\log(p)}), \\
\pi_1^{(q)} &= P(\|z + \sqrt{n}\tau_p^* e_1\|^2 > n + 2\sqrt{qn\log(p)}), \\
a_p^{(q)} &= E\big[(z(1))^2 \cdot 1\{\|z\|^2 > n + 2\sqrt{qn\log(p)}\}\big], \\
b_p^{(q)} &= E\big[(z(2))^2 \cdot 1\{\|z + \sqrt{n}\tau_p^* e_1\|^2 > n + 2\sqrt{qn\log(p)}\}\big], \\
c_p^{(q)} &= E\big[(z(1))^2 \cdot 1\{\|z + \sqrt{n}\tau_p^* e_1\|^2 > n + 2\sqrt{qn\log(p)}\}\big].
\end{aligned}$$

For notation simplicity, we omit all the superscripts. In the following lemma, $m^{(q)}(\ell, \mu)$, $m_*^{(q)}(\ell, \mu)$ and the event D_p are defined in Section 2.

Lemma B.3 *Let $S(\mu)$ denote the support of μ, $\kappa_m = E(|z(1)|^m)$ and $\kappa_{2m}(n) = E(\|z\|^{2m})$, where $z \sim N(0, I_n)$. Below, all the probabilities are conditioning on (ℓ, μ), and the $o(1)$ terms are uniform for all realizations of (ℓ, μ) in the event D_p.*

(a) Fix $j \notin S(\mu)$. For any $v \in \mathcal{S}^{n-1}$ and any integer $m \ge 1$

$$\begin{aligned}
&E[(v' z_j^{(q)})^2] = a_p, \\
&E(|v' z_j^{(q)}|^m) \le \kappa_m \pi_0 (1 + o(1)), \\
&E(\|z_j^{(q)}\|^2) = n a_p, \\
&E(\|z_j^{(q)}\|^{2m}) = \kappa_{2m}(n) \pi_0 (1 + o(1)), \\
&a_p = \pi_0 (1 + L_p n^{-1/2}).
\end{aligned}$$

(b) Fix $j \in S(\mu)$. For any $v \in \mathcal{S}^{n-1}$ and any integer $m \ge 1$

$$\begin{aligned}
&E[(v' z_j^{(q)})^2] = b_p + (c_p - b_p)\tfrac{(v'\ell)^2}{\|\ell\|^2} \\
&E(|v' z_j^{(q)}|^m) \le \kappa_m \pi_1 (1 + o(1)), \\
&E(\|z_j^{(q)}\|^2) = n b_p + (c_p - b_p), \\
&E(\|z_j^{(q)}\|^{2m}) \le 2^m \kappa_{2m}(n) \pi_1 (1 + o(1)), \\
&b_p = \pi_1 (1 + L_p n^{-1/4}), \quad c_p = \pi_1 (1 + L_p n^{-1/4}).
\end{aligned}$$

(c) $m^{(q)}(\ell,\mu) = (p-|S(\mu)|)\pi_0 + |S(\mu)|\pi_1,$
$m_*^{(q)}(\ell,\mu) = (p-|S(\mu)|)a_p + |S(\mu)|[b_p + n^{-1}(c_p - b_p)].$

Remark. Lemma B.3 allows us to characterize the quantities $m^{(q)}$ and $m_*^{(q)}$. First, by (a)-(b), $a_p \sim \pi_0$ and $b_p \sim c_p \sim \pi_1$. Combining them with (c) gives that $m_*^{(q)} \sim m^{(q)}$. Second, we look at $m^{(q)}$. By Lemma D.1 and Mills' ratio [6], $\pi_0 \sim \bar{\Phi}(\sqrt{2q\log(p)}) = L_p p^{-q}$. Similarly, $\pi_1 \sim L_p p^{-[(\sqrt{r}-\sqrt{q})_+]^2}$. Plugging them into (c) gives

$$m^{(q)} \sim L_p p^{1-q} + p\epsilon_p \cdot L_p p^{-[(\sqrt{r}-\sqrt{q})_+]^2}.$$

This is the equation (2.1) in the main text of [5].

B.2. Proof of Lemma 2.1. Fix $q > 0$ and write $H_0 = Z^{(q)}(Z^{(q)})'$ and $\hat{S} = \hat{S}_q^{(if)}$ for short. Fix a realization (ℓ,μ). With probability at least $1 - O(p^{-3})$,

$$||\hat{S}| - m^{(q)}| \leq \sqrt{6m^{(q)}\log(p)}. \tag{B.5}$$

First, we consider $q > \tilde{q}(\beta,\theta,r)$, so that $m^{(q)} \leq np^{-\delta}$ for some $\delta > 0$. Let $k = \lceil m^{(q)} + \sqrt{6m^{(q)}\log(p)}\rceil$. Under (B.5),

$$\lambda_{\max}(H_0) \leq \max_{T\subset\{1,\cdots,p\},|T|\leq k} \lambda_{\max}((ZZ')^{T,T}),$$
$$\lambda_{\min}^+(H_0) \geq \min_{T\subset\{1,\cdots,p\},|T|\leq k} \lambda_{\min}^+((ZZ')^{T,T}),$$

where for a matrix A, $\lambda_{\min}^+(A)$ denotes the minimum non-zero eigenvalue and $A^{T,T}$ is the submatrix restricted to rows and columns in T. For each fixed T, we can write $(Z'Z)^{T,T} = Z_T(Z_T)'$, where $Z_T = (z_j, j\in T)$ is an $n\times|T|$ matrix with *iid* entries of $N(0,1)$. Using Lemma B.2, for each T, with probability at least $1-O(p^{-(k+3)})$, all non-zero eigenvalues of $(ZZ')^{T,T}$ fall into

$$\Big[\big(\sqrt{n}+\sqrt{|T|}+\sqrt{6k\log(p)}\big)^2,\ \big(\sqrt{n}-\sqrt{|T|}-\sqrt{6k\log(p)}\big)^2\Big] = n \pm C\sqrt{nk\log(p)}.$$

Note that the number of subsets T such that $|T|\leq k$ is no more than p^k. Combining the above results, we find that with probability at least $1 - O(p^{-3})$, all non-zero eigenvalues of H_0 fall into

$$n \pm C\sqrt{nm^{(q)}\log(p)}. \tag{B.6}$$

The claim then follows.

Next, we consider $q < \tilde{q}(\beta,\theta,r)$, so that $m^{(q)} \geq np^{\delta}$ for some $\delta > 0$. Write for short

$$\omega_p = \sqrt{nm^{(q)}} + o(1)|S(\mu)|\pi_1,$$

where π_1 is as in Lemma B.3 and $m_1^{(q)} = |S(\mu)|\pi_1$ by definition. It suffices to show that with probability at least $1 - O(p^{-3})$,

$$\|H_0 - m_*^{(q)} I_n\| \leq C\omega_p. \tag{B.7}$$

We now show (B.7). Fix $\alpha > 0$. A subset $\mathcal{M}_\alpha$ of the unit sphere $\mathcal{S}^{n-1}$ is called an α-net if for any $v \in \mathcal{S}^{n-1}$, there exits $u \in \mathcal{M}_\alpha$ such that $\|u - v\| \leq \alpha$. The following lemma states some well-known results and its proof can be found in [7, Page 8].

Lemma B.4 *Fix $\alpha \in (0, 1/2)$. For any $\mathcal{M}_\alpha$, an α-net of $\mathcal{S}^{n-1}$, and any symmetric matrix $A \in R^{n,n}$, $\|A\| \leq (1-2\alpha)^{-1} \sup_{u \in \mathcal{M}_\alpha}\{|u'Au|\}$. Moreover, there exists an α-net $\mathcal{M}_\alpha^*$ of $\mathcal{S}^{n-1}$ such that $|\mathcal{M}_\alpha^*| \leq (1 + 2/\alpha)^n$.*

By Lemma B.4 with $\alpha = 1/4$, there exists a subset $\mathcal{M}^*$, such that $|\mathcal{M}^*| \leq 9^n$ and $\sup_{v \in \mathcal{M}^*} v'Av \geq \|A\|/2$ for any $n \times n$ matrix A. Therefore, to show the claim, it suffices to show that for each fixed $v \in \mathcal{M}^*$, with probability $\geq 1 - O(9^{-n}p^{-3})$,

$$|v'(H_0 - m_*^{(q)} I_n)v| \leq C\omega_p. \tag{B.8}$$

We now show (B.8). Fix v and define

$$W_j = (v'z_j^{(q)})^2 - a_p, \ \text{ for } j \notin S(\mu); \qquad W_j = (v'z_j^{(q)})^2 - b_p, \ \text{ for } j \in S(\mu),$$

where a_p and b_p are defined in Section B.1. By (c) of Lemma B.3, $m_*^{(q)} = (p - |S(\mu)|)a_p + |S(\mu)|b_p + n^{-1}(c_p - b_p)|S(\mu)|$. Since $|c_p - b_p| = o(\pi_1)$, we can rewrite

$$v'(H_0 - m_*^{(q)} I_n)v = \sum_{j=1}^{p} W_j + o(n^{-1}|S(\mu)|\pi_1). \tag{B.9}$$

Here W_j's are independent of each other. Applying Lemma B.3, we get the following results. For $j \notin S(\mu)$, $E(W_j) = 0$, $\mathrm{var}(W_j) \leq 3\pi_0(1 + o(1))$ and $E(|W_j|^m) \leq \kappa_{2m}\pi_0(1 + o(1))$. For $j \in S(\mu)$, $|E(W_j)| \leq |b_p - c_p| = \pi_1 \cdot o(1)$, $\mathrm{var}(W_j) \leq 3\pi_1(1 + o(1))$ and $E(|W_j|^m) \leq \kappa_{2m}\pi_1(1 + o(1))$. So we have

$$|\sum_{j=1}^{p} E(W_j)| = o(1)|S(\mu)|\pi_1, \qquad \sum_{j=1}^{p} \mathrm{var}(W_j) \lesssim 3m^{(q)}.$$

We apply Lemma B.1 with $\lambda = \sqrt{9p^{-1}m^{(q)}(n\log(9) + 2\log(p) + \log(2))}$. To check the moment conditions, we note that $\kappa_{2m} = E_{Y\sim N(0,1)}(|Y|^{2m}) \leq 2^m m!$ for all $m \geq 1$. Furthermore, since $m^{(q)}/n \to \infty$, we have $\sum_j \mathrm{var}(W_j)/p \sim 3m^{(q)}/p \gg \lambda/\sqrt{p}$. It follows that with probability $\geq 1 - O(9^{-n}p^{-3})$,

$$|\sum_{j=1}^{p} W_j| \lesssim 3\sqrt{\log(9)}\sqrt{nm^{(q)}} + o(1)|S(\mu)|\pi_1.$$

This gives (B.8), and the proof is now complete. □

B.3. Proof of Lemma 2.2. We have shown the first claim in (B.7), noting that $\omega_p \sim \sqrt{nm^{(q)}}$ when $r < \rho^*_\theta(\beta)$ (see also (B.15)).

