Submitted to the Annals of Statistics

PCA CONSISTENCY IN HIGH DIMENSION, LOW SAMPLE SIZE CONTEXT

BY SUNGKYU JUNG AND J. S. MARRON

University of North Carolina at Chapel Hill

Principal Component Analysis (PCA) is an important tool of dimension reduction especially when the dimension (or the number of variables) is very high. Asymptotic studies where the sample size is fixed, and the dimension grows (i.e. High Dimension, Low Sample Size (HDLSS)) are becoming increasingly relevant. We investigate the asymptotic behavior of the Principal Component (PC) directions. HDLSS asymptotics are used to study consistency, strong inconsistency and subspace consistency. We show that if the first few eigenvalues of a population covariance matrix are large enough compared to the others, then the corresponding estimated PC directions are consistent or converge to the appropriate subspace (subspace consistency) and most other PC directions are strongly inconsistent. Broad sets of sufficient conditions for each of these cases are specified and the main theorem gives a catalogue of possible combinations. In preparation for these results, we show that the geometric representation of HDLSS data holds under general conditions, which includes a ρ-mixing condition and a broad range of sphericity measures of the covariance matrix.

1. Introduction and summary. The High Dimension, Low Sample Size (HDLSS) data situation occurs in many areas of modern science and the asymptotic studies of this type of data are becoming increasingly relevant. We will focus on the case that the dimension d increases while the sample size n is fixed as done in Hall et al. [7] and Ahn et al. [1]. The d-dimensional covariance matrix is challenging to analyze in general since the number of parameters is $\frac{d(d+1)}{2}$, which increases even faster than d. Instead of assessing all of the parameter estimates, the covariance matrix is usually analyzed by Principal Component Analysis (PCA). PCA is often used to visualize important structure in the data and also used to reduce dimensionality by approximating the data with the first few principal components. Thus the asymptotic behavior of Principal Component (PC) directions is important. In this paper, we study the covariance matrix in terms of PC directions.

AMS 2000 subject classifications: Primary 62H25, 34L20; secondary 62F12

Keywords and phrases: Principal Component Analysis, Sample Covariance matrix, ρ-mixing, High dimension, low sample size data, Non-standard asymptotics, Consistency and strong inconsistency, Spiked population model

imsart-aos ver. 2007/12/10 file: PCAconsistency.tex date: June 26, 2008

Our focus is on the underlying mechanism which determines when the sample PC directions converge to their population counterparts as $d \to \infty$. In general we assume $d > n$. Since the size of the covariance matrix depends on d, the population covariance matrix is denoted as Σ_d and similarly the sample covariance matrix, S_d, so that their dependency on the dimension is emphasized. PCA is done by eigen-decomposition of a covariance matrix. The eigen-decomposition of Σ_d is

$$\Sigma_d = U_d \Lambda_d U_d',$$

where Λ_d is a diagonal matrix of eigenvalues $\lambda_{1,d} \geq \lambda_{2,d} \geq \cdots \geq \lambda_{d,d}$ and U_d is a matrix of corresponding eigenvectors so that $U_d = [u_{1,d}, u_{2,d}, \ldots, u_{d,d}]$. S_d is similarly decomposed as

$$S_d = \hat{U}_d \hat{\Lambda}_d \hat{U}_d'.$$

Ahn et al. [1] showed that HDLSS consistency is a very illuminating asymptotic property. Our main results are formulated in terms of three related concepts:

1. *consistency*: The direction $\hat{u}_{i,d}$ is *consistent* with its population counterpart $u_{i,d}$ if $\text{Angle}(u_{i,d}, \hat{u}_{i,d}) \longrightarrow 0$ as $d \to \infty$. The growth of dimension can be understood as adding more variation. The consistency of sample eigenvectors occurs when the added variation supports the existing structure in the covariance or is small enough to be ignored.
2. *strong inconsistency*: In situations where $\hat{u}_{i,d}$ is not consistent, a perhaps counter-intuitive HDLSS phenomenon frequently occurs. In particular, $\hat{u}_{i,d}$ is said to be *strongly inconsistent* with its population counterpart $u_{i,d}$ in the sense that it tends to be as far away from $u_{i,d}$ as possible, that is, $\text{Angle}(u_{i,d}, \hat{u}_{i,d}) \longrightarrow \frac{\pi}{2}$ as $d \to \infty$. Strong inconsistency occurs when the added variation obscures the underlying structure of the population covariance matrix.
3. *subspace consistency*: When several population eigenvalues indexed by $j \in J$ are similar, the corresponding sample eigenvectors may not be distinguishable. In this case, $\hat{u}_{j,d}$ will not be consistent for $u_{j,d}$ but will tend to lie in the linear span, $\text{span}\{u_{j,d} : j \in J\}$. This motivates the definition of convergence of a direction $\hat{u}_{i,d}$ to a subspace, called *subspace consistency*;

   $$\text{Angle}(\hat{u}_{i,d}, \text{span}\{u_{j,d} : j \in J\}) \longrightarrow 0$$

   as $d \to \infty$. This definition essentially comes from the theory of *canonical angles* discussed by Gaydos [6]. That theory also gives a notion of convergence of subspaces, that could be developed here.

In recent years, substantial work has been done on the asymptotic behavior of eigenvalues of the sample covariance matrix in the limit as $d \to \infty$, see Baik et al. [2], Johnstone [10] and Paul [13] for Gaussian assumptions and Baik and Silverstein [3] for non-Gaussian results when d and n increase at the same rate, i.e. $\frac{n}{d} \to c > 0$. Many of these focus on the *spiked covariance model*, introduced by Johnstone [10]. The spiked covariance model assumes that the first few eigenvalues of the population covariance matrix are greater than 1 and the rest are set to be 1 for all d. HDLSS asymptotics, where only $d \to \infty$ while n is fixed, have been studied by Hall et al. [7] and Ahn et al. [1]. They explored conditions which give the *geometric representation* of HDLSS data (i.e. modulo rotation, data tend to lie at vertices of a regular simplex.) as well as strong inconsistency of eigenvectors. Strong inconsistency is also found in the context of $\frac{n}{d} \to c$, in the study of *phase transition*, see e.g. Paul [13], Johnstone and Lu [11], and Baik et al. [2].

In this paper, a broad and general set of conditions for consistency and strong inconsistency are provided. Section 2 develops conditions that guarantee the non-zero eigenvalues of the sample covariance matrix tend to a increasing constant, which are much more general than those of Hall et al. [7] and Ahn et al. [1]. This asymptotic behavior of the sample covariance matrix is the basis of the geometric representation of HDLSS data. Our result gives broad new insight into this representation as discussed in section 3. The central issue of consistency and strong inconsistency is developed in section 4, as a series of theorems. For a fixed number κ, we assume the first κ eigenvalues are much larger than the others. We show that when $\kappa = 1$, the first sample eigenvector is consistent and the others are strongly inconsistent. We also generalize to the $\kappa > 1$ case, featuring two different types of results (consistency and subspace consistency) according to the asymptotic behaviors of the first κ eigenvalues. All results are combined and generalized in the main theorem (Theorem 2). Proofs of theorems are given in section 5.

1.1. *General setting.* Suppose we have a $d \times n$ data matrix $X_{(d)} = [X_{1,(d)}, \ldots, X_{n,(d)}]$ with $d > n$, where the d-dimensional random vectors $X_{1,(d)}, \ldots, X_{n,(d)}$ are independent and identically distributed. We assume that each $X_{i,(d)}$ follows a multivariate distribution (which does not have to be Gaussian) with mean zero and covariance matrix Σ_d. Define the sphered data matrix $Z_{(d)} = \Lambda_d^{-\frac{1}{2}} U_d' X_{(d)}$. Then the components of the $d \times n$ matrix $Z_{(d)}$ have univariate variances, and are uncorrelated with each other. We shall regulate the dependency (recall for non-Gaussian data, uncorrelated variables can still be dependent) of the random variables in $Z_{(d)}$ by a ρ-mixing condition. This allows serious weakening of the assumptions of

Gaussianity while still enabling the law of large numbers that lie behind the geometric representation results of Hall et al. [7].

