Additional Simulations and Proofs for "The Statistics and Mathematics of High Dimension Low Sample Size Asymptotics"

Dan Shen, Haipeng Shen, Hongtu Zhu, and J. S. Marron

Supplementary Material

The supplementary material is organized as follows.

- Section S1 discusses the curious case of n = 1 in HDLSS contexts.
- Section S2 shows additional simulation results for Example 5.1 in the main paper. These results provide empirical support for the theoretical convergence stated in Theorem 5.1.
- Section S3 defines the angle between a vector and a linear space which is introduced in Section 5.1.1 of the paper.
- Section S4 presents the extensions of Theorems 5.1 and 5.2 under the growing sample size context.
- Section S5 shows several extensions of Theorem 5.3 under the HDLSS setting.
- Section S6 presents the detailed proofs of the growing sample size asymptotic results. Specifically, Section S6.1 presents the proof of sample eigenvalues' properties in Theorem 5.2, Section S6.2 provides the proofs of Propositions S4.1 and S4.2, Section S6.3 shows the proof of Theorem 4.2, and Section S6.4 provides the proofs of several Lemmas.
- Section S7 shows the detailed proofs of HDLSS results. Section S7.1 provides the proof of Theorem 5.3, Sections S7.2-S7.5 include the proofs of Propositions S5.1-S5.4, Section S7.6 provides the proof of Theorem 4.1, and Section S7.7 includes the detailed proofs of needed lemmas.

S1 The Curious Case of n = 1

Interesting context for HDLSS results comes from the observation that the mathematical results hold for any sample size, including $n = 1$. Of course no statistician would think of attempting to do meaningful inference with only one data point. Furthermore the contemplation of consistency in such a context at first sounds absurd to most classically trained mathematical statisticians. However the mathematics in Section 3 of the main paper still apply in this case. The reason is that in the limit as $d \to \infty$, because of the growing stretch of the spike distribution, the first sample eigenvector (just the unit vector in the direction of the single data point) will tend to lie increasingly in the direction of the first population eigenvector.

This phenomenon as a demonstration that the spike assumption, in particular the $\alpha > 1$ case in Section 3, is unrealistically strong for any practical setting, thus casting doubt on the whether any practically useful insights can be gained from the mathematical results in Section 3. But while studying this issue two major points need to be kept in mind:

- There are real data situations where PCA reveals scientifically important structure in data. Figure 1 shows just one example of this, but there are many other cases as well.
- The mathematics of (3.1) in Section 3 gives a clear dichotomy, which shows that in HDLSS cases PCA will tend to either find important structure in data or else will find random directions which will show pure noise as projections.

Taking both sides of the data analytic and mathematical components together, leads to the conclusion that while the $\alpha > 1$ spike may feel very strong, in fact it is a reasonable model for many natural phenomena.

S2 Additional simulation results

As mentioned in Example 5.1 of the main paper, we carried out 100 simulation runs under various settings with sample size n = 50, 100, 200, 500, 1000, 2000; $\frac{d}{n} = 50$; and ratios $c_1 = 0.2, c_2 = 0.4, c_3 = 1$, to study the convergence of the sample eigenvectors to the respective cones around the corresponding population eigenvectors. The simulation results are plotted in Figure A.

Each panel in the first column is for a particular sample size, and plots the estimated angles between the sample eigenvectors and the corresponding population eigenvectors (as jittered points), along with the corresponding kernel density estimates and the theoretical angles (in vertical dash lines). Note that Panel (A) of Figure 3 in the main paper is for the case of $n = 200$. The plots clearly show that, as sample size n increases from 50 to 2000, the kernel density estimators are getting more and more concentrated around the corresponding theoretical angles θ_j, $j = 1, 2, 3$, which validates the convergence results.

The panels in the right column depicts the randomness of the sample eigen-directions within the cones, as in Panel (B) of Figure 3 for the case of $n = 200$. Within each panel, the jittered points are the angles between a random pair of sample eigenvectors (that correspond to the same population eigenvector), superimposed with the corresponding kernel density estimates (colored accordingly). As sample size n increases, the pairwise angles are increasingly close to 90 degrees, which is consistent with the randomness in high dimensions that has been identified in the literature.

S3 Angle between a vector and a linear space

This section rigorously defines the angle between a vector and a linear space, which was introduced in Section 5.1.1 of the paper. Denote H to be an index set, e.g. $H = \{m + 1, \cdots, d\}$,

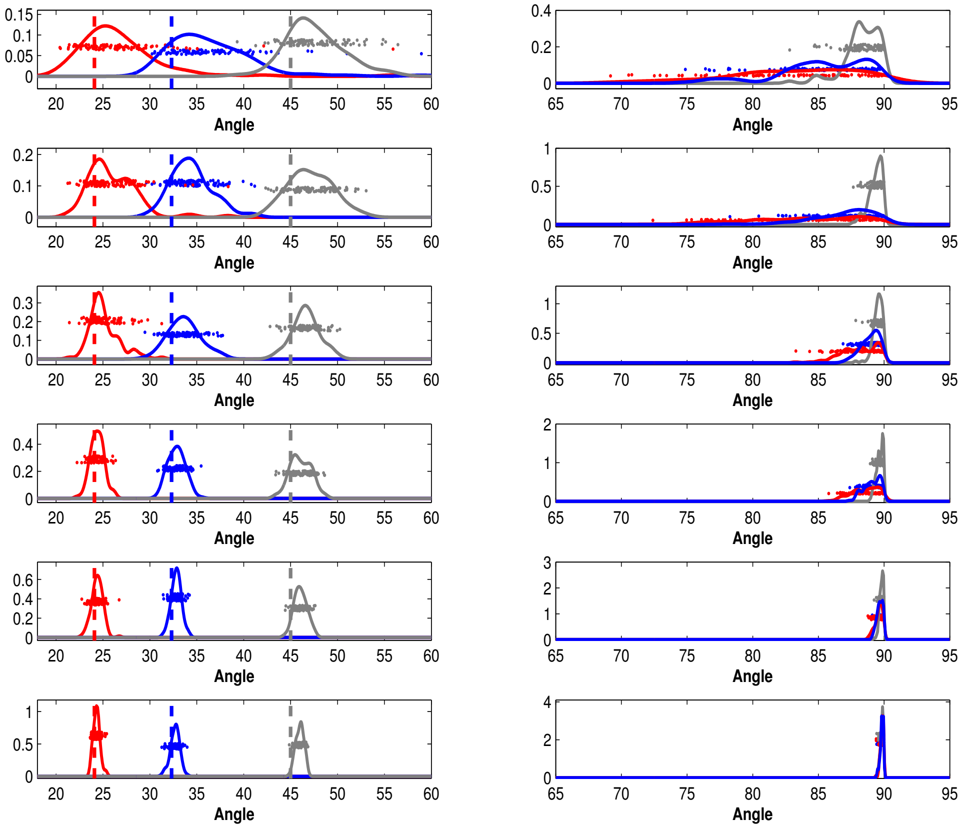

Figure A: Additional simulation plots for Example 5.1 in the main paper. Panels from top to bottom correspond to $n = 50, 100, 200, 500, 1000$, and 2000, respectively. In particular, the panels in the third row correspond to Figure 3 in the main paper.

and $S = \text{span}\{u_k, k \in H\}$ to be the subspace generated by $\{u_k, k \in H\}$, shown as the gray space in Figure B. For each sample eigenvector $\hat{u}_j$, $j \in H$, define the projection vector onto the subspace S as $\hat{u}_j^{proj} = \sum_{k\in H} < \hat{u}_j, u_k > u_k$. Then, the angle between $\hat{u}_j$ and S is defined as

$$\text{angle} < \hat{u}_j, S >= \text{angle} < \hat{u}_j, \hat{u}_j^{proj} >= \cos^{-1}\sqrt{\sum_{k\in H} < \hat{u}_j, u_k >^2},$$

which is shown as θ in Figure B.

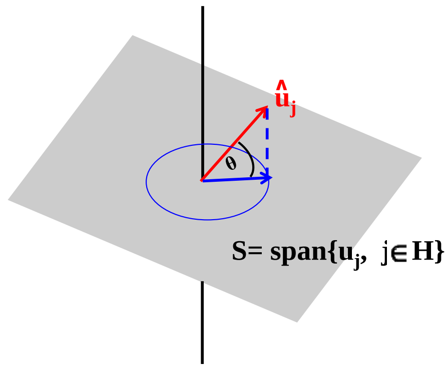

Figure B: Angle between the sample eigenvector $\hat{u}_j$ and the space $\mathbb{S}$. The blue vector is the projection of the red vector $\hat{u}_j$ onto the space $\mathbb{S}$.

S4 Extensions under the growing sample size asymptotic framework

In this section, we discuss the extensions of Theorems 5.1 and 5.2 under the growing sample size asymptotic framework. Section S4.1 is about the extensions of Theorem 5.1. Section S4.2 shows the extensions of Theorem 5.2.