We now show the second claim. Write for short $H_0 = Z^{(q)}(Z^q)'$. The key is the following lemma, which is proved in Section D.

Lemma B.5 *Under conditions of Lemma 2.2, as $p \to \infty$, conditioning on any realization of (ℓ, μ) on the event D_p, with probability at least $1 - O(n^{-2})$,*

$$|n^{-1}\mathrm{tr}(H_0) - m_*^{(q)}| \leq C\sqrt{m^{(q)}\log(p)},$$
$$\|H_0\|_F^2 \geq n^{-1}[\mathrm{tr}(H_0)]^2 + Cn^2 m^{(q)}.$$

Let k be the largest integer that is no larger than $m^{(q)}/2$. Since $k \gg n$, for each fixed $k \times k$ submatrix of ZZ', its rank is n with probability 1. Using (B.5), the rank of H_0 is n with probability at least $1 - O(p^{-3})$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ be the eigenvalues of H_0 and write $\bar{\lambda} = n^{-1}\sum_{i=1}^n \lambda_i$. For $\delta_p \equiv \sqrt{nm^{(q)}}$, (B.7) and Lemma B.5 imply

$$|\lambda_1 - \lambda_n| \leq A_1\delta_p, \qquad \bar{\lambda} = m_*^{(q)} + o(\delta_p), \qquad \sum_{i=1}^n \lambda_i^2 \geq n\bar{\lambda}^2 + A_2 n\delta_p^2, \tag{B.10}$$

for some constants $A_1, A_2 > 0$. On one hand,

$$\sum_{i=1}^n (\lambda_i - \bar{\lambda})^2 \geq A_2 n\delta_p^2.$$

On the other hand, $\lambda_i - \bar{\lambda} \leq \lambda_1 - \bar{\lambda}$ for i satisfying $\lambda_i \geq \bar{\lambda}$; moreover, $\bar{\lambda} - \lambda_i \leq A_1\delta_p$ for i such that $\lambda_i < \bar{\lambda}$. It follows that

$$\begin{aligned}\sum_{i=1}^n (\lambda_i - \bar{\lambda})^2 &\leq (\lambda_1 - \bar{\lambda})\sum_{i:\lambda_i \geq \bar{\lambda}}(\lambda_i - \bar{\lambda}) + A_1\delta_p \sum_{i:\lambda_i < \bar{\lambda}}(\bar{\lambda} - \lambda_i) \\ &= [(\lambda_1 - \bar{\lambda}) + A_1\delta_p]\sum_{i:\lambda_i \geq \bar{\lambda}}(\lambda_i - \bar{\lambda}) \\ &\leq n[(\lambda_1 - \bar{\lambda}) + A_1\delta_p](\lambda_1 - \bar{\lambda}).\end{aligned}$$

Now, if we write $x = \lambda_1 - \bar{\lambda}$, then $x(x + A_1\delta_p) \geq A_2\delta_p^2$. It follows that

$$\lambda_1 - \bar{\lambda} \geq \frac{\sqrt{A_1^2 + 4A_2} - A_1}{2}\delta_p. \tag{B.11}$$

Combining it with the second equation in (B.10), we obtain that $\lambda_1 \geq m_*^{(q)} + C\sqrt{nm^{(q)}}$. ☐

B.4. Proof of Lemma 2.3. Write for short $A = \ell(Z^{(q)}\mu^{(q)})' + (Z^{(q)}\mu^{(q)})\ell'$. Since

$$\|A\| \leq 2\|\ell\|\|Z^{(q)}\mu^{(q)}\| \leq 2n\|Z^{(q)}\mu^{(q)}\|_\infty,$$

it suffices to show that with probability $1 - O(p^{-3})$,

$$\|Z^{(q)}\mu^{(q)}\|_\infty \leq C\tau_p^*\sqrt{m_1^{(q)}}. \tag{B.12}$$

Note that

$$\|Z^{(q)}\mu^{(q)}\|_\infty = \max_{1\leq i\leq n} |\sum_{j\in S(\mu)} \mu(j)z_j^{(q)}(i)| = \tau_p^* \max_{1\leq i\leq n} |\sum_{j\in S(\mu)} z_j^{(q)}(i)|. \tag{B.13}$$

Fix i and write $V_j = z_j^{(q)}(i)$ for short. Then V_j's are independent and $E(V_j) = 0$ by symmetry. We apply Lemma B.3 with $v = e_1$ and find that $\mathrm{var}(V_j) \leq b_p + 2|c_p - b_p| = \pi_1(1 + o(1))$. By Lemma B.1 (the moment conditions can be verified using Lemma B.3), $|\sum_{j\in S(\mu)} V_j| \leq 2\sqrt{|S(\mu)|\pi_1}$ with probability $1 - O(p^{-4})$. It follows that with probability $1 - O(p^{-3})$,

$$\max_{1\leq i\leq n} |\sum_{j\in S(\mu)} z_j^{(q)}(i)| \leq C\sqrt{|S(\mu)|\pi_1} = C\sqrt{m_1^{(q)}}. \tag{B.14}$$

Combining (B.13)-(B.14) gives (B.12). ☐

B.5. Proof of Lemma 2.4. Introduce

$$\Delta^\dagger(q, \beta, r, \theta) = \begin{cases} \beta - \frac{1}{2}\min\{q, \beta - \frac{\theta}{2}\}, & q \leq r, \\ \beta + (\sqrt{q} - \sqrt{r})^2 - \frac{1}{2}\min\{q, \beta - \frac{\theta}{2} + (\sqrt{q} - \sqrt{r})^2\}, & q > r. \end{cases}$$

By elementary algebra,

$$r < \rho_\theta^*(\beta) \qquad \Longleftrightarrow \qquad \min_{q>0} \Delta^\dagger(q, \beta, r, \theta) > 1/2,$$

Moreover, using Mills' ratio,

$$\min\left\{\frac{n(\tau_p^*)^2|S(\mu)|\pi_1}{\sqrt{nm^{(q)}}},\ \ \tau_p^*\sqrt{|S(\mu)|\pi_1}\right\} = L_p p^{1/2-\Delta^\dagger(q,\beta,r,\theta)}.$$

It follows that for some $\delta > 0$,

$$r < \rho_\theta^*(\beta) \qquad \Longleftrightarrow \qquad |S(\mu)|\pi_1 \leq p^{-\delta}\max\{\sqrt{m^{(q)}}, \sqrt{n}\}. \tag{B.15}$$

Consider the first claim. The proof is similar to that of (B.8), except that we take $\lambda = C\sqrt{p^{-1}m^{(q)}\log(p)}$ when applying Lemma B.1. It follows that for any $v \in \mathcal{S}^{n-1}$, with probability at least $1 - O(p^{-3})$,

$$|v'(H_0 - m_*^{(q)}I_n)v| \leq C\sqrt{m^{(q)}\log(p)} + o(|S(\mu)|\pi_1).$$

By (B.15), the second term above is negligible and the claim follows.

Consider the second claim. $H - H_0 = \|\mu^{(q)}\|^2\ell\ell' + A$, where with probability at least $1 - O(p^{-3})$, $\|A\| \leq Cn\tau_p^*\sqrt{m_1^{(q)}}$ by Lemma 2.3. Furthermore, by elementary statistics, $\|\mu^{(q)}\|^2 \leq Cm_1^{(q)}(\tau_p^*)^2$. Note that $m_1^{(q)} = |S(\mu)|\pi_1$ due to the spherical symmetry of $N(0, I_n)$. Together, we see that

$$\|H - H_0\| \leq L_p(\sqrt{n}|S(\mu)|\pi_1 + n^{3/4}\sqrt{|S(\mu)|\pi_1}).$$

If $q > \tilde{q}(\beta, r, \theta)$, then $m^{(q)} = o(n)$ and (B.15) implies $\|H - H_0\| \leq p^{-\delta}n$. If $q < \tilde{q}(\beta, r, \theta)$, then $n = o(m^{(q)})$ and (B.15) implies $\|H - H_0\| \leq p^{-\delta}\sqrt{nm^{(q)}}$. □

APPENDIX C: PROOF OF LEMMAS IN SECTION 3

In this section, we prove Lemmas 3.1–3.3.

C.1. Proof of Lemma 3.1. We first show the claim for $B = I_p$ and then generalize it to any B satisfying $\max\{\|B\|, \|B^{-1}\|\} \leq L_p$.