The concept of ρ-mixing was first developed by Kolmogorov and Rozanov [12]. See Bradley [4] for a clear and insightful discussion. For $-\infty \le J \le L \le \infty$, let $\mathcal{F}_J^L$ denote the σ-field of events generated by the random variables $(Z_i,\ J \le i \le L)$. For any σ-field $\mathcal{A}$, let $L_2(\mathcal{A})$ denote the space of square-integrable, $\mathcal{A}$ measurable (real-valued) random variables. For each $m \ge 1$, define the maximal correlation coefficient

$$\rho(m) := \sup |\mathrm{corr}(f,g)|, \quad f \in L_2(\mathcal{F}_{-\infty}^j),\ g \in L_2(\mathcal{F}_{j+m}^\infty),$$

where sup is over all f, g and $j \in \mathbf{Z}$. The sequence $\{Z_i\}$ is said to be ρ-mixing if $\rho(m) \to 0$ as $m \to \infty$.

While the concept of ρ-mixing is useful as a mild condition for the development of laws of large numbers, its formulation is critically dependent on the ordering of variables. For many interesting data types, such as microarray data, there is clear dependence but no natural ordering of the variables. Hence we assume that there is some permutation of the data which is ρ-mixing. In particular, let $\{Z_{ij,(d)}\}_{i=1}^d$ be the components of the jth column vector of $Z_{(d)}$. We assume that for each d, there exists a permutation $\pi_d : \{1,\ldots,d\} \longmapsto \{1,\ldots,d\}$ so that the sequence $\{Z_{\pi_d(i)j,(d)} : i = 1,\ldots,d\}$ is ρ-mixing.

In the following, all the quantities depend on d, but the subscript d will be omitted for the sake of simplicity when it does not cause any confusion. The sample covariance matrix is defined as $S = n^{-1}XX'$. We do not subtract the sample mean vector because the population mean is assumed to be 0. Since the dimension of the sample covariance matrix S grows, it is challenging to deal with S directly. A useful approach is to work with the *dual* of S. The dual approach switches the role of columns and rows of the data matrix, by replacing X by X'. The $n \times n$ *dual sample covariance matrix* is defined as $S_D = n^{-1}X'X$. An advantage of this dual approach is that S_D and S share non-zero eigenvalues. If we write X as $U\Lambda^{\frac{1}{2}}Z$ and use the fact that U is a unitary matrix,

$$nS_D = (Z'\Lambda^{\frac{1}{2}}U')(U\Lambda^{\frac{1}{2}}Z) = Z'\Lambda Z = \sum_{i=1}^d \lambda_{i,d} z_i' z_i, \tag{1.1}$$

where the z_i's, $i = 1,\ldots,d$, are the row vectors of the matrix Z.

2. HDLSS asymptotic behavior of the sample covariance matrix.

In this section, we investigate the behavior of the sample covariance matrix S when $d \to \infty$ and n is fixed. Under mild and broad conditions, the eigenvalues of S, or the dual S_D, behave asymptotically as if they are from the identity matrix. That is, the set of sample eigenvectors tends to be an arbitrary choice. This lies at the heart of the geometric representation results of Hall et al. [7] and Ahn et al. [1] which are studied more deeply in section 3. We will see that this condition readily implies the strong inconsistency of sample eigenvectors, see Theorem 2.

The conditions for the theorem are conveniently formulated in terms of a measure of sphericity

$$\epsilon \equiv \frac{\mathrm{tr}^2(\Sigma)}{d\,\mathrm{tr}(\Sigma^2)} = \frac{(\sum_{i=1}^d \lambda_{i,d})^2}{d \sum_{i=1}^d \lambda_{i,d}^2},$$

proposed and used by John [8, 9] as the basis of a hypothesis test for equality of eigenvalues. Note that these inequalities always hold:

$$\frac{1}{d} \le \epsilon \le 1.$$

Also note that perfect sphericity of the distribution (i.e. equality of eigenvalues) occurs only when $\epsilon = 1$. The other end of the ϵ range is the most singular case where in the limit as the first eigenvalue dominates all others.