S4.1 Extensions of Theorem 5.1

This section presents two natural extensions of Theorem 5.1 that include random matrix ($\frac{n}{d} \to c$, as $n \to \infty$), and High Dimension Medium Sample Size (HDMSS, $d \gg n \to \infty$) cases.

The first extension of Theorem 5.1 includes the random matrix cases, by allowing $c_{m_0} = 0$ for some $m_0 \leq m$, which suggests that positive information dominates in the leading m_0 spikes. Then Assumptions $\mathcal{A}1$ and $\mathcal{A}2$ respectively become

$\mathcal{A}3$. as $n, d \to \infty$, the population eigenvalues satisfy

$$\lambda_1 > \cdots > \lambda_{m_0} \gg \lambda_{m_0+1} > \cdots > \lambda_m \gg \lambda_{m+1} \to \cdots \to \lambda_d = 1.$$

$\mathcal{A}4$. as $n, d \to \infty$, $\frac{d}{n\lambda_j} \to c_j$ for $j = 1, \cdots, m$, where $0 = c_1 = \cdots = c_{m_0} < c_{m_0+1} < \cdots < c_m < \infty$.

For $d \ll n$, $m_0 = m = d$ in Assumptions $\mathcal{A}3$ and $\mathcal{A}4$. For random matrix cases with $n \sim d$, $m_0 = m$ in Assumptions $\mathcal{A}3$ and $\mathcal{A}4$. Since $c_1 = \cdots = c_{m_0} = 0$ in Assumption $\mathcal{A}3$, if the eigenvalue index is less than or equal to m_0, the corresponding sample eigenvalues and eigenvectors are consistent. These results are summarized in the following Proposition S4.1(a).

Another extension of Theorem 5.1 is to allow $c_{m_0+1} = \infty$ for some $m_0 \leq m$, i.e. negative information dominates in higher-order spikes. This contains the HDMSS cases (Cabanski et al., 2010; Yata and Aoshima, 2012; Aoshima and Yata, 2015), where $d \gg n \to \infty$. Assumption $\mathcal{A}1$ then becomes Assumption $\mathcal{A}3$, and Assumption $\mathcal{A}2$ becomes

$\mathcal{A}5$. as $n, d \to \infty$, $\frac{d}{n\lambda_j} \to c_j$ for $j = 1, \cdots, m$, where $0 < c_1 < \cdots < c_{m_0} < c_{m_0+1} = \cdots = c_m = \infty$.

Since $c_{m_0+1} = \cdots = c_m = \infty$, for index $j \geq m_0 + 1$, the proportional error between the sample and population eigenvalues goes to infinity, and the angle between the corresponding sample and population eigenvectors converges to 90 degrees. These results are summarized in Proposition S4.1(b).

Proposition S4.1. *(a) Under Assumptions 4.1, $\mathcal{A}3$ and $\mathcal{A}4$, the sample eigenvalues and eigenvectors satisfy*

$$\frac{\hat{\lambda}_j}{\lambda_j} \quad \text{and} \quad |< \hat{u}_j, u_j >| \quad \xrightarrow{a.s} 1, \quad 1 \leq j \leq m_0,$$

and the properties of the other sample eigenvalues and eigenvectors remain the same as in Theorem 5.1.

(b) Let $H = \{m_0 + 1, \cdots, d\}$ and define $\mathbb{S}$ as in (5.1). If Assumption $\mathcal{A}4$ in (a) is replaced by Assumption $\mathcal{A}5$, the sample eigenvalues satisfy

$$\frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{a.s}} 1, \quad m_0 + 1 \leq j \leq m,$$

and the sample eigenvectors satisfy

$$\begin{cases} |< \hat{u}_j, u_j >| = \mathrm{O}_{\text{a.s}} \left\{ (\frac{n\lambda_j}{d})^{\frac{1}{2}} \right\}, \\ \text{angle} < \hat{u}_j, \mathbb{S} > \xrightarrow{a.s} 0, \end{cases} \qquad m_0 \leq j \leq [n \wedge d];$$

the properties of the other sample eigenvalues and eigenvectors remain the same as in Theorem 5.1.

S4.2 Extensions of Theorem 5.2

In this section, we discuss two natural extensions of Theorem 5.2 that include $d \ll n$, random matrix ($\frac{n}{d} \to c$, as $n \to \infty$), and High Dimension Medium Sample Size (HDMSS, $d \gg n \to \infty$) cases.

One way to extend Theorem 5.2 is to allow $c_{r_0} = 0$ for some $r_0 \leq r$. Then, Assumptions $\mathcal{B}2$ and $\mathcal{B}3$ of Section 5.1.2 (main paper), respectively, become

$\mathcal{B}4$. as $n, d \to \infty$, the eigenvalues in different tiers have different limits:

$$\delta_1 > \cdots > \delta_{r_0} \gg \delta_{r_0+1} > \cdots > \delta_r \gg \lambda_{m+1} \to \cdots \to \lambda_d = 1.$$

$\mathcal{B}5$. as $n, d \to \infty$, $\frac{d}{n\delta_k} \to c_k$, for $k = 1, \cdots, r$, where $0 = c_1 = \cdots = c_{r_0} < c_{r_0+1} \cdots < c_r < \infty$.

This scenario contains $d \ll n$ and the random matrix cases. For $d \ll n$, $r_0 = r$ and $m = d$ in Assumptions $\mathcal{B}4$ and $\mathcal{B}5$. For random matrix cases with $n \sim d$, $r_0 = r$ in Assumptions $\mathcal{B}4$ and $\mathcal{B}5$. Since $c_1 = \cdots = c_{r_0} = 0$, the sample eigenvalues and eigenvectors, whose subgroup index is less than or equal to r_0, are, respectively, consistent and subspace consistent. These results are summarized in Proposition S4.2(a).

Another extension of Theorem 5.2 is obtained by allowing $c_{r_0+1} = \infty$. Then Assumption $\mathcal{B}2$ becomes Assumption $\mathcal{B}4$, and Assumption $\mathcal{B}3$ becomes

$\mathcal{B}6$. as $n, d \to \infty$, $\frac{d}{n\delta_k} \to c_k$, for $k = 1, \cdots, r$, where $0 < c_1 < \cdots < c_{r_0} < c_{r_0+1} = \cdots = c_r = \infty$.

This scenario contains the HDMSS case. Since $c_{r_0+1} = \cdots = c_r = \infty$, for eigenvalue index $k \geq r_0+1$, the ratios between the sample and corresponding population eigenvalues go to infinity. Furthermore, the corresponding angles between the sample eigenvectors and their population counterparts converge to 90 degrees. These results are summarized in Proposition S4.2(b).

Proposition S4.2.

(a) *Under Assumptions 4.1, $\mathcal{B}1$, $\mathcal{B}4$ and $\mathcal{B}5$, the sample eigenvalues and eigenvectors satisfy that*

$$\frac{\hat{\lambda}_j}{\lambda_j} \quad \text{and} \quad \text{angle} < \hat{u}_j, \mathbb{S}_k > \quad \xrightarrow{a.s} 1, \quad j \in H_k,\ k = 1, \cdots, r_0; \tag{S4.1}$$

in addition, the properties of the rest sample eigenvalues and eigenvectors remain the same as those in Theorem 5.2.

(b) *Let $H = \bigcup_{k=r_0+1}^{r+1} H_k$ and define $\mathbb{S}$ as in (5.1). If Assumption $\mathcal{B}5$ in (a) is replaced by Assumption $\mathcal{B}6$, the sample eigenvalues satisfy:*

$$\frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{a.s}} 1, \quad \sum_{k=1}^{r_0} q_k + 1 \leq j \leq m; \tag{S4.2}$$

the sample eigenvectors satisfy that

$$\begin{cases} |< \hat{u}_j, u_j >| = \mathrm{O}_{\text{a.s}}\left\{ (\frac{n\lambda_j}{d})^{\frac{1}{2}} \right\}, \\ \text{angle} < \hat{u}_j, \mathbb{S} > \xrightarrow{a.s} 0, \end{cases} \quad \sum_{k=1}^{r_0} q_k + 1 \leq j \leq [n \wedge d]; \tag{S4.3}$$

in addition, the properties of the rest sample eigenvalues and eigenvectors remain the same as those in Theorem 5.2.

Remark S4.1. If every tier only contains one eigenvalue, then $r = m$ and Proposition S4.2 reduces to Proposition S4.1.

S5 Extensions under the HDLSS asymptotic framework

We discuss two extensions of Theorem 5.3 under the HDLSS contexts. In ***Scenario 1***, we allow $c_{m_0} = 0$ for some $m_0 \le m$, and summarize the corresponding results in Propositions S5.1 and S5.2, which, respectively, correspond to Scenario (a) in Propositions S4.2 and S4.1 in the growing sample size asymptotic contexts. For ***Scenario 2***, we allow $c_{m_0+1} = \infty$, and state the corresponding results in Propositions S5.3 and S5.4, which similarly correspond to Scenario (b) in Propositions S4.2 and S4.1.