Fix $B = I_p$. We use $\delta > 0$ to denote a generic constant which only depends on (α, β, θ) but may change from occurrence to occurrence. In our model, $X = \ell\mu' + Z$. Let $H_0 = ZZ' - pI_n$. It is seen

$$XX' - pI_n = [\|\mu\|^2\ell\ell' + \ell\mu'Z' + Z\mu\ell'] + ZZ' - pI_n = [\|\mu\|^2\ell\ell' + \ell\mu'Z' + Z\mu\ell'] + H_0. \tag{C.16}$$

Since ξ is a left singular vector of X, $\lambda\xi = [\|\mu\|^2(\xi,\ell) + (\xi, Z\mu)]\ell + (\xi,\ell)Z\mu + H_0\xi$. Rearranging it, we have

$$\sqrt{n}\xi = (I_n - (1/\lambda)H_0)^{-1}[b_1\ell + b_2 Z(\mu/\|\mu\|)], \tag{C.17}$$

where $b_1 = b_1(\ell, Z, \mu) = (1/\lambda)\cdot[\sqrt{n}\|\mu\|^2(\xi,\ell)+\sqrt{n}(\xi, Z\mu)]$ and $b_2 = b_2(\ell, Z, \mu) = (1/\lambda)\sqrt{n}\|\mu\|(\xi,\ell)$. Therefore, $\min\{\|\sqrt{n}\xi - \ell\|_\infty, \|\sqrt{n}\xi + \ell\|_\infty\}$ is no greater than

(C.18)
$$\min\{|b_1-1|, |b_1+1|\}+|b_1|\|\ell-(I_n-(1/\lambda)H_0)^{-1}\ell\|_\infty+|b_2|\|(I_n-(1/\lambda)H_0)^{-1}Z(\mu/\|\mu\|)\|_\infty.$$

To show the claim, it is sufficient to show that with probability at least $1 - o(p^{-3})$,

$$\min\{|b_1 - 1|, |b_1 + 1|\} \le p^{-\delta}, \qquad |b_2| \le p^{-\delta}, \tag{C.19}$$

and

(C.20)
$$\|\ell-(I_n-(1/\lambda)H_0)^{-1}\ell\|_\infty \le p^{-\delta}, \qquad \|(I_n-\frac{1}{\lambda}H_0)^{-1}Z(\mu/\|\mu\|)\|_\infty \le C\sqrt{\log(p)}.$$

We now show (C.19). Consider the first item. Since Z and μ are independent, we have that with probability at least $1 - o(p^{-3})$, $|(\xi, Z\mu)| \le \|\mu\| \cdot \|Z(\mu/\|\mu\|)\| \le 2\|\mu\|\sqrt{n}$. Combining this with the triangle inequality,

$$\begin{aligned}
&\min\{b_1 - 1, b_1 + 1\} \\
\le&(n\|\mu\|^2/\lambda)|\cos(\ell,\xi) - 1| + |1 - (n\|\mu\|^2/\lambda)| + (\sqrt{n}/\lambda)|(\xi, Z\mu)| \\
\le&(n\|\mu\|^2/\lambda)|\cos(\ell,\xi) - 1| + |1 - (n\|\mu\|^2/\lambda)| + 2n\|\mu\|/\lambda. \qquad \text{(C.21)}
\end{aligned}$$

At the same time, we rewrite (C.16) as

$$XX' - pI_n = A + H_0, \text{ where } A = \|\mu\|^2\ell\ell' + \ell\mu'Z' + Z\mu\ell' \text{ for short.} \tag{C.22}$$

Note that A is a symmetric matrix of rank 2. For short, write $\nu = \|\mu\|^{-2}\mu$ and $a = a(\ell,\mu,Z) = (1+4n^{-1}[\ell'Z\nu+\|Z\nu\|^2])^{1/2}$. Let $\lambda_\pm$ be the two nonzero eigenvalues of A, and let $\eta_\pm$ be the corresponding eigenvectors. By elementary algebra,

$$\lambda_\pm(A) = n\|\mu\|^2[(1/2)(1\pm a)+n^{-1}\ell'Z\nu], \qquad \eta_\pm \propto (1/2)(1\pm a)\ell+Z\nu. \tag{C.23}$$

By elementary statistics, it is seen that with probability at least $1 - o(p^{-3})$ that $n^{-1}[|\ell'Z\nu| + \|Z\nu\|^2]$ does not exceed

(C.24)
$$C\sqrt{\log(p)}n^{-1}[(\sqrt{n}\|\mu\|^{-1}) + n\|\mu\|^{-2}] = C\sqrt{\log(p)}[(\sqrt{n}\|\mu\|)^{-1} + \|\mu\|^{-2}].$$

Note that for (α,β,θ) in our range of interest, $p\epsilon_p\tau_p^2 \ge \sqrt{p/n} = p^{(1-\theta)/2}$. By the way μ is generated, $\|\mu\|^2 \sim p\epsilon_p\tau_p^2$. Therefore, with probability at least $1 - o(p^{-3})$,

$$\|\mu\|^2 \sim p\epsilon_p\tau_p^2 \ge p^\delta\sqrt{p/n}, \qquad n\|\mu\|^2 \sim np\epsilon_p\tau_p^2 \ge p^\delta\sqrt{pn}. \tag{C.25}$$

Inserting (C.25) into (C.24) gives that with probability at least $1 - o(p^{-3})$, $|a-1| \leq Cp^{-\delta}$. Combining this with (C.23),

(C.26)
$$|(n\|\mu\|^2/\lambda_+) - 1| \leq p^{-\delta}, \qquad (\lambda_-/\lambda_+) \leq p^{-\delta}, \qquad |\cos(\ell, \eta_+) - 1| \leq p^{-\delta}.$$

At the same time, by a direct use of the elementary Random Matrix Theory [7], $\|H_0\| = \|ZZ' - pI_p\| \leq C\sqrt{pn}$. Combining these with (C.25)-(C.26) gives

$$\|(1/\lambda_+)H_0\| \leq C\sqrt{pn}/(n\|\mu\|^2) \leq Cp^{-\delta}. \tag{C.27}$$

This says that in (C.22), the leading eigenvalue of A is larger than that of H_0 by p^{δ} times. By matrix perturbation theory, we have that with probability at least $1 - o(p^{-3})$,

$$|\lambda_+/\lambda - 1| \leq p^{-\delta}, \qquad |\cos(\eta_+, \xi) - 1| \leq p^{-\delta}. \tag{C.28}$$

Combining (C.26) and (C.28) gives

$$|(n\|\mu\|^2/\lambda) - 1| \leq p^{-\delta}, \qquad |\cos(\ell, \xi) - 1| \leq p^{-\delta}. \tag{C.29}$$

In particular, combining (C.25), (C.27), and (C.28) gives that with probability at least $1 - o(p^{-3})$,

$$\|(1/\lambda)H_0\| \leq Cp^{-\delta}, \qquad \sqrt{pn}/\lambda \leq p^{-\delta}. \tag{C.30}$$

Inserting (C.29) into (C.21) gives the first item of (C.19).

Consider the second item of (C.19). Note that $|b_2| \leq (n\|\mu\|/\lambda)$, where by (C.29), the right hand side $\leq \|\mu\|^{-1}$. The claim follows directly from (C.25).

We now show (C.20). Since the proofs are similar, we only show the first item. Let e_1 be the first base vector of R^n. Note that by symmetry and by using the union bound, it is sufficient to show that with probability at least $1 - o(p^{-4})$,

(C.31)
$$\left|e_1'(I_n - \frac{1}{\lambda}H_0)^{-1}e_1 - 1\right| \leq p^{-\delta}, \qquad |e_1'(I_n - \frac{1}{\lambda}H_0)^{-1}(\ell - \ell_1 e_1)| \leq C\sqrt{\log(p)}.$$

The first claim follows easily by (C.30) and basic algebra. For the second claim, write $\ell = (\ell_1, \tilde{\ell})'$, and let $\tilde{Z}$ be the $(n-1) \times p$ matrix consisting all but the first row of Z, and let $\tilde{H}_0 = \tilde{Z}\tilde{Z}' - pI_{n-1}$. It follows that

$$I_n - (1/\lambda)H_0 = \begin{pmatrix} 1 - (1/\lambda)[\|Z_1\|^2 - p], & -(1/\lambda)Z_1'\tilde{Z} \\ -(1/\lambda)\tilde{Z}Z_1, & I_{n-1} - (1/\lambda)\tilde{H}_0 \end{pmatrix},$$

and
(C.32)
$$e_1'(I_n - \frac{1}{\lambda}H_0)^{-1}(\ell - \ell_1 e_1) = (e_1'[I_n - (1/\lambda)H_0]^{-1}e_1)\cdot(1/\lambda)Z_1'\tilde{Z}'[I_{n-1} - (1/\lambda)\tilde{H}_0]^{-1}\tilde{\ell}.$$

Now, since rows of Z are independent, Z_1 and $\tilde{Z}[I_{n-1} - (1/\lambda)\tilde{H}_0]^{-1}\tilde{\ell}$ are two vectors that almost independent of each other; the only issue is that Z_1 is correlated with λ. To overcome the difficulty, we write
(C.33)
$$\frac{Z_1'\tilde{Z}'[I_{n-1} - (1/\lambda)\tilde{H}_0]^{-1}\tilde{\ell}}{\lambda} = \sum_{k=0}^{\infty} \frac{Z_1'\tilde{Z}\tilde{H}_0^k\tilde{\ell}}{\lambda^{k+1}} = \sum_{k=0}^{\infty} \frac{\|\tilde{Z}\tilde{H}_0^k\tilde{\ell}\|}{\lambda^{k+1}} \cdot \frac{Z_1'\tilde{Z}\tilde{H}_0^k\tilde{\ell}}{\|\tilde{Z}\tilde{H}_0^k\tilde{\ell}\|}.$$