Ahn et al. [1] claimed that if $\epsilon \gg \frac{1}{d}$, in the sense that $\epsilon^{-1} = o(d)$, then the eigenvalues of S_D tend to be identical in probability as $d \to \infty$. However, they needed an additional assumption (e.g. a Gaussian assumption on $X_{(d)}$) to have independence among components of $Z_{(d)}$, as described in example 3.1. In this paper, we extend this result to the case of arbitrary distributions with dependency regulated by the ρ-mixing condition as in section 1.1, which is much more general than either a Gaussian or an independence assumption. We also explore convergence in the almost sure sense with stronger assumptions. Our results use a measure of sphericity for part of the eigenvalues for conditions of a.s. convergence and also for later use in section 4. In particular, define the measure of sphericity for $\{\lambda_{k,d}, \dots, \lambda_{d,d}\}$ as

$$\epsilon_k \equiv \frac{(\sum_{i=k}^d \lambda_{i,d})^2}{d \sum_{i=k}^d \lambda_{i,d}^2}.$$

For convenience, we name several assumptions used in this paper made about the measure of sphericity ϵ:

- *The ϵ-condition*: $\epsilon \gg \frac{1}{d}$, i.e.

$$(d\epsilon)^{-1} = \frac{\sum_{i=1}^d \lambda_{i,d}^2}{(\sum_{i=1}^d \lambda_{i,d})^2} \to 0 \text{ as } d \to \infty. \tag{2.1}$$

- *The ϵ_k-condition*: $\epsilon_k \gg \frac{1}{d}$, i.e.

$$(d\epsilon_k)^{-1} = \frac{\sum_{i=k}^d \lambda_{i,d}^2}{(\sum_{i=k}^d \lambda_{i,d})^2} \to 0 \text{ as } d \to \infty. \tag{2.2}$$

- *The strong ϵ_k-condition*: For some fixed $l \geq k$, $\epsilon_l \gg \frac{1}{\sqrt{d}}$, i.e.

$$d^{-\frac{1}{2}}\epsilon_l^{-1} = \frac{d^{\frac{1}{2}} \sum_{i=l}^d \lambda_{i,d}^2}{(\sum_{i=l}^d \lambda_{i,d})^2} \to 0 \text{ as } d \to \infty. \tag{2.3}$$

REMARK. Note that the ϵ_k-condition is identical to the ϵ-condition when $k = 1$. Similarly, the strong ϵ_k-condition is also called *the strong ϵ-condition* when $k = 1$. The strong ϵ_k-condition is stronger than the ϵ_k condition if the minimum of l's which satisfy (2.3), l_o, is as small as k. But, if $l_o > k$, then this is not necessarily true. We will use the strong ϵ_k-condition combined with the ϵ_k-condition.

Note that the ϵ-condition is quite broad in the spectrum of possible values of ϵ: It only avoids the most singular case. The strong ϵ-condition further restricts ϵ_l to essentially in the range $(\frac{1}{\sqrt{d}}, 1]$.

The following theorem states that if the (strong) ϵ-condition holds for Σ_d, then the sample eigenvalues behave as if they are from a scaled identity matrix. It uses the notation I_n for the $n \times n$ identity matrix.

THEOREM 1. *For a fixed n, let $\Sigma_d = U_d \Lambda_d U_d'$, $d = n+1, n+2, \ldots$ be a sequence of covariance matrices. Let $X_{(d)}$ be a $d \times n$ data matrix from a d-variate distribution with mean zero and covariance matrix Σ_d. Let $S_d = \hat{U}_d \hat{\Lambda}_d \hat{U}_d'$ be the sample covariance matrix estimated from $X_{(d)}$ for each d and let $S_{D,d}$ be its dual.*

(1) Assume that the components of $Z_{(d)} = \Lambda_d^{-\frac{1}{2}} U_d' X_{(d)}$ have uniformly bounded fourth moments and are ρ-mixing under some permutation. If (2.1) holds, then

$$c_d^{-1} S_{D,d} \longrightarrow I_n, \tag{2.4}$$

in probability as $d \to \infty$, where $c_d = n^{-1} \sum_{i=1}^d \lambda_{i,d}$.