S5.1 Scenario 1

This subsection studies the HDLSS asymptotic properties of PCA under the first scenario, where the first m population eigenvalues are further partitioned into two groups, and the positive information dominates in the first subgroup, e.g $\frac{d}{n\lambda_j} \to 0$ for $j \le m_0$. Then we can obtain different asymptotic properties of PCA in these two groups.

Scenario 1 furthermore contains two different scenarios. The first one is where the sample eigenvalues and eigenvectors within the first subgroup are asymptotically indistinguishable, which corresponds to Proposition S4.2 (a). Now Assumptions $\mathcal{C}1$ and $\mathcal{C}2$ in Theorem 5.3, respectively, become

$\mathcal{C}3$. For fixed n, as $d \to \infty$, the population eigenvalues satisfy

$$\lambda_1 \ge \cdots \ge \lambda_{m_0} \gg \lambda_{m_0+1} \ge \cdots \ge \lambda_m \gg \lambda_{m+1} \to \cdots \to \lambda_d = 1.$$

$\mathcal{C}4$. For fixed n, as $d \to \infty$, $\frac{d}{n\lambda_j} \to c_j$, for $j = 1, \cdots, m$, where $0 = c_1 = \cdots = c_{m_0} < c_{m_0+1} \le \cdots \le c_m < \infty$.

Assumption $\mathcal{C}3$ assumes that the first m population eigenvalues are further separated into two groups. Assumption $\mathcal{C}4$ guarantees that the positive information dominates in the first subgroup.

We now define several non-negative random matrices, whose eigenvalues describe the HDLSS asymptotic properties of PCA. Denote

$$c_j^* = \lim_{d\to\infty} \frac{\lambda_{m_0}}{\lambda_j}, \quad j = 1, \cdots, m_0. \tag{S5.4}$$

Then define $m_0 \times d$ and $(m - m_0) \times d$ matrices $\mathbb{M}_1^1$ and $\mathbb{M}_2^1$ as following:

$$\begin{aligned} \mathbb{M}_1^1 &= \left[\mathbb{C}_1^1, 0_{m_0\times(d-m_0)}\right]_{m_0\times d}, \\ \mathbb{M}_2^1 &= \left[0_{(m-m_0)\times m_0}, \mathbb{C}_2^1, 0_{(m-m_0)\times(d-m)}\right]_{(m-m_0)\times d}, \end{aligned}$$

where $\mathbb{C}_1^1 = \mathrm{diag}({c^*}_1^{-\frac{1}{2}}, \cdots, {c^*}_{m_0}^{-\frac{1}{2}})$, $\mathbb{C}_2^1 = \mathrm{diag}(c_{m_0+1}^{-\frac{1}{2}}, \cdots, c_m^{-\frac{1}{2}})$, and $0_{k\times l}$ is the $k\times l$ zero matrix. Finally, we define random matrices

$$\mathcal{W}_1^1 = \mathbb{M}_1^1 Z^T Z {\mathbb{M}_1^1}^T \quad \text{and} \quad \mathcal{W}_2^1 = \mathbb{M}_2^1 Z^T Z {\mathbb{M}_2^1}^T, \tag{S5.5}$$

whose eigenvalues describe the HDLSS limiting behavior of PCA.

Since the sample eigenvalues are asymptomatically indistinguishable, similar to Theorem 5.3, we will study subspace consistency and need to define spaces

$$\mathbb{S}_k^1 = \text{span}\{u_j, j \in H_k^1\}, \quad k = 1, 2, 3,$$

where $H_1^1 = \{1, \cdots, m_0\}$, $H_2^1 = \{m_0 + 1, \cdots, m\}$ and $H_3^1 = \{m + 1, \cdots, d\}$.

Proposition S5.1. *Under Assumptions 4.1, $\mathcal{C}3$ and $\mathcal{C}4$, for fixed n, as $d \to \infty$, we have*

$$\begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} \frac{c_j^*}{n}\lambda_j(\mathcal{W}_1^1), \quad 1 \le j \le m_0, \\ \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} \frac{c_j}{n}\lambda_{j-m_0}(\mathcal{W}_2^1) + c_j, \quad m_0 + 1 \le j \le m, \end{cases} \tag{S5.6}$$

and

$$\begin{cases} angle < \hat{u}_j, \mathbb{S}_1^1 > \xrightarrow{a.s} 1, \quad 1 \le j \le m_0, \\ angle < \hat{u}_j, \mathbb{S}_2^1 > \xrightarrow{a.s} \arccos\left\{\left(1 + \frac{n}{\lambda_{j-m_0}(\mathcal{W}_2^1)}\right)^{-\frac{1}{2}}\right\}, \quad m_0 + 1 \le j \le m. \end{cases} \tag{S5.7}$$

Moreover, the properties of the rest sample eigenvalues and eigenvectors remain the same as those in Theorem 5.3.

Remark S5.1. If $m_0 = 0$, then $0 < c_j < \infty$ for $j = 1, \cdots, m$ and Proposition S5.1 reduces to Theorem 5.3. If $m_0 = m$, the angles between domain sample eigenvectors and the corresponding subspace converge to 0.

Remark S5.2. Since the positive information dominates in the first subgroup, it follows from (S5.7) that the angles between the sample eigenvectors within the first subgroup and the corresponding space converge to 0. However, the angles between the sample eigenvectors within the second subgroup and the corresponding space converge to non-degenerate random variables.

The second sub-scenario within Scenario 1 is where the sample eigenvalues and eigenvectors within the first subgroup are asymptotically distinguishable. It corresponds to Proposition S4.1 (a). Then Assumption $\mathcal{C}1$ becomes

$\mathcal{C}5$. For fixed n, as $d \to \infty$, the population eigenvalues satisfy

$$\lambda_1 \gg \cdots \gg \lambda_{m_0} \gg \lambda_{m_0+1} \ge \cdots \ge \lambda_m \gg \lambda_{m+1} \to \cdots \to \lambda_d = 1.$$

Proposition S5.2. *Under Assumptions 4.1, $\mathcal{C}4$ and $\mathcal{C}5$, for fixed n, as $d \to \infty$, we have*

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} R_j, \quad 1 \le j \le m_0, \tag{S5.8}$$

where R_j is in (4.8), and

$$|< \hat{u}_j, u_j >| \xrightarrow{a.s} 1, \quad 1 \le j \le m_0. \tag{S5.9}$$

Moreover, the properties of the rest sample eigenvalues and eigenvectors remain the same as those in Proposition S5.1. If Assumption 4.1 is strengthened to normal distribution, then R_j has the same distribution with $\frac{\chi_n^2}{n}$.

Remark S5.3. If $m_0 = 0$, then $0 < c_j < \infty$ for $j = 1, \cdots, m$ and Proposition S5.2 becomes Theorem 5.3. If $m_0 = m$, the first m sample eigenvalues and eigenvectors are all asymptotically distinguishable, and the angles between the sample and corresponding population eigenvectors converge to 0.

Remark S5.4. The difference between Propositions S5.1 and S5.2 is that $\lambda_1 \geq \cdots \geq \lambda_{m_0}$ is replaced by $\lambda_1 \gg \cdots \gg \lambda_{m_0}$. This determines whether the sample eigenvalues and eigenvectors within the first subgroup can be asymptotically identified or not.

S5.2 Scenario 2

This subsection studies the HDLSS asymptotic properties of PCA in Scenario 2. In this case, the negative information dominates in the second subgroup (e.g $\frac{d}{n\lambda_j} \to \infty$ for $j > m_0$). Similar to Scenario 1, Scenario 2 also contains two sub-scenarios. The first one is where the sample eigenvalues and eigenvectors within the first subgroup are asymptotically indistinguishable, which corresponds to Proposition S4.2 (b). Then Assumption $\mathcal{C}1$ becomes Assumption $\mathcal{C}3$ and Assumption $\mathcal{C}4$ becomes

$\mathcal{C}6$. For fixed n, as $d \to \infty$, $\frac{d}{n\lambda_j} \to c_j$, for $j = 1, \cdots, m$, where $0 < c_1 < \cdots < c_{m_0} < c_{m_0+1} = \cdots = c_m = \infty$.

For $j = 1, \cdots, m_0$, replace c_j^* by c_j in the definition of $\mathcal{W}_1^1$ in (S5.5) to define $\mathcal{W}^2$, whose eigenvalues determine the HDLSS asymptotic behavior of PCA within the first subgroup. In addition, we need to define spaces $\mathbb{S}_1^2 = \mathbb{S}_1^1$ and $\mathbb{S}_2^2 = \text{span}\{u_j, m_0 + 1 \leq j \leq d\}$ to explore the subspace consistency.