Now, for each k, Z_1 and $\tilde{Z}\tilde{H}_0^k\tilde{\ell}$ are independent, and so

$$Z_1'(\tilde{Z}'\tilde{H}_0^k\tilde{\ell}/\|\tilde{Z}'\tilde{H}_0^k\tilde{\ell}\|) \sim N(0,1).$$

For k-th term, with probability $1 - o(p^{-4(k+1)})$, there is $|Z_1'(\tilde{Z}'\tilde{H}_0^k\tilde{\ell}/\|\tilde{Z}'\tilde{H}_0^k\tilde{\ell}\|)| \leq \sqrt{8(k+1)\log(p)}$. Additionally, by basics in RMT [7], with probability at least $1 - o(p^{-4})$, $\|\tilde{Z}\tilde{H}_0^k\| \leq \sqrt{p}(C\sqrt{np})^k$ for all k.

$$(1/\lambda)^{k+1}\|\tilde{Z}\tilde{H}_0^k\tilde{\ell}\| \leq \sqrt{(n-1)}(1/\lambda)^k\|\tilde{Z}\tilde{H}_0^k\| \leq (C\sqrt{np}/\lambda)^{k+1}.$$

Combining these with (C.33) and the second term of (C.30), it is seen that with probability at least $1 - o(p^{-4})$,

$$|(1/\lambda)Z_1'\tilde{Z}'[I_{n-1} - (1/\lambda)\tilde{H}_0]^{-1}\tilde{\ell}| \leq p^{-\delta}.$$

Inserting this into (C.32) and using the first item of (C.31), the second item of (C.31) follows.

For a general B, the proof is similar by noting that $\|ZB\| \leq L_p\|Z\|$ and the following lemma, which is proved below.

Lemma C.1 *As $n, p \to \infty$ and $p/n \to \infty$, for an $n \times p$ random matrix Z where $Z(i,j) \overset{iid}{\sim} N(0,1)$ and any non-random matrix $B \in R^{p,p}$ such that $\max\{\|B\|, \|B^{-1}\|\} \leq L_p$, with probability $1 - O(p^3)$, $\|ZBB'Z' - \mathrm{tr}(BB')I_n\| \leq C\sqrt{np}$.*

C.2. Proof of Lemma 3.2. Letting Φ be the CDF of $N(0,1)$, denote the mean and variance of $|z_i + h|$ by $u(h)$ and $\sigma^2(h)$, respectively. It is seen that

$$u(h) = \sqrt{2/\pi}e^{-h^2/2} + h[1 - 2\Phi(-h)], \qquad \sigma^2(h) = 1 + h^2 - \mu^2(h). \tag{C.34}$$

By Jensen's inequality, $E|z_i + h| \geq |E(z_i + h)| = h$. It follows that

$$u(h) \geq h, \qquad \sigma^2(h) \leq 1, \tag{C.35}$$

At the same time, we claim that as $n \to \infty$, for any $0 \leq x \leq \sqrt{n}/\log(n)$,

$$P\big(\big|\sum_{i=1}^{n}\big(|z_i + h| - u(h)\big)\big| \geq \sqrt{n}x\big) \leq 2\exp\big(-(1 + o(1))\frac{x^2}{2\sigma^2(h)}\big), \tag{C.36}$$

where $o(1) \to 0$ as $n \to \infty$, uniformly for all $h > 0$ and $0 < x \leq \sqrt{n}/\log(n)$. Combining (C.35) and (C.36) gives Lemma 3.2.

We now show (C.36). Write for short $Y_i = |z_i + h|$. It is sufficient to show that

$$P\big(\sum_{i=1}^{n} Y_i \geq nu(h) + \sqrt{n}x\big) \leq \exp\big(-(1 + o(1))\frac{x^2}{2\sigma^2(h)}\big), \tag{C.37}$$

and

$$P\big(\sum_{i=1}^{n} Y_i \leq nu(h) - \sqrt{n}x\big) \leq \exp\big(-(1 + o(1))\frac{x^2}{2\sigma^2(h)}\big). \tag{C.38}$$

Since the proofs are similar, we only show (C.37). By elementary calculations, the moment generating function of Y_i is

$$M_Y(s) = E[e^{sY}] = e^{s^2/2}[e^{hs}\Phi(s + h) + e^{-hs}\Phi(s - h)], \tag{C.39}$$

By Cramer-Chernoff Theorem ([2]), for any $s > 0$ and any y,

$$P(\sum_{i=1}^{n} Y_i \geq ny) \leq e^{-n(ys - \log M_Y(s))}. \tag{C.40}$$

We now show this (C.37) for the cases of $h < 2\log(\sqrt{n}/x)$ and $h \geq 2\log(\sqrt{n}/x)$ separately.

Consider the case where $h < 2\log(\sqrt{n}/x)$. We wish to use (C.40) with

$$s = \frac{1}{\sigma^2(h)}\frac{x}{\sqrt{n}}, \qquad y = u(h) + x/\sqrt{n}.$$

By our assumptions of $h < 2\log(\sqrt{n}/x)$ and $0 < x \leq \sqrt{n}/\log(n)$,

$$s = O(x/\sqrt{n}) = o(1), \qquad hs \leq 2\log(\sqrt{n}/x)(x/\sqrt{n}) = o(1).$$

Now, on one hand, since $y\log^3(1/y) \to 0$ as $y \to 0+$, $h^3 s = o(1)$ and $h^3 s^3 = o(s^2)$. Combining this with elementary Taylor expansion,

$$e^{\pm hs} = 1 \pm hs + \frac{(hs)^2}{2} + o(s^2). \tag{C.41}$$

On the other hand, applying Taylor expansion to $\Phi(s \pm h)$ and noting that ϕ is a symmetric function,

$$\Phi(s \pm h) = \Phi(\pm h) + \phi(h)s - h\phi(h)s^2 + o(s^2). \tag{C.42}$$

where we have used that the third derivative of Φ is a bounded function. Combining (C.41)-(C.42) and re-arranging,

$$\begin{aligned} & e^{hs}\Phi(s+h) + e^{-hs}\Phi(s-h) \\ =& 1 + 2s\phi(h) + hs[\Phi(h) - \Phi(-h)] + h^2 s^2/2 + o(s^2) \\ =& 1 + u(h)s + h^2 s^2/2 + o(s^2), \end{aligned} \tag{C.43, C.44}$$

where in the first step, we have used $\Phi(h) + \Phi(-h) = 1$, and in the second step, we have used the expression of $u(h)$ given in (C.34).

We now analyze $\log[e^{hs}\Phi(s+h) + e^{-hs}\Phi(s-h)]$. Write for short $w = e^{hs}\Phi(s+h) + e^{-hs}\Phi(s-h) - 1$. By (C.44) and $|u(h)| \le h+1$ from (C.34), $|w| \le C\max\{(h+1)s, h^2 s^2\}$, and so

$$|\log(1+w) - w + w^2/2| \le C|w|^3 \le C\max\{(h+1)^3 s^3, h^6 s^6\},$$

where by similar argument as above, $\max\{(h+1)^3 s^3, h^6 s^6\} = o(s^2)$. Combining this with (C.44),

$$\begin{aligned} & \log[e^{hs}\Phi(s+h) + e^{-hs}\Phi(s-h)] \\ =& \log(1+w) \\ =& w - w^2/2 + o(s^2) \\ =& u(h)s + h^2 s^2/2 - [u(h)s + h^2 s^2/2]^2/2 + o(s^2) \\ =& u(h)s + (h^2 - u(h)^2)s^2/2 - [u(h)h^2 s^3 + h^4 s^4/4]/2 + o(s^2), \end{aligned}$$

where we note $|u(h)h^2 s^3 + h^4 s^4/4| \le C(h+1)h^2 s^3 + h^4 s^4/4 = o(s^2)$. As a result,

$$\log[e^{hs}\Phi(s+h) + e^{-hs}\Phi(s-h)] = u(h)s + (h^2 - u(h)^2)s^2/2 + o(s^2).$$

Combining this with (C.39) and the expression of $\sigma(h)$ given in (C.34) and rearranging it,

$$ys - \log[M_Y(s)] = (y - u(h))s - (1 + h^2 - u(h)^2)\frac{s^2}{2} + o(s^2) = (y - u(h))s - \sigma^2(h)\frac{s^2}{2} + o(s^2).$$

Now, invoking $s = \frac{1}{\sigma^2(h)} x/\sqrt{n}$ and $y = u(h) + x/\sqrt{n}$ gives

$$ys - \log[M_Y(s)] = \frac{1}{2\sigma^2(h)}(x/\sqrt{n})^2(1 + o(1)).$$

Combining this with (C.40) gives the claim.