(2) Assume that the components of $Z_{(d)} = \Lambda_d^{-\frac{1}{2}} U_d' X_{(d)}$ have uniformly bounded eighth moments and are independent to each other. If both (2.1) and (2.3) hold, then $c_d^{-1} S_{D,d} \to I_n$ almost surely as $d \to \infty$.

The (strong) ϵ-condition holds for quite general settings. The strong ϵ-condition combined with the ϵ-condition holds under;

(a) Null case: All eigenvalues are the same.
(b) Mild spiked model: The first m eigenvalues are moderately larger than the others, for example, $\lambda_{1,d} = \cdots = \lambda_{m,d} = C_1 \cdot d^\alpha$ and $\lambda_{m+1,d} = \cdots = \lambda_{d,d} = C_2$, where $m < d$, $\alpha < 1$ and $C_1, C_2 > 0$.

The ϵ-condition fails when;

(c) Singular case: Only the first few eigenvalues are non-zero.
(d) Exponential decrease: $\lambda_{i,d} = c^{-i}$ for some $c > 1$.
(e) Sharp spiked model: The first m eigenvalues are much larger than the others. One example is the same as (b) but $\alpha \geq 1$.

The polynomially decreasing case, $\lambda_{i,d} = i^{-\beta}$, is interesting because it depends on the power β;

(f-1) The strong ϵ-condition holds when $0 \leq \beta < \frac{3}{4}$.
(f-2) The ϵ-condition holds but the strong ϵ-condition fails when $\frac{3}{4} \leq \beta \leq 1$.
(f-3) The ϵ-condition fails when $\beta > 1$.

Another family of examples that includes all three cases is the spiked model with the number of spikes increasing, for example, $\lambda_{1,d} = \cdots = \lambda_{m,d} = C_1 \cdot d^\alpha$ and $\lambda_{m+1,d} = \cdots = \lambda_{d,d} = C_2$, where $m = \lfloor d^\beta \rfloor$, $0 < \beta < 1$ and $C_1, C_2 > 0$;

(g-1) The strong ϵ-condition holds when $0 \leq 2\alpha + \beta < \frac{3}{2}$;
(g-2) The ϵ-condition holds but the strong ϵ-condition fails when $\frac{3}{2} \leq 2\alpha + \beta < 2$;
(g-3) The ϵ-condition fails when $2\alpha + \beta \geq 2$.

3. Geometric representation of HDLSS data.

Suppose $X \sim \mathcal{N}_d(0, I_d)$. When the dimension d is small, most of the mass of the data lies near origin. However with a large d, Hall et al. [7] showed that Euclidean distance of X to the origin is described as

$$\|X\| = \sqrt{d} + o_p(\sqrt{d}). \tag{3.1}$$

Moreover the distance between two samples is also rather deterministic, i.e.

$$\|X_1 - X_2\| = \sqrt{2d} + o_p(\sqrt{d}). \tag{3.2}$$

These results can be derived by the law of large numbers. Hall et al. [7] generalized those results under the assumptions that $d^{-1} \sum_{i=1}^{d} \mathrm{Var}(X_i) \to 1$ and $\{X_i\}$ is ρ-mixing.

Application of part (1) of Theorem 1 generalizes these results. Let $X_{1,(d)}$, $X_{2,(d)}$ be two samples that satisfy the assumptions of Theorem 1 part (1). Assume without loss of generality that $\lim_{d\to\infty} d^{-1}\sum_{i=1}^{d}\lambda_{i,d} = 1$. The scaled squared distance between two data points is

$$\frac{\|X_{1,(d)} - X_{2,(d)}\|^2}{\sum_{i=1}^{d}\lambda_{i,d}} = \sum_{i=1}^{d}\tilde{\lambda}_{i,d}z_{i1}^2 + \sum_{i=1}^{d}\tilde{\lambda}_{i,d}z_{i2}^2 + \sum_{i=1}^{d}\tilde{\lambda}_{i,d}z_{i1}z_{i2},$$

where $\tilde{\lambda}_{i,d} = \frac{\lambda_{i,d}}{\sum_{i=1}^{d}\lambda_{i,d}}$. Note that by (1.1), the first two terms are diagonal elements of $c_d^{-1}S_{D,d}$ in Theorem 1 and the third term is an off-diagonal element. Since $c_d^{-1}S_{D,d} \to I_n$, we have (3.2). (3.1) is derived similarly.