Under Assumptions $\mathcal{C}3$ and $\mathcal{C}6$, the sample eigenvalues $\hat{\lambda}_j$ for $j = m_0 + 1, \cdots, m$ cannot be asymptotically identified. This leads to that the angles between $\hat{u}_j$ and u_j for $j = m_0 + 1, \cdots, m$ converges to 90 degrees. These results are summarized in the following proposition.

Proposition S5.3. *Under Assumptions 4.1, $\mathcal{C}3$ and $\mathcal{C}6$, for fixed n, as $d \to \infty$, the sample eigenvalues satisfy*

$$\begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} \frac{c_j}{n}\lambda_j(\mathcal{W}^2) + c_j, \quad 1 \leq j \leq m_0, \\ \frac{n\hat{\lambda}_j}{d\lambda_d} \xrightarrow{\text{a.s}} 1, \quad m_0 + 1 \leq j \leq m; \end{cases} \tag{S5.10}$$

in addition, the properties of the rest sample eigenvalues remain the same as those in Theorem 5.3. Moreover, we have

$$\begin{cases} angle < \hat{u}_j, \mathbb{S}_1^2 > \xrightarrow{a.s} \arccos\left\{\left(1 + \frac{n}{\lambda_j(\mathcal{W}^2)}\right)^{-\frac{1}{2}}\right\}, \quad 1 \leq j \leq m_0, \\ |< \hat{u}_j, u_j >| = \mathrm{O}_{\text{a.s}}\left\{(\frac{\lambda_j}{d})^{\frac{1}{2}}\right\}, \quad m_0 + 1 \leq j \leq n, \\ angle < \hat{u}_j, \mathbb{S}_2^2 > \xrightarrow{a.s} 1, \quad m_0 + 1 \leq j \leq n. \end{cases} \tag{S5.11}$$

Remark S5.5. If there is no $c_j = \infty$ for $j = 1, \cdots, m$, Proposition S5.3 reduces to Theorem 5.3. If all $c_j = \infty$ for $j = 1, \cdots, m$, then all sample eigenvalues cannot be asymptotically identified and the corresponding angles between the sample and population eigenvectors converge to 90°.

The second sub-scenario in Scenario 2 is where the sample eigenvalues and eigenvectors within the first subgroup are asymptotically distinguishable, which corresponds to Proposition S4.1 (b). Now Assumption $\mathcal{C}1$ becomes Assumption $\mathcal{C}5$ and Assumption $\mathcal{C}2$ becomes

$\mathcal{C}7$. For fixed n, as $d \to \infty$, $\frac{d}{n\rho_j} \to c_j$, for $j = 1, \cdots, m$, where $0 = c_1 = \cdots = c_{m_0} < c_{m_0+1} = \cdots = c_m = \infty$.

The results can be summarized as

Proposition S5.4. *Under Assumptions 4.1, $\mathcal{C}5$ and $\mathcal{C}7$, for fixed n, as $d \to \infty$, the sample eigenvalues and eigenvectors, whose index is less than or equal to m_0, have the same properties as those in Proposition S5.2. Moreover, the properties of the rest sample eigenvalues and eigenvectors remain the same as those in Proposition S5.3.*

Remark S5.6. Proposition S5.4 is a combination of Propositions S5.2 and S5.3. Proposition S5.4 is consistent with the results in Jung and Marron (2009); Shen et al. (2012).

S6 Proofs for the growing sample size asymptotic results

This section provides detailed proofs of the theorems, propositions and lemmas in the growing sample size contexts. Section S6.1.1 proves the asymptotic properties of the sample eigenvalues as stated in Theorem 5.2. Section S6.2 shows the proofs of Propositions S4.1 and S4.2. Section S6.3 presents the proof of Theorem 4.2. Section S6.4 proves the lemmas needed in proving Theorem 5.2.

S6.1 The proof of sample eigenvalues' properties in Theorem 5.2

This subsection shows the detailed proof of asymptotic properties of sample eigenvalues. Without loss of generality (WLOG), we assume that $\lambda_{m+1} = \cdots = \lambda_d = 1$. Due to the invariance property of the angle between the sample and population eigenvectors, see Shen et al. (2012), we assume WLOG that the population eigenvectors $u_j = e_j$, $j = 1, \ldots, d$, where the j-th component of e_j equals 1 and the rest are zero. The sample eigenvalue properties are studied through the dual matrix $\hat{\Sigma}_D$, which shares the same nonzero eigenvalues with the sample covariance matrix $\hat{\Sigma}$. Since $u_j = e_j$, $j = 1, \ldots, d$, then it follows from (4.2) and (4.8) that the dual matrix can be written as

$$\hat{\Sigma}_D = \frac{1}{n} X^T X = \frac{1}{n} \sum_{j=1}^{d} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T,$$

which is partitioned into two matrices as follows:

$$\hat{\Sigma}_D = A + B, \quad \text{with} \quad A = \frac{1}{n} \sum_{j=1}^{m} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T, \quad B = \frac{1}{n} \sum_{j=m+1}^{d} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T. \tag{S6.12}$$

The following proof contains two steps. The first one is to show the asymptotic properties of the eigenvalues of A and B in Lemmas S6.1 and S6.2, respectively. The second one is to use the *Wielandt's Inequality* (Rao, 2002), now restated as Lemma S6.3, to study the asymptotic properties of the sample eigenvalues.

We first give out three preliminary lemmas.

Lemma S6.1. *As $n, d \to \infty$, the eigenvalues of the matrix A in* (S6.12) *satisfy*

$$\frac{\lambda_k(A)}{\lambda_k} \xrightarrow{\text{a.s}} 1, \quad k = 1, \cdots, m,$$

where $\lambda_k(A)$ denotes the kth largest eigenvalue of the matrix A.

Lemma S6.2. *As $n, d \to \infty$, the eigenvalues of the matrix B in* (S6.12) *satisfy*

$$\frac{n\lambda_k(B)}{d\lambda_{k+m}} \xrightarrow{\text{a.s}} 1, \quad k = 1, \cdots, [n \wedge (d-m)].$$

Lemma S6.3. *Assume that A, B are $m \times m$ real symmetric matrices, then for all $k = 1, \ldots, m$,*

$$max_{i+j=k+m} \{\lambda_i(A) + \lambda_j(B)\} \leq \lambda_k(A+B) \leq min_{i+j=k+1} \{\lambda_i(A) + \lambda_j(B)\}.$$

The proofs of Lemmas S6.1 and S6.2 are, respectively, shown in Sections S6.4.1 and S6.4.2.

S6.1.1 Asymptotic properties of the sample eigenvalues

According to Lemma S6.3, we have that

$$\frac{\lambda_j(A)}{\lambda_j} + \frac{\lambda_n(B)}{\lambda_j} \leq \frac{\hat{\lambda}_j}{\lambda_j} \leq \frac{\lambda_j(A)}{\lambda_j} + \frac{\lambda_1(B)}{\lambda_j}. \tag{S6.13}$$

According to Lemma S6.2 and Assumption $\mathcal{B}3$ and $\lambda_{m+1} = \cdots = \lambda_d = 1$, we have that as $n, d \to \infty$

$$\frac{\lambda_1(B)}{\lambda_j} = \frac{d\lambda_{1+m}}{n\lambda_j} \times \frac{n\lambda_1(B)}{d\lambda_{1+m}} \xrightarrow{a.s} c_k, \quad j \in H_k,\ k = 1, \cdots, r,$$
$$\frac{\lambda_n(B)}{\lambda_j} = \frac{d\lambda_{n+m}}{n\lambda_j} \times \frac{n\lambda_n(B)}{d\lambda_{n+m}} \xrightarrow{a.s} c_k, \quad j \in H_k,\ k = 1, \cdots, r,$$

which, together with (S6.13) and Lemma S6.1, yields that

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{a.s} 1 + c_k, \quad j \in H_k,\ k = 1, \cdots, r. \tag{S6.14}$$

In addition, since rank$(A) \leq m$, then $\lambda_j(A) = 0$ for $j > m$. Combining above with (S6.13) and Lemma S6.2, we have

$$\frac{n\hat{\lambda}_j}{d\lambda_j} \xrightarrow{a.s} 1, \quad j = m+1, \cdots, [n \wedge d]. \tag{S6.15}$$

which, together with (S6.14), yields the asymptotic properties of sample eigenvalues (5.6) in Theorem 5.2.

S6.2 Proofs of Propositions S4.1 and S4.2

This section provides the proofs of Propositions S4.1 and S4.2. Since Proposition S4.1 is a special case of Proposition S4.2, we just need to show the proof of Proposition S4.2.