We now consider the case of $h \geq 2\log(\sqrt{n}/x)$. We wish to use (C.40) again, with the same y but a different s: $s = x/\sqrt{n}$. In the current case, since $x \leq \sqrt{n}/\log(n)$,

$$h \to \infty, \qquad s \to 0.$$

By the assumptions of $h \geq 2\log(\sqrt{n}/x)$ and $s = x/\sqrt{n}$, and

$$\phi(h/2) \leq C\exp(-(\log(\sqrt{n}/x))^2/2) = o(s^2),$$

it follows that $\max\{\Phi(-s-h), \Phi(s-h)\} \leq \Phi(-h/2) = o(1)\phi(h/2)$, where the right hand side is $o(s^2)$. As a result,

(C.45)
$$e^{hs}\Phi(s+h) + e^{-hs}\Phi(s-h) = e^{hs}[1 - \Phi(-s-h) + e^{-2hs}\Phi(s-h)] = e^{hs}[1 + o(s^2)],$$

and so

$$\log[e^{hs}\Phi(s+h) + e^{-hs}\Phi(s-h)] = hs + o(s^2).$$

Combining this with (C.39) and (C.47) and invoking $s = x/\sqrt{n}$ and $y = u(h) + x/\sqrt{n}$,

$$\begin{aligned} ys - \log M_Y(s) &= (u(h) + x/\sqrt{n} - h)s - s^2/2 + o(s^2) \\ &= s^2/2 + o(s^2) \\ &= s^2/(2\sigma^2(h)) + o(s^2), \end{aligned} \tag{C.46}$$

where in the last two steps, we have used

(C.47)
$$h - u(h) = 2h\Phi(-h) - 2\phi(h) = o(s), \qquad \sigma^2(h) = 1 + h^2 - u(h)^2 = 1 + o(s).$$

Inserting (C.46) into (C.40) gives the claim. ∎

C.3. Proof of Lemma 3.3. Denote by Φ the CDF of $N(0,1)$. By direct calculations,

$$u(h) = \sqrt{2/\pi}e^{-h^2/2} + h[1 - 2\Phi(-h)].$$

This implies $u(h) \to \sqrt{2/\pi}$ when $h \to 0$ and $u(h)/h \to 1$ when $h \to \infty$. Furthermore,

$$\begin{aligned} u'(h) &= -2h\phi(h) + [1 - 2\Phi(-h)] + 2h\phi(-h) = 1 - 2\Phi(-h), \\ u''(h) &= 2\phi(-h) > 0. \end{aligned}$$

So $u(h)$ is strictly convex and monotony increasing for $h \in (0, \infty)$.

Let h_0 be the unique solution of $u'(h) = 0.9$. Fix (h_1, h_2) such that $h_2 > h_1 > 0$. If $h_1 > h_0$, by convexity,

$$u(h_2) - u(h_1) \geq u'(h_1)(h_2 - h_1) \geq 0.9(h_2 - h_1).$$

If $h_2 < h_0$, using the Taylor expansion, for some $\tilde{h} \in [h_1, h_2]$,

$$u(h_2) - u(h_1) = u'(h_1)(h_2 - h_1) + \frac{1}{2}u''(\tilde{h})(h_2 - h_1)^2 \geq \frac{1}{2}u''(h_0)(h_2 - h_1)^2.$$

If $h_1 < h_0 < h_2$, then we decompose the difference into $u(h_2) - u(h_0) + u(h_0) - u(h_1)$ and combine with the two cases we just dicussed, then we have that

$$u(h_2) - u(h_1) \geq 0.9(h_2 - h_0) + C_1(h_0 - h_1)^2.$$

When $h_2 - h_0 \geq h_0 - h_1$, then we have $u(h_2) - u(h_1) \geq 0.45(h_2 - h_0) + 0.45(h_0 - h_1) = 0.45(h_2 - h_1)$; otherwise, there is $u(h_2) - u(h_1) \geq \frac{C_1}{2}[(h_2 - h_0)^2 + (h_0 - h_1)^2] \geq C(h_2 - h_1)^2$. Combining the three cases gives the claim. □

APPENDIX D: PROOF OF SECONDARY LEMMAS

In this section, we show the proof of Lemmas B.3, B.5 and C.1.

D.1. Proof of Lemma B.3. The following lemma is useful, which is proved below.

Lemma D.1 *For any fixed $q > 0$,*

$$\pi_0^{(q)} = \bar{\Phi}\big(\sqrt{2q\log(p)}\big)\big(1 + L_p n^{-1/2}\big),$$
$$\pi_1^{(q)} = \begin{cases} 1 - L_p p^{-(\sqrt{r}-\sqrt{q})^2}, & r > q, \\ \bar{\Phi}\big((\sqrt{q} - \sqrt{r})\sqrt{2\log(p)}\big)\big(1 + L_p n^{-1/4}\big), & r \leq q. \end{cases}$$

First, we prove (a). Write for short $z_j = z$ and $z^{(q)} = z_j^{(q)}$. Since the distribution of $z^{(q)}$ is spherically symmetric, $v'z^{(q)}$ has the same distribution as $e_1'z^{(q)}$, for any $v \in \mathcal{S}^{n-1}$. It follows that $E[(v'z^{(q)})^2] = E[(z^{(q)}(1))^2] = a_p$. Furthermore, $E(\|z^{(q)}\|^2) = nE[(z^{(q)}(1))^2] = na_p$.

Consider $E(|v'z^{(q)}|^m)$. Again, by spherical symmetry,

$$E(|v'z^{(q)}|^m) = E\big(|z^{(q)}(1)|^m\big) = E\big(|z(1)|^m 1\{z^2(1) + \|\tilde{z}\|^2 > n + 2\sqrt{qn\log(p)}\}\big),$$

where $\tilde{z} = (z(2), \cdots, z(n))'$. Note that $\tilde{z}$ is independent of $z(1)$ and $\|\tilde{z}\|^2 \sim \chi^2_{n-1}$. Let B_1 be the event that $|z(1)| \leq \sqrt{2\delta_1 \log(p)}$, for some δ_1 to determine. From basic properties of the $N(0,1)$ distribution, $P(B_1^c) = L_p p^{-\delta_1}$ and $E(|z(1)|^m I_{B_1^c}) = L_p p^{-\delta_1}$. It follows that

$$\begin{aligned} E(|v'z^{(q)}|^m) &\leq E\big(|z(1)|^m 1\{z^2(1) + \|\tilde{z}\|^2 > n + 2\sqrt{qn\log(p)}, B_1\}\big) + L_p p^{-\delta_1} \\ &\leq E\big(|z(1)|^m 1\{\|\tilde{z}\|^2 > n + 2\sqrt{qn\log(p)} - 2\sqrt{\delta_1 \log(p)}\}\big) + L_p p^{-\delta_1} \\ &= E(|z(1)|^m) \cdot P(\|\tilde{z}\|^2 > n + 2\sqrt{qn\log(p)}(1+o(1))) + L_p p^{-\delta_1} \\ &= E(|z(1)|^m)\pi_0(1+o(1)) + L_p p^{-\delta_1}. \end{aligned}$$

By choosing δ_1 appropriately large, we find that the first term dominates.

Consider $E(\|z^{(q)}\|^{2m})$. Denote by f_n the density of χ^2_n, where $f_n(y) = \frac{y^{n/2-1}e^{-y}}{2^{n/2}\Gamma(n/2)}$. Note that $y^m f_n(y) = \frac{2^m \Gamma(m+n/2)}{\Gamma(n/2)} f_{n+2m}(y)$. It follows that

$$E(\|z^{(q)}\|^{2m}) = \frac{2^m \Gamma(m+n/2)}{\Gamma(n/2)} P(\chi^2_{n+2m} > n + 2\sqrt{qn\log(p)}).$$

First, by letting $q = 0$ on both hand sides, we have $\kappa_{2m}(n) = E(\|z\|^{2m}) = \frac{2^m\Gamma(m+n/2)}{\Gamma(n/2)}$. Second, since $n+2\sqrt{qn\log(p)} = n_* + 2\sqrt{qn_*\log(p)}(1+L_p n^{-1/2})$ for $n_* = n+2m$, Lemma D.1 implies that $P(\chi^2_{n+2m} > n + 2\sqrt{qn\log(p)}) = \pi_0(1+o(1))$. Together, the above right hand side is $\kappa_{2m}(n)\pi_0(1+o(1))$.

Consider a_p. Similarly to the above, for $n_* = n+2$,

$$\begin{aligned} a_p = n^{-1}E(\|z^{(q)}\|^2) &= \frac{2\Gamma(1+n/2)}{n\Gamma(n/2)} P(\chi^2_{n+2} > n_* + 2\sqrt{qn_*\log(p)}(1+L_p n^{-1/2})) \\ &= P(\chi^2_{n+2} > n_* + 2\sqrt{qn_*\log(p)}(1+L_p n^{-1/2})) \\ &= \pi_0(1+L_p n^{-1/2}). \end{aligned}$$

Second, we prove (b). We first state an approximation of π_1. From basic properties of chi-square distributions, for all $q, r \geq 0$,

$$\begin{aligned} &P(\chi^2_n(0) > n + 2\sqrt{qn\log(p)}) = \bar{\Phi}(\sqrt{2q\log(p)})(1+L_p n^{-1/2}), \\ &P(\chi^2_n(2r\log(p)) > n + 2\sqrt{qn\log(p)}) = \bar{\Phi}\big((\sqrt{q}-\sqrt{r})\sqrt{2\log(p)}\big)(1+L_p n^{-1/4}). \end{aligned}$$

Therefore, we find that

$$\begin{aligned} \pi_1 &= P(\chi^2_n(2r\log(p)) > n + 2\sqrt{qn\log(p)}) \\ &= P(\chi^2_n(0) > 2(\sqrt{q}-\sqrt{r})\sqrt{n\log(p)}) \cdot (1+o(1)). \end{aligned} \tag{D.48}$$

Consider $E[(v'z_j^{(q)})^2]$. Fix v and introduce

$$w_1 = \ell/\|\ell\|, \qquad w_2 = (1-(v'\ell)^2/\|\ell\|^2)^{-1/2}[v-(v'\ell)\ell/\|\ell\|^2].$$