REMARK. If $\lim_{d\to\infty} d^{-1}\sum_{i=1}^{d}\lambda_{i,d} = 1$, then the conclusion (2.4) of Theorem 1 part (1) holds if and only if the representations (3.1) and (3.2) hold under the same assumptions in the theorem.

In this representation, the ρ-mixing assumption plays a very important role. The following example, due to John Kent, shows that some type of mixing condition is important.

EXAMPLE 3.1 (Strong dependency via a scale mixture of Gaussian). Let $X = Y_1U + \sigma Y_2(1-U)$, where Y_1, Y_2 are two independent $\mathcal{N}_d(0, I_d)$ random variables, $U = 0$ or 1 with probability $\frac{1}{2}$ and independent of Y_1, Y_2, and $\sigma > 1$. Then,

$$\|X\| = \begin{cases} d^{\frac{1}{2}} + O_p(1) & \text{w.p. } \frac{1}{2} \\ \sigma d^{\frac{1}{2}} + O_p(1) & \text{w.p. } \frac{1}{2} \end{cases}$$

Thus, (3.1) does not hold. Note that since $\mathrm{Cov}(X) = \frac{1+\sigma^2}{2}I_d$, the ϵ-condition holds and the variables are uncorrelated. However, there is strong dependency, i.e. $\mathrm{Cov}(z_i^2, z_j^2) = (\frac{1+\sigma^2}{2})^{-2}\mathrm{Cov}(x_i^2, x_j^2) = (\frac{1-\sigma^2}{1+\sigma^2})^2$ for all $i \neq j$ which implies that $\rho(m) > c$ for some $c > 0$, for all m. Thus, the ρ-mixing condition does not hold for all permutation. Note that, however, under Gaussian assumption, given any covariance matrix Σ, $Z = \Sigma^{-\frac{1}{2}}X$ has independent components.

Note that in the case $X = (X_1, \ldots, X_d)$ is a sequence of i.i.d. random variables, the results (3.1) and (3.2) can be considerably strengthened to $\|X\| = \sqrt{d} + O_p(1)$, and $\|X_1 - X_2\| = \sqrt{2d} + O_p(1)$. The following example shows that strong results are beyond the reach of reasonable assumption.

EXAMPLE 3.2 (Varying sphericity). Let $X \sim \mathcal{N}_d(0, \Sigma_d)$, where $\Sigma_d = \text{diag}(d^\alpha, 1, \ldots, 1)$ and $\alpha \in (0,1)$. Define $Z = \Sigma_d^{-\frac{1}{2}} X$. Then the components of Z, z_i's, are independent standard Gaussian random variables. We get $\|X\|^2 = d^\alpha z_1^2 + \sum_{i=2}^d z_i^2$. Now for $0 < \alpha < \frac{1}{2}$, $d^{-\frac{1}{2}}(\|X\|^2 - d) \Rightarrow \mathcal{N}(0,1)$ and for $\frac{1}{2} < \alpha < 1$, $d^{-\alpha}(\|X\|^2 - d) \Rightarrow z_1^2$, where $\Rightarrow$ denotes convergence in distribution. Thus by the delta-method, we get

$$\|X\| = \begin{cases} \sqrt{d} + O_p(1), & \text{if } 0 < \alpha < \frac{1}{2}, \\ \sqrt{d} + O_p(d^{\alpha - \frac{1}{2}}), & \text{if } \frac{1}{2} < \alpha < 1. \end{cases}$$

In both cases, the representation (3.1) holds.