In order to prove Proposition S4.2, we need to redefine A and B in (S6.12) as follows:

$$A = \frac{1}{n}\sum_{j=1}^{r_0} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T, \quad B = \frac{1}{n}\sum_{j=r_0+1}^{d} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T.$$

Then similar as (S6.14), we have

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{a.s} 1 + c_l, \quad j \in H_l,\ l = 1, \cdots, r_0. \tag{S6.16}$$

Under Assumption $\mathcal{B}5$, we have that $c_l = 0$ for $l = 1, \cdots, r_0$. Then it follows from (S6.16) that

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{a.s} 1, \quad j \in H_l,\ l = 1, \cdots, r_0. \tag{S6.17}$$

Under Assumption $\mathcal{B}6$, we have that $c_{r_0+1} = \cdots = c_r = \infty$. Note that $\lambda_{m+1} \to \cdots \to \lambda_d = 1$, and similar as (S6.15), we have

$$\frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{a.s}} 1, \quad \sum_{k=1}^{r_0} q_k \le j \le m. \tag{S6.18}$$

Now we consider the asymptotic properties of sample eigenvectors. Under Assumption $\mathcal{B}5$, we have that $c_l = 0$ for $l = 1, \cdots, r_0$ in (7.13). Then it follows that

$$\sum_{k \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s} 1, \quad j \in H_l,\ l = 1, \cdots, r_0,$$

which, together with (7.2), yields

$$\text{angle} < \hat{u}_j, \mathbb{S}_l > \xrightarrow{a.s} 1, \quad j \in H_l,\ l = 1, \cdots, r_0. \tag{S6.19}$$

In addition, since $\lambda_{m+1} \to \cdots \to \lambda_d = 1$, then similar as the proof of (7.3) and (7.4) in the main paper, we have

$$\begin{cases} |< \hat{u}_j, u_j >| = \mathrm{O}_{\text{a.s}}\left\{ (\frac{n\lambda_j}{d})^{\frac{1}{2}} \right\}, \\ \text{angle} < \hat{u}_j, \mathbb{S} > \xrightarrow{a.s} 0, \end{cases} \quad \sum_{k=1}^{r_0} q_k + 1 \le j \le [n \wedge d], \tag{S6.20}$$

where S is defined in Proposition S4.2.

Thus, according to (S6.17) and (S6.19), (S4.1) in Proposition S4.2 (a) is established. According to (S6.18) and (S6.20), (S4.2) and (S4.3) in Proposition S4.2 (b) are established.

S6.3 Proof of Theorem 4.2

Similar as in the Proof of Theorem 5.2, we assume WLOG that the population eigenvectors $u_j = e_j$. Then it follows from (4.2) that X_i has the following decomposition

$$X_i = \sum_{j=1}^{d} \lambda_j^{\frac{1}{2}} e_j z_{i,j}. \tag{S6.21}$$

According to the definition of population PC scores (4.3), the j-th population PC scores are

$$S_j = (S_{1,j}, \cdots, S_{n,j})^T = (z_{1,j}, \cdots, z_{n,j})^T. \tag{S6.22}$$

According to the definition of sample PC scores (4.5), the j-th sample PC scores are

$$\hat{S}_j = (\hat{S}_{1,j}, \cdots, \hat{S}_{n,j}) = \hat{\lambda}_j^{-\frac{1}{2}} (\hat{u}_j^T X_1, \cdots, \hat{u}_j^T X_n). \tag{S6.23}$$

From (S6.21), (S6.22) and (S6.23), the ratios between the sample and population PC scores are, for $i = 1, \cdots, n$, $j = 1, \cdots, m$,

$$\frac{\hat{S}_{i,j}}{S_{i,j}} = \frac{\lambda_j^{\frac{1}{2}}}{\hat{\lambda}_j^{\frac{1}{2}}} \hat{u}_{j,j} + \sum_{1 \le k \le m, k \ne j}^{m} \frac{\lambda_k^{\frac{1}{2}} z_{i,k}}{\hat{\lambda}_j^{\frac{1}{2}} z_{i,j}} \hat{u}_{k,j} + \sum_{k=m+1}^{d} \frac{\lambda_k^{\frac{1}{2}} z_{i,k}}{\hat{\lambda}_j^{\frac{1}{2}} z_{i,j}} \hat{u}_{k,j}. \tag{S6.24}$$

We denote the three terms on the right-hand-side of (S6.24) as Γ_k, $k = 1, 2, 3$. In order to obtain the asymptotic properties of the sample PC scores (4.11), it follows from (S6.24) that we just need to show that for $i = 1, \cdots, n$, $j = 1, \cdots, m$,

$$|\Gamma_1| \xrightarrow{a.s} 1, \tag{S6.25}$$

$$|\Gamma_k| \xrightarrow{a.s} 0, \quad k = 2, 3. \tag{S6.26}$$

We first prove (S6.25). Note that under spike model (4.10), Lemma S6.1 still holds and Lemma S6.2 becomes that almost surely

$$\frac{n\lambda_k(B)}{d\lambda_{k+m}} \sim 1, \quad k = 1, \cdots, [n \wedge (d-m)],$$

which, together with Lemma S6.3 (Wielandt's Inequality) and $\frac{d}{n^{\frac{1}{2}}\lambda_m} \to 0$, yields

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1, \quad j = 1, \cdots, m. \tag{S6.27}$$

In addition, (7.15) becomes

$$\hat{u}_{j,j}^2 \xrightarrow{\text{a.s}} 1, \quad j = 1, \cdots, m. \tag{S6.28}$$

It follows from (S6.27) and (S6.28) that (S6.25) is established.

Secondly, to prove (S6.26) for $k = 2$, we need to show

$$\lambda_k^{\frac{1}{2}} \hat{\lambda}_j^{-\frac{1}{2}} \hat{u}_{k,j} \xrightarrow{a.s} 0, \quad 1 \le j \ne k \le m. \tag{S6.29}$$

Note that for $1 \le j \ne k \le m$, we have $\lambda_k^{\frac{1}{2}}\hat{\lambda}_j^{-\frac{1}{2}} \xrightarrow{a.s} \lambda_k^{\frac{1}{2}}\lambda_j^{-\frac{1}{2}} \to c_{k,j}$, where $c_{k,j} = 0$ or ∞. Under such a scenario, Shen et al. (2012) has showed (S6.29).

Finally, we show the proof of (S6.26) for $k = 3$. According to Cauchy-Schwarz Inequality, we have that

$$\begin{aligned}\Gamma_3^2 &\le \frac{\lambda_{m+1}}{\hat{\lambda}_j z_{i,j}^2}\left\{\sum_{k=m+1}^{d} z_{i,k}^2\right\}\left\{\sum_{k=m+1}^{d} \hat{u}_{k,j}^2\right\} \\ &\le \frac{\lambda_{m+1}}{z_{i,j}^2}\frac{\lambda_j}{\hat{\lambda}_j}\frac{d^2}{n\lambda_j^2}\left\{\frac{1}{d-m}\sum_{k=m+1}^{d} z_{i,k}^2\right\}\left\{\frac{n\lambda_j}{d}\sum_{k=m+1}^{d} \hat{u}_{k,j}^2\right\}. \end{aligned} \tag{S6.30}$$

Since $\sum_{k=m+1}^{d} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}}(\frac{d}{n\lambda_j})$ from Shen et al. (2012), it follows that almost surely $\frac{n\lambda_j}{d}\sum_{k=m+1}^{d} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}}(1)$. In addition, note that $\lambda_{m+1} \sim 1$, $\frac{\lambda_j}{\hat{\lambda}_j} \xrightarrow{a.s} 1$, $\frac{d^2}{n\lambda_j^2} \to 0$ and $\frac{1}{d-m}\sum_{k=m+1}^{d} z_{i,k}^2 \xrightarrow{a.s} 1$. Then it follows from (S6.30) that (S6.26) is established for $k = 3$.

S6.4 Proofs of Lemmas

We now prove the lemmas that are used to prove the Theorem 5.2 in the grow sample size asymptomatic contexts. Sections S6.4.1- S6.4.3 respectively provide the proofs of Lemmas S6.1, S6.2 and 7.2.