Both w_1 and w_2 are unit vectors and $w_1'w_2 = 0$. Let Q be any orthogonal matrix whose first two columns are w_1 and w_2. By direct calculations, $Q'v = (x_0, \sqrt{1-x_0^2}, 0, \cdots, 0)'$ and $Q'\ell = (\sqrt{n}, 0, \cdots, 0)$, where $x_0 = (v'\ell)/\|\ell\|$. Since $Q'z$ and z have the same distribution,

$$\begin{aligned} v'z_j^{(q)} &= v'QQ'z\cdot 1\{\|Q'z+\mu(j)Q'\ell\|^2 > n+2\sqrt{qn\log(p)}\}] \\ &\overset{(d)}{=} v'Qz\cdot 1\{\|z+\mu(j)Q'\ell\|^2 > n+2\sqrt{qn\log(p)}\}] \\ &= \big[x_0 z(1) + (1-x_0^2)^{1/2} z(2)\big]\cdot 1\{\|z+\sqrt{n}\tau_p^* e_1\| > n\}. \end{aligned} \tag{D.49}$$

It follows that

$$\begin{aligned} E[(v'z_j^{(q)})^2] &= E\big[(x_0 z(1) + (1-x_0^2)^{1/2} z(2))^2 1\{\|z+\sqrt{n}\tau_p^* e_1\|^2 > n+2\sqrt{qn\log(p)}\}\big] \\ &= (1-x_0^2)E\big[(z(2))^2 1\{\|z+\sqrt{n}\tau_p^* e_1\|^2 > n+2\sqrt{qn\log(p)}\}\big] \\ &\quad + x_0^2\ E\big[(z(1))^2 1\{\|z+\sqrt{n}\tau_p^* e_1\|^2 > n+2\sqrt{qn\log(p)}\}\big] \\ &= b_p + (c_p - b_p)(v'\ell)^2/\|\ell\|^2, \end{aligned}$$

where the second equality comes from the symmetry on $z(2)$ (so the cross term disappears).

Consider b_p and c_p. Let $\tilde{z} = (z(2), \cdots, z(n))'$, where $\|\tilde{z}\|^2 \sim \chi^2_{n-1}$ and it is independent of $z(1)$. We write

$$c_p = E\big[(z(1))^2 1\{\|\tilde{z}\|^2 > n+2\sqrt{qn\log(p)} - g(z(1))\}\big], \quad g(x) \equiv (x+\sqrt{n}\tau_p^*)^2.$$

For a constant $\delta_2 > 0$ to be determined, let B_2 be the event that $|z(1)| \leq \sqrt{2\delta_2\log(p)}$. From basic properties of normal distributions, $P(B_2^c) = L_p p^{-\delta_2}$ and $E[z^2(1)I_{B_2^c}] = L_p p^{-\delta_2}$. Over the event B_2, we have $g(z(1)) = [z(1) - (2\sqrt{nr\log(p)})^{1/2}]^2 = 2\sqrt{rn\log(p)}(1+L_p n^{-1/4})$. It follows that

$$\begin{aligned} c_p &\leq E\big[(z(1))^2\cdot P(B_2\cap\{\|\tilde{z}\|^2 > n+2\sqrt{qn\log(p)} - g(z(1))\}|z(1))\big] + L_p p^{-\delta_2} \\ &\leq E[(z(1))^2]\cdot P\big(\chi^2_{n-1} > n + 2(\sqrt{q}-\sqrt{r})\sqrt{n\log(p)}(1+L_p n^{-1/4})\big) + L_p p^{-\delta_2} \\ &= \pi_1(1+L_p n^{-1/4}), \end{aligned}$$

where the last inequality comes from (D.48) and that δ_2 is chosen appropriately large. To compute b_p, we write

$$\begin{aligned} b_p &= E\big[(z(2))^2 1\{\|\tilde{z}\|^2 > n+2\sqrt{qn\log(p)} - g(z(1))\}\big] \\ &= (n-1)^{-1}E\big[\|\tilde{z}\|^2 1\{\|\tilde{z}\|^2 > n+2\sqrt{qn\log(p)} - g(z(1))\}\big]. \end{aligned}$$

Let B_2 be the same event. Let $q_* = [(\sqrt{q}-\sqrt{r})_+]^2$. We have

$$\begin{aligned}b_p &= (n-1)^{-1}E\big[\|\tilde{z}\|^2 1\{\|\tilde{z}\|^2 > n + 2\sqrt{q_* n\log(p)}(1+L_p n^{-1/4})\}\big] + L_p p^{-\delta_2}\\ &= (n-1)^{-1}E(\|\tilde{z}^{(q_*)}\|^2)(1+L_p n^{-1/4}) = \pi_1(1+L_p n^{-1/4}),\end{aligned}$$

where in the last equality, we have applied the result in (a) with $q = q_*$.

Consider $E(|v'z^{(q)}|^m)$. Let $\tilde{w} = (z(3), \cdots, z(n))'$. Then $\|\tilde{w}\|^2 \sim \chi^2_{n-2}$ and it is independent of $(z(1), z(2))$. By (D.49),

$$\begin{aligned}E(|v'z^{(q)}|^m) = E\big(&|x_0 z(1) + (1-x_0^2)^{1/2}z(2)|^m\\ &\cdot 1\{\|\tilde{w}\|^2 > n + 2\sqrt{qn\log(p)} - g(z(1)) - (z(2))^2\}\big).\end{aligned}$$

Let B_3 be the event that $\max\{|z(1)|, |z(2)|\} \leq \sqrt{2\delta_3\log(p)}$. Then $P(B_3^c) = L_p p^{-\delta_3}$ and over B_3, $g(z(1)) + (z(2))^2 = 2\sqrt{rn\log(p)}(1+o(1))$. Applying similar arguments as above, we find that

$$E(|v'z^{(q)}|^m) \leq E(|x_0 z(1) + (1-x_0^2)^{1/2}z(2)|^m)\cdot \pi_1(1+o(1)) = \kappa_m \pi_1(1+o(1)).$$

Here the last inequality is because $x_0 z(1) + (1-x_0)^{1/2}z(2) \sim N(0,1)$. The claim then follows.

Consider $E(\|z^{(q)}\|^2)$ and $E(\|z^{(q)}\|^{2m})$. Using Q defined above (for an arbitrary v)

$$\begin{aligned}E(\|z^{(q)}\|^2) &= E\big(\|Q'z\|^2 1\{\|Q'z + \tau_p^* Q'\ell\|^2 > n + 2\sqrt{qn\log(p)}\}\big)\\ &= E\big(\|z\|^2 1\{\|z + \sqrt{n}\tau_p^* e_1\|^2 > n + 2\sqrt{qn\log(p)}\}\big)\\ &= E\big((z(1))^2 1\{\|z + \sqrt{n}\tau_p^* e_1\|^2 > n + 2\sqrt{qn\log(p)}\}\big)\\ &\quad + (n-1)E\big((z(2))^2 1\{\|z + \sqrt{n}\tau_p^* e_1\|^2 > n + 2\sqrt{qn\log(p)}\}\big)\\ &= c_p + (n-1)b_p.\end{aligned}$$

Recall that $\tilde{z} = (z(2), \cdots, z(n))'$, $q_* = [(\sqrt{q}-\sqrt{r})_+]^2$ and $g(x) = (x+\sqrt{n}\tau_p^*)^2$ for any $x \in R$. Note that $(x+y)^m \leq 2^m(|x|^m + |y|^m)$ for any $x, y \in R$. We have

$$\begin{aligned}E(\|z^{(q)}\|^{2m}) &= E\big(\|z\|^{2m} 1\{\|z + \sqrt{n}\tau_p^* e_1\|^2 > n + 2\sqrt{qn\log(p)}\}\big)\\ &\leq 2^m E\big((z(1))^{2m} 1\{\|\tilde{z}\|^2 > n + 2\sqrt{qn\log(p)} - g(z(1))\}\big)\\ &\quad + 2^m E\big(\|\tilde{z}\|^{2m} 1\{\|\tilde{z}\|^2 > n + 2\sqrt{qn\log(p)} - g(z(1))\}\big)\\ &= 2^m \kappa_{2m}\pi_1(1+o(1)) + 2^m E\big(\|\tilde{z}^{(q_*)}\|^{2m}\big)(1+o(1))\\ &= 2^m(\kappa_{2m} + \kappa_{2m}(n-1))\cdot \pi_1(1+o(1)). \qquad \text{(D.50)}\end{aligned}$$

Here, we have applied the result in (a) for $E(\|z^{(q)}\|^{2m})$ with $q = q_*$.

Last, we prove (c). Using the spherical symmetry of $z^{(q)}$ and the Q defined above, we have already seen that $\|z + \tau_p^* \ell\| \overset{(d)}{=} \|z + \sqrt{n}\tau_p^* e_1\|$ and

$$z \cdot 1\{\|z + \tau_p^* \ell\| > n + 2\sqrt{qn\log(p)}\} \overset{(d)}{=} z \cdot 1\{\|z + \sqrt{n}\tau_p^* e_1\| > n + 2\sqrt{qn\log(p)}\}.$$

Then the claims follow from the definitions and (a)-(b). □

D.2. Proof of Lemma B.5. Let $\pi(j) = \pi_0$ for $j \notin S(\mu)$ and $\pi(j) = \pi_1$ for $j \in S(\mu)$, where π_0, π_1 are defined in Section B.1. Then $\sum_{j=1}^p \pi(j) = m^{(q)}$ by (c) of Lemma B.3.