4. Consistency and strong inconsistency of PC directions. In this section, conditions for consistency or strong inconsistency of the sample PC direction vectors are investigated, in the general setting of section 1.1. The generic eigen-structure of the covariance matrix that we assume is the following. For a fixed number κ, we assume the first κ eigenvalues are much larger than others. (The precise meaning of *large* will be addressed shortly.) The rest of eigenvalues are assumed to satisfy the ϵ-condition, which is very broad in the range of sphericity. We begin with the case $\kappa = 1$ and generalize the result for $\kappa > 1$ in two distinct ways. The main theorem (Theorem 2) contains and combines those previous results and also embraces various cases according to the magnitude of the first κ eigenvalues. We also investigate the sufficient conditions for a stronger result, i.e. almost sure convergence, which involves use of the strong ϵ-condition.

4.1. *Criteria for consistency or strong inconsistency of the first PC direction.* Consider the simplest case that only the first PC direction of S is of interest. Section 3 gives some preliminary indication of this. As an illustration, consider a spiked model as in Example 3.2 but now let $\alpha > 1$. Let $\{u_i\}$ be the set of eigenvectors of Σ_d and V_{d-1} be the subspace of all eigenvectors except the first one. Then the projection of X onto u_1 has a norm $\|\text{Proj}_{u_1} X\| = \|X_1\| = O_p(d^{\frac{\alpha}{2}})$. The projection of X onto V_{d-1} has a norm $\sqrt{d} + o_p(\sqrt{d})$ by (3.1). Thus when $\alpha > 1$, if we scale the whole data space $\mathbf{R}^d$ by dividing by $d^{\frac{\alpha}{2}}$, then $\text{Proj}_{V_{d-1}} X$ becomes negligible compared to $\text{Proj}_{u_1} X$. (See Figure 1.) Thus for a large d, $\Sigma_d \approx \lambda_1 u_1 u_1'$ and the variation of X is mostly along u_1. Therefore the sample eigenvector corresponding to the largest eigenvalue, $\hat{u}_1$, will be similar to u_1.

To generalize this, suppose all eigenvalues except the first one satisfy the ϵ-condition, i.e. $\epsilon_2 \gg \frac{1}{d}$. The following proposition states that under the general setting in section 1.1, the first sample eigenvector $\hat{u}_1$ converges to its

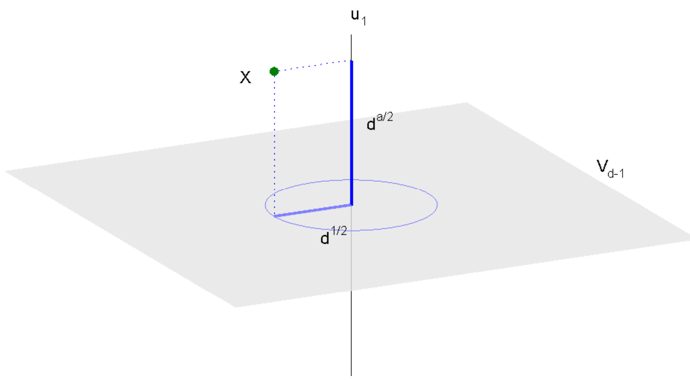

FIG 1. *Projection of a d-dimensional random variable X onto u_1 and V_{d-1}. If $\alpha > 1$, then the subspace V_{d-1} becomes negligible compared to u_1 when $d \to \infty$*

population counterpart u_1 (consistency) or tends to be perpendicular to u_1 (strong inconsistency) according to the magnitude of the first eigenvalue λ_1, while all the other sample eigenvectors are strongly inconsistent regardless of the magnitude λ_1.

PROPOSITION 1. *For a fixed n, let $\Sigma_d = U_d \Lambda_d U_d'$, $d = n+1, n+2, \ldots$ be a sequence of covariance matrices. Let $X_{(d)}$ be a $d \times n$ data matrix from a d-variate distribution with mean zero and covariance matrix Σ_d. Let $S_d = \hat{U}_d \hat{\Lambda}_d \hat{U}_d'$ be the sample covariance matrix estimated from $X_{(d)}$ for each d. Assume the following:*

(a) The components of $Z_{(d)} = \Lambda_d^{-\frac{1}{2}} U_d' X_{(d)}$ have uniformly bounded fourth moments and are ρ-mixing for some permutation.