S6.4.1 Proof of Lemma S6.1

A in (S6.12) can be written as following:

$$A = \frac{1}{n}[\lambda_1^{\frac{1}{2}}\widetilde{Z}_1, \cdots, \lambda_m^{\frac{1}{2}}\widetilde{Z}_m][\lambda_1^{\frac{1}{2}}\widetilde{Z}_1, \cdots, \lambda_m^{\frac{1}{2}}\widetilde{Z}_m]^T,$$

whose $m \times m$ dual matrix A_D is

$$\begin{aligned} A_D &= \frac{1}{n}[\lambda_1^{\frac{1}{2}}\widetilde{Z}_1, \cdots, \lambda_m^{\frac{1}{2}}\widetilde{Z}_m]^T[\lambda_1^{\frac{1}{2}}\widetilde{Z}_1, \cdots, \lambda_m^{\frac{1}{2}}\widetilde{Z}_m] \\ &= \lambda_m \begin{pmatrix} \frac{\lambda_1}{\lambda_m}\frac{1}{n}\sum_{i=1}^n z_{i,1}^2 & \cdots & (\frac{\lambda_1}{\lambda_m})^{\frac{1}{2}}\frac{1}{n}\sum_{i=1}^n z_{i,1}z_{i,m} \\ \vdots & \ddots & \vdots \\ (\frac{\lambda_1}{\lambda_m})^{\frac{1}{2}}\frac{1}{n}\sum_{i=1}^n z_{i,1}z_{i,m} & \cdots & \frac{1}{n}\sum_{i=1}^n z_{i,m}^2 \end{pmatrix}. \end{aligned} \tag{S6.31}$$

Since $z_{i,j}$ are i.i.d and have zero mean, unit variance, then it follows that as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} z_{i,k}z_{i,l} \xrightarrow{\mathrm{a.s}} \begin{cases} 1 & 1 \le k = l \le m \\ 0 & 1 \le k \ne l \le m \end{cases}. \tag{S6.32}$$

In addition, according to Assumptions $\mathcal{B}1$-$\mathcal{B}3$, as $n, d \to \infty$,

$$\frac{\lambda_i}{\lambda_j} \to \frac{c_l}{c_k}, \quad i \in H_k,\ j \in H_l,\ 1 \le k, l \le m, \tag{S6.33}$$

which, together with (S6.31) and (S6.32), yields that as $n, d \to \infty$

$$\frac{A_D}{\lambda_m} \xrightarrow{\mathrm{a.s}} \begin{pmatrix} \frac{c_r}{c_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}. \tag{S6.34}$$

Combining (S6.33) and (S6.34), we have that as $n, d \to \infty$

$$\frac{\lambda_k(A)}{\lambda_k} = \frac{\lambda_k(A_D)}{\lambda_k} \xrightarrow{a.s} 1, \quad k = 1, \cdots, m,$$

which yields Lemma S6.1.

S6.4.2 Proof of Lemma S6.2

Note that for $k = 1, \cdots, [n \wedge (d-m)]$,

$$\frac{d\lambda_d}{n} \times \lambda_k(\frac{1}{d}\sum_{j=m+1}^{d} \widetilde{Z}_j \widetilde{Z}_j^T) \leq \lambda_k(B) \leq \frac{d\lambda_{m+1}}{n} \times \lambda_k(\frac{1}{d}\sum_{j=m+1}^{d} \widetilde{Z}_j \widetilde{Z}_j^T). \tag{S6.35}$$

Since $\frac{n}{d} \to 0$, then it follows from Lemma 7.1 (Bai-Yin's law) that

$$\lambda_k(\frac{1}{d}\sum_{j=m+1}^{d} \widetilde{Z}_j \widetilde{Z}_j^T) \xrightarrow{\text{a.s}} 1, \quad k = 1, \cdots, [n \wedge (d-m)],$$

which, together with (S6.35) and $\lambda_{m+1} \to \cdots \to \lambda_d$, yields Lemma S6.2.

S6.4.3 Proof of Lemma 7.2

We now show the detailed proof of Lemma 7.2 in the main paper, which is used to prove the asymptotic properties of sample eigenvectors in Theorem 5.2. Note that the proof of Lemma 7.2 depends on the asymptotic properties of the sample eigenvalues, which have been proved in Section S6.1.

Firstly, we prove (7.14) in Lemma 7.2. According to Assumptions $\mathcal{B}1$, $\mathcal{B}3$ and sample eigenvalues properties (5.6), we have that for $k \in H_h,\ h = 1, \cdots, r$,

$$\lambda_k^{-1}\hat{\lambda}_j \xrightarrow{a.s} (1 + c_l)c_h c_l^{-1}, \quad j \in H_l,\ l \leq r. \tag{S6.36}$$

Since $\lambda_j(A) = 0$ for $j > m$ and (S6.13), then we have

$$\max_{m+1 \leq j \leq n} |\frac{n}{d}\hat{\lambda}_j - 1| \leq |\frac{n}{d}\lambda_n(B) - 1| + |\frac{n}{d}\lambda_1(B) - 1|. \tag{S6.37}$$

Since $\frac{n}{d} \to 0$, then it follows from Lemma 7.1 that

$$|\frac{n}{d}\lambda_n(B) - 1| \quad \text{and} \quad |\frac{n}{d}\lambda_1(B) - 1| \quad \xrightarrow{a.s} 0,$$

which, together with (S6.37), Assumptions $\mathcal{B}1$, $\mathcal{B}3$ and $[n \wedge d] = n$, yields that for $k \in H_h,\ h = 1, \cdots, r$,

$$\max_{m+1 \leq j \leq n} |\lambda_k^{-1}\hat{\lambda}_j - c_h| \xrightarrow{a.s} 0. \tag{S6.38}$$

Combining (S6.36), (S6.38), (7.5) and $\frac{1}{n}\sum_{i=1}^{n} z_{i,k}^2 \xrightarrow{a.s} 1$, we have that for $k \in H_h$, $h = 1, \cdots, r$,

$$\sum_{l=1}^{r}(1 + c_l)c_h c_l^{-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 + c_h \sum_{j=m+1}^{[n \wedge d]} \hat{u}_{k,j}^2 \xrightarrow{a.s} 1. \tag{S6.39}$$

According to (7.8) and $\sum_{k=1}^{d} \hat{u}_{k,j}^2 = 1$, we have that $\sum_{k=1}^{m} \hat{u}_{k,j}^2 \xrightarrow{a.s} 0$ for $j = m+1, \cdots, [n \wedge d]$. Then it follows that for $k \in H_h$, $h = 1, \cdots, r$,

$$\sum_{j=m+1}^{[n\wedge d]} \hat{u}_{k,j}^2 \xrightarrow{a.s} 0. \tag{S6.40}$$

According to (S6.39) and (S6.40), we have (7.14) in Lemma 7.2.

Secondly, we prove (7.15) in Lemma 7.2. According to (7.9), we have that for $j \in H_l,\ l = 1, \cdots, r$,

$$\frac{\lambda_d}{\hat{\lambda}_j} \lambda_{\min}(\frac{1}{n} Z Z^T) \leq \sum_{k=1}^{d} \lambda_d \lambda_k^{-1} \hat{u}_{k,j}^2 \leq \frac{\lambda_d}{\hat{\lambda}_j} \lambda \max(\frac{1}{n} Z Z^T). \tag{S6.41}$$

According to (5.6), Assumptions $\mathcal{B}1$, $\mathcal{B}3$ and Lemma 7.1, we have that for $j \in H_l,\ l = 1, \cdots, r$,

$$\frac{\lambda_d}{\hat{\lambda}_j} \lambda_{\min}(\frac{1}{n} Z Z^T) \quad \text{and} \quad \frac{\lambda_d}{\hat{\lambda}_j} \lambda \max(\frac{1}{n} Z Z^T) \quad \xrightarrow{a.s} \frac{c_l}{1 + c_l}. \tag{S6.42}$$

Combining (S6.41), (S6.42) and the fact that $\lambda_d \lambda_k^{-1} \to 0$ for $k \leq m$ and 1 for $k > m$, we have that

$$\sum_{k=m+1}^{d} \hat{u}_{k,j}^2 \xrightarrow{a.s} \frac{c_l}{1 + c_l}, \quad j \in H_l,\ l = 1, \cdots, r,$$

which further yields

$$\sum_{h=1}^{r} \sum_{k \in H_h} \hat{u}_{k,j}^2 = 1 - \sum_{k=m+1}^{d} \hat{u}_{k,j}^2 \xrightarrow{a.s} \frac{1}{1 + c_l}, \quad j \in H_l,\ l = 1, \cdots, r,$$

Thus (7.15) in Lemma 7.2 is established.