First, consider $\mathrm{tr}(H_0)$. Write $M_j = n^{-1}[\|z_j^{(q)}\|^2 - E(\|z_j^{(q)}\|^2)]$. By definition,

$$n^{-1}\mathrm{tr}(H_0) - m_*^{(q)} = \sum_{j=1}^p M_j. \tag{D.51}$$

By Lemma B.3, $E(\|z_j^{(q)}\|^2) \lesssim n\pi(j)$ and $E(\|z_j^{(q)}\|^{2m}) \le 2^m \kappa_{2m}(n)\pi(j) \le C4^m \pi(j) n^m$, where $\kappa_{2m}(n)$ is the m-th moment of the χ_n^2 distribution and we have used $\kappa_{2m}(n) \le C2^m n^m$. Noting that $(a+b)^m \le 2^m(a^m + b^m)$ for any real values a and b, by direct calculations,

$$E(M_j) = 0, \qquad \mathrm{var}(M_j) \le C\pi(j), \qquad E(|M_j|^m) \le C8^m.$$

By Lemma B.1 (Bernstein inequality), with probability at least $1 - O(p^{-3})$,

$$|\sum_{j=1}^p M_j| \le C\sqrt{m^{(q)}\log(p)}. \tag{D.52}$$

Combining (D.51)-(D.52) gives the first claim.

Second, consider $\|H_0\|_F^2$. By direct calculations,

$$\begin{aligned} \|H_0\|_F^2 - n^{-1}[\mathrm{tr}(H_0)]^2 &= \sum_{1\le j,k\le p} [(z_j^{(q)})' z_k^{(q)}]^2 - n^{-1}(\sum_{j=1}^p \|z_j^{(q)}\|^2)^2 \\ &= \frac{n-1}{n}\sum_{j=1}^p \|z_j^{(q)}\|^4 + 2\sum_{1\le j<k\le p}\left([(z_j^{(q)})' z_k^{(q)}]^2 - \frac{1}{n}\|z_j^{(q)}\|^2\|z_k^{(q)}\|^2\right) \\ &\equiv (I) + (II). \end{aligned} \tag{D.53}$$

We now study (I). Write $U_j = n^{-1}\|z_j^{(q)}\|^4$ for short. By Lemma B.3, $E(U_j) = n^{-1}\kappa_4(n)\pi_0(1+o(1))$ and $\mathrm{var}(U_j) \le Cn^2\pi_0$ for $j \notin S(\mu)$; moreover, $\mathrm{var}(U_j) \le$

$Cn^2\pi_1$ for $j \in S(\mu)$. We also claim that $E(U_j) \geq n^{-1}\kappa_4(n-1)\pi_1(1+o(1))$ for $j \in S(\mu)$. The proof is similar to that for (D.50), but in the second line of (D.50), we instead use the inequality $\|z\|^{2m} \geq \|\tilde{z}\|^{2m}$. Note that $\kappa_4(n)$ is the second moment of χ^2_n and so $\kappa_4(n) = n^2 + 2n$. It follows that

$$\sum_{j=1}^{p} E(U_j) \gtrsim nm^{(q)}, \qquad \sum_{j=1}^{p} \mathrm{var}(U_j) \leq Cn^2 m^{(q)}.$$

Using Lemma B.1, with probability at least $1 - O(p^{-3})$, $\sum_{j=1}^{p} U_j \gtrsim nm^{(q)} - Cn\sqrt{m^{(q)}\log(p)} \gtrsim Cnm^{(q)}$. Since $(I) = (n-1)\sum_{j=1}^{p} U_j$,

$$(I) \geq C_1 n^2 m^{(q)}, \qquad \text{for some constant } C_1 > 0. \tag{D.54}$$

We then study (II). Let $V_{jk} = [(z_j^{(q)})' z_k^{(q)}]^2 - \frac{1}{n}\|z_j^{(q)}\|^2\|z_k^{(q)}\|^2$. Introduce

$$W_j(v) = [v' z_j^{(q)}]^2 - n^{-1}\|z_j^{(q)}\|^2, \qquad \text{for any } v \in \mathcal{S}^{n-1}.$$

Let $v_j = z_j/\|z_j\|$. Then v_j is independent of $\|z_j^{(q)}\|$ and $V_{jk} = \|z_j^{(q)}\|^2 W_k(v_j)$. By Lemma B.3, for any fixed $v \in \mathcal{S}^{n-1}$,

$$\begin{array}{ll} E[W_j(v)] = 0, \;\; E[(W_j(v))^2] \leq C\pi_0, & j \notin S(\mu), \\ E[W_j(v)] = (c_p - b_p)[(v'\tilde{\ell})^2 - n^{-1}], \;\; E[(W_j(v))^2] \leq C\pi_1, & j \in S(\mu), \end{array}$$

where $\tilde{\ell} = \ell/\|\ell\|$. As a result, if either $j \notin S(\mu)$ or $k \notin S(\mu)$, then $E(V_{jk}) = 0$; if both $j, k \in S(\mu)$, then $E(V_{jk}) = (c_p - b_p)E\{\|z_j^{(q)}\|^2[(v_j'\tilde{\ell})^2 - n^{-1}]\} = (c_p - b_p)E[W_j(\tilde{\ell})] = (1 - n^{-1})(c_p - b_p)^2 \geq 0$. It follows that

$$E[(II)] \geq 0. \tag{D.55}$$

To compute $\mathrm{var}((II))$, we calculate $E(V_{jk}V_{j'k'})$ for all (j,k,j',k') such that $j \neq k$ and $j' \neq k'$. Since $V_{jk} = V_{kj}$, we assume $j \neq k'$ and $j' \neq k$ without loss of generality. We have the following observations: (1) $E(V_{jk}) \leq (c_p - b_p)^2$ if both $j, k \in S(\mu)$ and $E(V_{jk}) = 0$ otherwise. (2) $E(V_{jk}^2) = E(\|z_j^{(q)}\|^2)E[W_k^2(v_j)] \leq Cn^2\pi(j)\pi(k)$ for any $j \neq k$. (3) When $j \neq j'$ and $k \neq k'$, V_{jk} is independent of $V_{j'k'}$, so $E(V_{jk}V_{j'k'}) = E(V_{jk})E(V_{j'k'})$. (4) When $j = j'$ and $k \neq k'$, $E(V_{jk}V_{jk'}) = E[\|z_j^{(q)}\|^4 W_k(v_j)W_{k'}(v_j)]$; as a result, $E(V_{jk}V_{jk'}) = 0$ when either $k \notin S(\mu)$ or $k' \notin S(\mu)$; if $k, k' \in S(\mu)$,

$E(V_{jk}V_{jk'}) = (c_p - b_p)^2 E[(W_j(\tilde{\ell}))^2] \leq C(c_p - b_p)^2 \pi(j)$. Therefore,

$$\sum_{\substack{(j,j',k,k'):\\ j\neq k, j'\neq k'}} E(V_{jk}V_{j'k'}) \leq \sum_{(j,k):j\neq k} Cn^2 \pi(j)\pi(k) + \sum_{\substack{(j,k,k'):j\notin\{k,k'\}\\ \{k,k'\}\subset S(\mu), k\neq k'}} C(c_p - b_p)^2 \pi(j)$$
$$+ \sum_{\substack{(j,j',k,k'):\{j,j',k,k'\}\subset S(\mu)\\ j,j',k,k' \text{ are different}}} (c_p - b_p)^4$$
$$\leq Cn^2 (m^{(q)})^2 + C(c_p - b_p)^2 m^{(q)} |S(\mu)|^2 + C(c_p - b_p)^4 |S(\mu)|^4$$
$$\leq Cn^2 (m^{(q)})^2 + m^{(q)} (|S(\mu)|\pi_1)^2 \cdot o(1) + C(|S(\mu)|\pi_1)^4 \cdot o(1),$$

where the last inequality is due to that $c_p - b_p = o(\pi_1)$. Using (B.15), when $r < \rho^*_\theta(\beta)$ ("impossibility") and $q < \tilde{q}(\beta, \theta, r)$ ("fat" case), $(|S(\mu)|\pi_1)^2 = o(m^{(q)})$ and so the first term in the above dominates the other two. It implies

$$\text{(D.56)} \qquad \mathrm{var}((II)) \leq C_2 n^2 (m^{(q)})^2, \qquad \text{for some constant } C_2 > 0.$$

We combine (D.55)-(D.56) and apply the Markov inequality. It follows that with probability at least $1 - 4n^{-2} C_2 / C_1^2$,

$$\text{(D.57)} \qquad (II) \geq -C_1 n^2 m^{(q)}/2.$$

The second claim follows by plugging (D.54) and (D.57) into (D.53). □

D.3. Proof of Lemma C.1. The proof is similar to that of Lemma 2.1. By Lemma B.4, there exists an $(1/4)$-net of $\mathcal{S}^{n-1}$, denoted as $\mathcal{M}^*_{1/4}$, such that $|\mathcal{M}^*_{1/4}| \leq 9^n$ and $\sup_{v\in\mathcal{M}^*_{1/4}} v'Av \geq 2\|A\|$ for any $n \times n$ matrix A. Therefore, to show the claim, it suffices to show that for each fixed $v \in \mathcal{M}^*_{1/4}$, with probability $\geq 1 - O(9^{-n}p^{-2})$,

$$\text{(D.58)} \qquad |v'(ZBB'Z' - \mathrm{tr}(BB')I_n)v| \leq C\sqrt{np}.$$

Denote the eigenvalue decomposition of BB' by $V'\Lambda V$, where Λ is diagonal matrix with diagonals $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Fix v, we can write

$$v'ZBB'Z'v = v'ZV'\Lambda VZ'v = \sum_{i=1}^p \lambda_i \eta_i^2, \qquad \eta_i \stackrel{iid}{\sim} N(0,1).$$

The last equation comes from $VZ'v \sim N(0, I_p)$. So we have $E[v'ZBB'Z'v] = \mathrm{tr}(BB')$ for any fixed v with $\|v\| = 1$. Let $W_j = \lambda_i\eta_i^2/\lambda_1 - \lambda_i/\lambda_1$, then W_j's are independent of each other, $E(W_j) = 0$, $\mathrm{var}(W_j) \leq 2$ and $E(|W_j|^m) \leq$