Finally, we prove (7.16) in Lemma 7.2. Define the diagonal matrix $\Lambda^* = \mathrm{diag}(\lambda_1^*, \cdots, \lambda_d^*)$, where $\lambda_j^* = 1$ for $j \in H_l$, $l = 1$ or $r+1$ and $\lambda_j^* = c_1^{-1} c_l$ for $l = 2, \cdots, r$. Let $Z^* = Z(\Lambda^*)^{\frac{1}{2}}$ and $W^* = W(\Lambda^*)^{\frac{1}{2}}$. Consider the k-diagonal entries $W^*(W^*)^T = \frac{1}{n}(Z^*)^T Z^*$ on two sides, we have

$$\sum_{j=1}^{d} \lambda_k^{-1} \lambda_j^* \hat{\lambda}_j \hat{u}_{k,j}^2 = \frac{1}{n} \sum_{i=1}^{n} (Z_{i,k}^*)^2 = \frac{1}{n} \sum_{i=1}^{n} \lambda_k^* Z_{i,k}^2. \tag{S6.43}$$

According to (S6.36) and (S6.38), we have that, for $k \in H_1$,

$$\begin{cases} \lambda_k^{-1} \lambda_j^* \hat{\lambda}_j \xrightarrow{a.s} (1 + c_l), \quad j \in H_l,\ l \leq r, \\ \max_{m+1 \leq j \leq [n\wedge d]} |\lambda_k^{-1} \lambda_j^* \hat{\lambda}_j - c_1| \xrightarrow{a.s} 0, \quad m+1 \leq j \leq [n \wedge d]. \end{cases} \tag{S6.44}$$

Since $\lambda_k^* = 1$ for $k \in H_1$, then $\frac{1}{n} \sum_{i=1}^{n} \lambda_k^* z_{i,k}^2 = \frac{1}{n} \sum_{i=1}^{n} z_{i,k}^2 \xrightarrow{a.s} 1$ for $k \in H_1$, then it follows from (S6.43) and (S6.44) that

$$\sum_{l=1}^{r} (1 + c_l) \sum_{j \in H_l} \hat{u}_{k,j}^2 + c_1 \sum_{j=m+1}^{[n\wedge d]} \hat{u}_{k,j}^2 \xrightarrow{a.s} 1, \quad k \in H_1. \tag{S6.45}$$

Combining (S6.40) and (S6.45), we have (7.16) in Lemma 7.2.

S7 Proofs for the HDLSS asymptotic results

This section contains the detailed proofs of the theorems, propositions and lemmas under the HDLSS contexts. Section S7.1 provides the proof of Theorem 5.3. Sections S7.2- S7.5 present the proofs of Propositions S5.1-S5.4. Section S7.6 shows the Proof of Theorem 4.1. Section S7.7 provides the proofs of the lemmas.

S7.1 Proof of Theorem 5.3

Proof of Theorem 5.3 contains two parts. The first one is to show the sample eigenvalue properties in Section S7.1.1. The second one is to show the sample eigenvector properties in Section S7.1.2.

S7.1.1 Asymptotic properties of sample eigenvalues in Theorem 5.3

Similar as in Section S6.1, the sample eigenvalues are also studied through the dual matrix $\hat{\Sigma}_D$, which are partitioned as A and B in (S6.12). Now we give out the asymptotic properties of the eigenvalues of A and B.

Lemma S7.1. *For fixed n, as $d \to \infty$, the eigenvalues of the matrix A in* (S6.12) *satisfy*

$$\frac{\lambda_k(A)}{\lambda_k} \xrightarrow{\text{a.s}} \frac{c_k}{n}\lambda_k(\mathcal{W}), \quad k = 1, \cdots, m,$$

where $\mathcal{W}$ is defined in (5.8).

The detailed proof of Lemma S7.1 is in Section S7.7.1.

The HDLSS context doesn't change the asymptotic properties of the eigenvalues of B, shown in Lemma S6.2, with the only difference being that $n \to \infty$ is replaced by $d \to \infty$ and $[n \wedge (d-m)] = n$. For completeness, we restate the asymptotic properties of eigenvalues of B below.

Lemma S7.2. *For fixed n, as $d \to \infty$, the eigenvalues of the matrix B in* (S6.12) *satisfy*

$$\frac{n\lambda_k(B)}{d\lambda_{k+m}} \xrightarrow{\text{a.s}} 1, \quad k = 1, \cdots, n.$$

According to Lemma S7.2 and Assumption $\mathcal{C}2$, we have

$$\frac{\lambda_1(B)}{\lambda_j} \quad \text{and} \quad \frac{\lambda_n(B)}{\lambda_j} \quad \xrightarrow{a.s} c_j, \qquad j = 1, \cdots, m,$$

which, together with (S6.13) and Lemma S7.1, yields

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{a.s} \frac{c_j}{n}\lambda_j(\mathcal{W}) + c_j, \quad j = 1, \cdots, m. \tag{S7.46}$$

In addition, since rank$(A) \leq m$, then it follows from (S6.13) and S7.2 that

$$\frac{n\hat{\lambda}_j}{d\lambda_j} \xrightarrow{a.s} 1, \quad j = m+1, \cdots, n. \tag{S7.47}$$

Combining (S7.46) and (S7.47), we have proved the asymptotic properties of the sample eigenvalues, as stated in (5.10) of Theorem 5.3.

S7.1.2 Asymptotic properties of the sample eigenvectors in Theorem 5.3

Similar as in Section 7.1 of the main paper, the proof procedure is also separated into two steps.

Step 1 is to show the asymptotic properties of $\hat{u}_j$ for $j = m+1, \cdots, n$, which contains two small steps:

- The first one is to show that

$$|< \hat{u}_j, u_j >|^2 = \hat{u}_{j,j}^2 = \mathrm{O}_{\mathrm{a.s}}(\frac{1}{d}), \quad j = m+1, \cdots, n, \tag{S7.48}$$

  which means that the angle between $\hat{u}_j$ and u_j converges to 90 degrees.

- The second one is to show that

$$\text{angle} < \hat{u}_j, \mathbb{S}_2 > \xrightarrow{a.s} 0, \quad j = m+1, \cdots, n, \tag{S7.49}$$

  where $\mathbb{S}_2$ is defined in front of Theorem 5.3.

According to (7.7), we have that

$$\max_{m+1 \leq j \leq n} |< \hat{u}_j, u_j >|^2 = \max_{m+1 \leq j \leq n} \hat{u}_{j,j}^2 \leq \frac{\lambda_{m+1}}{\hat{\lambda}_n} \lambda_{\max}(\frac{1}{n}\bar{Z}^T\bar{Z}), \tag{S7.50}$$

Since n is a fixed number, then $\lambda_{\max}(\frac{1}{n}\bar{Z}^T\bar{Z})$ is a random variable. Then it follows from (S7.47) and (S7.50) that (S7.48) is established. Proof of (S7.49) is the same as (7.4) and is skipped here.

Step 2 is to show the asymptotic properties of $\hat{u}_j$ such that the angle between $\hat{u}_j$ and $\mathbb{S}_1$ in (5.10) almost surely converges to $\arccos\left\{\left(1 + \frac{n}{\lambda_j(\mathcal{W})}\right)^{-\frac{1}{2}}\right\}$, $j = 1, \cdots, m$. According to (7.2), we just need to show that

$$\sum_{k=1}^{m} \hat{u}_{k,j}^2 \xrightarrow{a.s} \frac{\lambda_j(\mathcal{W})}{n + \lambda_j(\mathcal{W})}, \quad j = 1, \cdots, m. \tag{S7.51}$$

According to (S7.46) and Lemma 7.1, we have that for $j = 1, \cdots, m$,

$$\frac{\lambda_d}{\hat{\lambda}_j}\lambda_{\min}(\frac{1}{n}Z^TZ) \quad \text{and} \quad \frac{\lambda_d}{\hat{\lambda}_j}\lambda_{\max}(\frac{1}{n}Z^TZ) \quad \xrightarrow{a.s} \frac{n}{n + \lambda_j(\mathcal{W})},$$

which, together with (S6.41) and the fact that $\lambda_d\lambda_k^{-1} \to 0$ for $k \leq m$ and 1 for $k > m$, yields

$$\sum_{k=m+1}^{d} \hat{u}_{k,j}^2 \xrightarrow{a.s} \frac{n}{n + \lambda_j(\mathcal{W})}, \quad j = 1, \cdots, m. \tag{S7.52}$$

Since $\sum_{k=1}^{d} \hat{u}_{k,j}^2 = 1$, then it follows from (S7.52) that (S7.51) is established.

According to (S7.48), (S7.49) and (S7.51), we have proved the asymptotic properties of the sample eigenvectors as stated in (5.10) of Theorem 5.3.

S7.2 Proof of Proposition S5.1

The sample eigenvalue properties are also studied through the dual matrix $\hat{\Sigma}_D$, which are partitioned into three parts as $\hat{\Sigma}_D = A_1 + A_2 + B$, where

$$A_1 = \frac{1}{n}\sum_{j=1}^{m_0} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T, \quad A_2 = \frac{1}{n}\sum_{j=m_0+1}^{m} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T, \quad \text{and} \quad B = \frac{1}{n}\sum_{j=m+1}^{d} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T. \tag{S7.53}$$

The eigenvalue properties of A_1 and A_2 are similar as A in Lemma S7.1, and are shown in the following lemma.

Lemma S7.3. *For fixed n, as $d \to \infty$, the eigenvalues of the matrices A_1 and A_2 in* (S7.53) *satisfy*

$$\begin{cases} \frac{\lambda_k(A_1)}{\lambda_k} \xrightarrow{\text{a.s}} \frac{c_k^*}{n}\lambda_k(\mathcal{W}_1^1), & 1 \le k \le m_0, \\ \frac{\lambda_k(A_2)}{\lambda_{k+m_0}} \xrightarrow{\text{a.s}} \frac{c_{k+m_0}}{n}\lambda_k(\mathcal{W}_2^1), & 1 \le k \le m - m_0, \end{cases}$$

where c_k^* are defined in (S5.4)*, and $\mathcal{W}_1^1$ and $\mathcal{W}_2^1$ are defined in* (S5.5).