κ_{2m}. We apply Lemma B.1 with $\lambda = 2\sqrt{n\log(9) + 2\log(p)}$. To check the moment conditions, we note that $\kappa_{2m} = E_{z\sim N(0,1)}(|z|^{2m}) \leq 2^m m!$ for all $m \geq 1$. It follows that with probability $\geq 1 - O(9^{-n}p^{-2})$,

$$\lambda_1|\sum_{j=1}^{p} W_j| \leq 2\sqrt{np\log(9) + 2p\log(p)}\lambda_1 \leq C\sqrt{np}.$$

The last inequality is because $\lambda_1 = \|B\|^2 \leq L_p$. This proves (D.58). □

D.4. Proof of Lemma D.1. We start from computing π_0. Using the density of the χ_n^2 distribution,

$$\pi_0 = \int_{n+2\sqrt{qn\log(p)}}^{\infty} \frac{x^{n/2-1}e^{-x/2}dx}{2^{n/2}\Gamma(n/2)} \equiv \frac{1}{2^{n/2}\Gamma(n/2)} \cdot (I).$$

Now, we calculate the integral (I). Write for short

$$t = \sqrt{2q\log(p)} \qquad \text{and} \qquad x_0 = n + \sqrt{2n}t.$$

With a variable change $x = n + \sqrt{2n}y$, we have

$$\begin{aligned}
(I) &= \sqrt{2n}x_0^{n/2-1}e^{-x_0/2}\int_0^{\infty}(1 + \sqrt{2n}y/x_0)^{n/2-1}e^{-\sqrt{2n}y/2}dy \\
&= \sqrt{2n}x_0^{n/2-1}e^{-x_0/2}\int_0^{\infty}\exp\Big\{(n/2-1)\log(1+\sqrt{2n}y/x_0) - \sqrt{2n}y/2\Big\}dy \\
&\equiv \sqrt{2n}x_0^{n/2-1}e^{-x_0/2}\big[(I_1) + (I_2) + (I_3)\big], \qquad \text{(D.59)}
\end{aligned}$$

where (I_1) contains the integral from 0 to ct, (I_2) contains that from ct to $x_0/\sqrt{2n}$ and (I_3) contains that from $x_0/\sqrt{2n}$ to infinity. We will determine the constant $c > 0$ later.

Consider (I_1). From the Taylor expansion, $\log(1+a) = a - a^2/2 + O(a^3)$ for small a. Moreover, $n/x_0 = 1 - \sqrt{2n}t/x_0 + O(t^2/n)$, $x_0 = O(n)$ and $y = O(t)$. As a result, for $0 < y < ct$, by simple calculations,

$$(n/2-1)\log(1+\sqrt{2n}y/x_0) - \sqrt{2n}y/2 = -ty - y^2/2 + O(t^3/\sqrt{n}).$$

Noting that $e^a = 1 + O(a)$ for small a, so (I_1) is equal to $\int_0^{ct} e^{-ty-y^2/2}dy \cdot [1 + O(t^3/\sqrt{n})]$. By direct calculation, $\int_0^{ct} e^{-ty-y^2/2}dy = e^{t^2/2}\int_t^{(1+c)t} e^{-y^2/2}dy = \sqrt{2\pi}e^{t^2/2}\big[\bar{\Phi}(t) - \bar{\Phi}((1+c)t)\big]$. By Mills' ratio, $\bar{\Phi}(t) = L_p p^{-q}$ and $\bar{\Phi}((1+c)t) =$

$L_p p^{-(1+c)^2 q}$. Therefore, when c is chosen large enough, $\bar{\Phi}((1+c)t) = o(1) \cdot \bar{\Phi}(t) L_p n^{-1/2}$. It follows that

$$(I_1) = \sqrt{2\pi} e^{t^2/2} \bar{\Phi}(t) \big(1 + L_p/\sqrt{n}\big). \tag{D.60}$$

Consider (I_2). Since $\log(1+a) - a \le a^2/4$ for $a \in [0,1]$, when $ct < y < x_0/\sqrt{2n}$,

$$(n/2-1)\log(1+\sqrt{2n}y/x_0) - \sqrt{2n}y/2 = -ty - y^2/4 + O(tx_0^2/(\sqrt{n})^3)$$

As a result, $(I_2) \le (1 + L_p n^{-1/2}) \int_{ct}^{\infty} e^{-yt-y^2/4} dy$, where $\int_{ct}^{\infty} e^{-yt-y^2/4} dy = 2\sqrt{\pi} e^{t^2} \bar{\Phi}((c+2)t/\sqrt{2}) = e^{t^2/2} L_p \sqrt{n} p^{-[(c+2)^2/2-1]q}$. By choosing c appropriately large, we have

$$(I_2) = o(1) \cdot e^{t^2/2} L_p n^{-1/2}. \tag{D.61}$$

Consider (I_3). Since $\log(1+a) \le t$ for all $a \ge 0$, when $t > x_0/\sqrt{2n}$,

$$(n/2-1)\log(1+\sqrt{2n}y/x_0) - \sqrt{2n}y/2 \le -(n/x_0)ty \le -ty/2.$$

It follows that $(I_3) \le \int_{x_0/\sqrt{2n}}^{\infty} e^{-ty/2} dy = (2/t) e^{-x_0^2/(2n)} = o(1) \cdot e^{t^2/2} L_p n^{-1/2}$.

Combining the above results for (I_1)-(I_3), we obtain that

$$\pi_0 = R_n(t) \cdot \bar{\Phi}(t) \big(1 + L_p n^{-1/2}\big), \qquad \text{where } R_n(t) \equiv \frac{2\sqrt{\pi n} x_0^{n/2-1} e^{-x_0/2 + t^2/2}}{2^{n/2}\Gamma(n/2)}.$$

We plug in $x_0 = n + \sqrt{2n}t$ and rewrite $R_n(t) = \frac{n}{x_0} \frac{\sqrt{\pi}(n/e)^{n/2}}{\Gamma(n/2)} \big(1 + t\sqrt{2/n}\big)^{n/2} e^{-t\sqrt{n/2} + t^2/2}$. Note that by Taylor expansion, $\log(1+a) = a - a^2/2 + O(a^3)$ for $a = t\sqrt{2/n}$. Therefore, we have $R_n(t) = \frac{n}{x_0} \frac{\sqrt{\pi}(n/e)^{n/2}}{\Gamma(n/2)} \exp\big\{O(t^3/\sqrt{n})\big\} = 1 + L_p n^{-1/2}$. This gives

$$\pi_0 = \bar{\Phi}(t)\big(1 + L_p n^{-1/2}\big).$$

Next, we compute π_1. Define

$$\tilde{r} = \frac{(z(1) - \sqrt{n}\tau_p)^2}{2\sqrt{n\log(p)}}, \qquad W = \sum_{i=2}^{n} z^2(i).$$

Then $\tilde{r}$ and W are independent; furthermore, W has a χ^2_{n-1} distribution. We rewrite

$$\pi_1 = E\left[P\left(\frac{W-n}{\sqrt{2n}} > (\sqrt{q} - \sqrt{\tilde{r}})\sqrt{2\log(p)} \middle| \tilde{r} \right) \right].$$

For a constant $c > 0$ to be determined, let B_1 be the event that $|z(1)| \leq \sqrt{2c\log(p)}$. Then $P(B_1^c) = L_p p^{-c}$. Over the event B_1, $\tilde{r} = r + L_p n^{-1/4}$. When $r > q$, utilizing the results for π_0, we get

$$\begin{aligned}\pi_1 &= \Phi\big((\sqrt{r} - \sqrt{q} + o(1))\sqrt{2\log(p)}\big)(1 + L_p n^{-1/2}) + L_p p^{-c} \\ &= 1 - L_p p^{-(\sqrt{r}-\sqrt{q})^2} + L_p p^{-c}.\end{aligned}$$

When $r < q$.

$$\begin{aligned}\pi_1 &= \bar{\Phi}\big((\sqrt{q} - \sqrt{r} + L_p n^{-1/4})\sqrt{2\log(p)}\big)(1 + L_p n^{-1/2}) + L_p p^{-c} \\ &= \bar{\Phi}\big((\sqrt{q} - \sqrt{r})\sqrt{2\log(p)}\big)(1 + L_p n^{-1/4}) + L_p p^{-c}.\end{aligned}$$

We choose c large enough so that $L_p p^{-c}$ is always dominated by any other term. This gives the claim for π_1. □

REFERENCES

[1] Azizyan, M., Singh, A. and Wasserman, L. (2013). Minimax theory for high-dimensional Gaussian mixtures with sparse mean separation. 2139–2147.

[2] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. OUP Oxford.

[3] Chaudhuri, K., Dasgupta, S. and Vattani, A. (2009). Learning mixtures of Gaussians using the k-means algorithm. *arXiv preprint arXiv:0912.0086*.

[4] Jin, J., Ke, Z. T. and Wang, W. (2014). Optimal spectral clustering by Higher Criticism Thresholding. *Manuscript*.

[5] Jin, J., Ke, Z. T. and Wang, W. (2015). Phase transitions for high dimensional clustering and related problems. *Manuscript*.

[6] Shorack, G. and Wellner, J. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons.

[7] Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing* 210–268.

J. Jin
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania, 15213
USA
E-mail: jiashun@stat.cmu.edu

Z. Ke
Department of Statistics
University of Chicago
Chicago, Illinois, 60637
USA
E-mail: zke@galton.uchicago.edu

W. Wang
Department of Biostatistics and Epidemiology
University of Pennsylvania, Perelman School of Medicine
Philadelphia, Pennsylvania, 19104
USA
E-mail: wanjiew@wharton.upenn.edu