The properties of the eigenvalues of B in (S7.53) remain the same as B in Lemma S7.2. According to Lemmas S6.3, S7.2 and S7.3 and Assumption $\mathcal{C}3$, the sample eigenvalues satisfy

$$\begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} \frac{c_j^*}{n}\lambda_j(\mathcal{W}_1^1), & 1 \le j \le m_0, \\ \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} \frac{c_j}{n}\lambda_{j-m_0}(\mathcal{W}_2^1) + c_j, & m_0 + 1 \le j \le m, \end{cases}$$

which yields (S5.6) in Proposition S5.1.

In addition, following the proof procedure of (S7.51), we have

$$\begin{cases} \sum_{k=1}^{m_0} \hat{u}_{k,j}^2 \xrightarrow{a.s} 1, & 1 \le j \le m_0, \\ \sum_{k=m_0+1}^{m} \hat{u}_{k,j}^2 \xrightarrow{a.s} \frac{\lambda_{j-m_0}(\mathcal{W}_2^1)}{n+\lambda_{j-m_0}(\mathcal{W}_2^1)}, & m_0 + 1 \le j \le m, \end{cases} \tag{S7.54}$$

which yields

$$\begin{cases} \text{angle} < \hat{u}_j, \mathbb{S}_1^1 > \xrightarrow{a.s} 1, & 1 \le j \le m_0, \\ \text{angle} < \hat{u}_j, \mathbb{S}_2^1 > \xrightarrow{a.s} \arccos\left\{\left(1 + \frac{n}{\lambda_{j-m_0}(\mathcal{W}_2^1)}\right)^{-\frac{1}{2}}\right\}, & m_0 + 1 \le j \le m. \end{cases}$$

It then follows that (S5.7) in Proposition S5.1 is established, which concludes the proof of Proposition S5.1.

S7.3 Proof of Proposition S5.2

The proof is again through the dual matrix $\hat{\Sigma}_D = A_1 + A_2 + B$, where the eigenvalue properties of A_2 and B remain the same as in Section S7.2. However, under the assumptions in Proposition S5.2, A_1 has different properties such that for fixed n, as $d \to \infty$,

$$\frac{\lambda_k(A_1)}{\lambda_k} \xrightarrow{\text{a.s}} R_k, \quad 1 \leq k \leq m_0. \tag{S7.55}$$

Since A_2 and B have the properties as in Section S7.2, it follows from (S7.55), Assumption $\mathcal{C}5$ and Lemma S6.3 that the sample eigenvalues satisfy

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} R_j, \quad 1 \leq j \leq m_0, \tag{S7.56}$$

which yields (S5.8) in Proposition S5.2.

Since (S7.56) shows that the first m_0 sample eigenvalues can be asymptotically distinguished, then their corresponding sample eigenvectors can be distinguished and are consistent with the corresponding population eigenvectors. In fact, similar as (S7.51) and (S7.54), we have

$$|< \hat{u}_j, u_j >|^2 = \hat{u}_{j,j}^2 \xrightarrow{a.s} 1, \quad 1 \leq j \leq m_0, \tag{S7.57}$$

which yields (S5.9) in Proposition S5.2.

S7.4 Proof of Proposition S5.3

The critical difference between the proofs of Proposition S5.3 and Theorem 5.3 is that the dual matrix $\hat{\Sigma}_D$ should be partitioned into the following two parts:

$$A = \frac{1}{n}\sum_{j=1}^{m_0} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T, \quad B = \frac{1}{n}\sum_{j=m_0+1}^{d} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T. \tag{S7.58}$$

Then following the proof procedure of (S7.46), we have

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{a.s} \frac{c_j}{n}\lambda_j(\mathcal{W}^2) + c_j, \quad j = 1, \cdots, m_0. \tag{S7.59}$$

Since $\lambda_{m+1} \to \cdots \to \lambda_d = 1$, then similar as (S7.47), we have

$$\frac{n\hat{\lambda}_j}{d\lambda_d} \xrightarrow{a.s} 1, \quad j = m_0 + 1, \cdots, n. \tag{S7.60}$$

According to (S7.59) and (S7.60), we have (S5.10) in Proposition S5.3.

In addition, following the proof procedure of (5.10) in Theorem 5.3, we have (S5.11) in Proposition S5.3.

S7.5 Proof of Proposition S5.4

The difference between the proofs of Propositions S5.4 and S5.3 is that A in (S7.58) has the following properties that for fixed n, as $d \to \infty$,

$$\frac{\lambda_k(A)}{\lambda_k} \xrightarrow{\text{a.s}} R_j, \quad 1 \le k \le m_0,$$

which yields

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} R_j, \quad 1 \le j \le m_0.$$

Thus the sample eigenvalues and eigenvectors whose index is less than or equal to m_0 have the same properties as in Proposition S5.2. Furthermore, the properties of the other sample eigenvalues and eigenvectors remain the same as in Proposition S5.3.

S7.6 Proof of Theorem 4.1

To prove Theorem 4.1, it follows from definition of Γ_k, $k = 1, 2, 3$ in (S6.24) that we just need to show that for $i = 1, \cdots, n$, $j = 1, \cdots, m$,

$$|\Gamma_1| \xrightarrow{a.s} R_j^{-\frac{1}{2}}, \tag{S7.61}$$

$$|\Gamma_k| \xrightarrow{a.s} 0, \quad k = 2, 3. \tag{S7.62}$$

Under spike model (4.6), (S7.56) and (S7.57) respectively become

$$\begin{aligned} \frac{\hat{\lambda}_j}{\lambda_j} &\overset{\text{a.s}}{\to} R_j, \quad j = 1, \cdots, m, \\ \hat{u}_{j,j}^2 &\overset{\text{a.s}}{\to} 1, \quad j = 1, \cdots, m, \end{aligned}$$

which yields (S7.61). In addition, following the proof procedure of (S6.26), we have (S7.62). According to (S6.24), (S7.61) and (S7.62), we have

$$\left| \frac{\hat{S}_{i,j}}{S_{i,j}} \right| \xrightarrow{a.s} R_j^{-\frac{1}{2}}, \quad i = 1, \cdots, n, \ j = 1, \cdots, m,$$

which yields (4.9) in Theorem 5.1.

S7.7 Proofs of the Lemmas

This section contains the detailed proofs of the necessary lemmas under the HDLSS contexts. Section S7.7.1 provides the proof of Lemma S7.1. We skip the proofs for Lemmas S7.2 and S7.3, as they are essentially the same as Lemmas S6.2 and S7.1, respectively.

S7.7.1 Proof of Lemma S7.1

The eigenvalue properties of A in Lemma S7.1 are also studied through its dual matrix A_D in (S6.31). According to Assumption $\mathcal{C}2$, we have

$$\frac{n\lambda_k}{d\lambda_d} \to c_k^{-1}, \quad k = 1, \cdots, m, \tag{S7.63}$$

which, together with (5.8) and (S6.31), yields that for fixed n, as $d \to \infty$,

$$\frac{n^2 A_D}{d\lambda_d} \xrightarrow{a.s} \mathcal{W}. \tag{S7.64}$$

Thus it follows from (S7.63) and (S7.64) that

$$\frac{\lambda_k(A)}{\lambda_k} = \frac{\lambda_k(A_D)}{\lambda_k} \xrightarrow{a.s} \frac{c_k}{n}\lambda_k(\mathcal{W}), \quad k = 1, \cdots, m,$$

which yields Lemma S7.1.

References

Aoshima, M. and Yata, K. (2015) Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. *Methodology and Computing in Applied Probability 17*(2), 419–439.

Cabanski, C., Qi, Y., Yin, X., Bair, E., Hayward, M., Fan, C., Li, J., Wilkerson, M., Marron, J. S., Perou, C. and Hayes, D. (2010) Swiss made: standardized within class sum of squares to evaluate methodologies and dataset elements. *PloS One 5*(3), e9905.

Jung, S. and Marron, J. S. (2009) PCA consistency in high dimension, low sample size context. *The Annals of Statistics 37*(6B), 4104–4130.

Rao, C. (2002) *Linear statistical inference and its applications.* John Willey & Sons, New York.

Shen, D., Shen, H. and Marron, J. S. (2012) A general framework for consistency of principal component analysis. *arXiv preprint arXiv:1211.2671*.

Yata, K. and Aoshima, M. (2012) Inference on high-dimensional mean vectors with fewer observations than the dimension. *Methodology and Computing in Applied Probability 14*(3), 459–476.