For an $\alpha_1 > 0$,

(b) $\dfrac{\lambda_{1,d}}{d^{\alpha_1}} \longrightarrow c_1$ for some $c_1 > 0$,
(c) The ϵ_2-condition holds and $\sum_{i=2}^{d} \lambda_{i,d} = O(d)$.

If $\alpha_1 > 1$, then the first sample eigenvector is consistent and the others are strongly inconsistent in the sense that

$$Angle(\hat{u}_1, u_1) \xrightarrow{p} 0 \text{ as } d \to \infty,$$

$$Angle(\hat{u}_i, u_i) \xrightarrow{p} \frac{\pi}{2} \text{ as } d \to \infty \; ^\forall i = 2, \ldots, n.$$

If $\alpha_1 \in (0,1)$, then all sample eigenvectors are strongly inconsistent, i.e.

$$Angle(\hat{u}_i, u_i) \xrightarrow{p} \frac{\pi}{2} \text{ as } d \to \infty \ \forall i = 1, \ldots, n.$$

Note that the gap between consistency and strong inconsistency is very thin, i.e. if we avoid $\alpha_1 = 1$, then we have either consistency or strong inconsistency. Thus in the HDLSS context, asymptotic behavior of PC directions is mostly captured by consistency and strong inconsistency. Now it makes sense to say λ_1 is much larger than the others when $\alpha_1 > 1$, which results in consistency. Also note that if $\alpha_1 < 1$, then the ϵ-condition holds, which is in fact the condition for Theorem 1.

4.2. *Generalizations.* In this section, we generalize Proposition 1 to the case that multiple eigenvalues are much larger than the others. This leads to two different types of result.

First is the case that the first p eigenvectors are each consistent. Consider a covariance structure with multiple spikes, that is, p eigenvalues, $p > 1$, which are much larger than the others. In order to have consistency of the first p eigenvectors, we require that each of p eigenvalues has a distinct order of magnitude, for example, $\lambda_{1,d} = d^3$, $\lambda_{2,d} = d^2$ and sum of the rest is order of d.

PROPOSITION 2. *For a fixed n, let Σ_d, $X_{(d)}$, and S_d be as before. Assume (a) of Proposition 1. Let $\alpha_1 > \alpha_2 > \cdots > \alpha_p > 1$ for some $p < n$. Suppose the following conditions hold:*

(b) $\dfrac{\lambda_{i,d}}{d^{\alpha_i}} \longrightarrow c_i$ *for some* $c_i > 0, \ \forall i = 1, \ldots, p$

(c) The ϵ_{p+1}-condition holds and $\sum_{i=p+1}^{d} \lambda_{i,d} = O(d)$.

Then, the first p sample eigenvectors are consistent and the others are strongly inconsistent in the sense that

$$Angle(\hat{u}_i, u_i) \xrightarrow{p} 0 \text{ as } d \to \infty \ \forall i = 1, \ldots, p,$$

$$Angle(\hat{u}_i, u_i) \xrightarrow{p} \frac{\pi}{2} \text{ as } d \to \infty \ \forall i = p+1, \ldots, n.$$

Consider now a distribution having a covariance structure with multiple spikes as before. Let k be the number of spikes. An interesting phenomenon happens when the first k eigenvalues are of the same order of magnitude, i.e. $\lim_{d\to\infty} \frac{\lambda_{1,d}}{\lambda_{k,d}} = c > 1$ for some fixed constant c. Then the first k sample eigenvectors are neither consistent nor strongly inconsistent. However, all of those random directions converge to the subspace spanned by the first k population eigenvectors. Essentially, when eigenvalues are of the same order, the eigen-directions can not be separated but are subspace consistent with the proper subspace.

PROPOSITION 3. *For a fixed n, let Σ_d, $X_{(d)}$, and S_d be as before. Assume (a) of Proposition 1. Let $\alpha_1 > 1$ and $k < n$. Suppose the following conditions hold:*

(b) $\dfrac{\lambda_{i,d}}{d^{\alpha_1}} \longrightarrow c_i$ *for some* $c_i > 0$, $\forall i = 1, \ldots, k$

(c) *The ϵ_{k+1}-condition holds and* $\sum_{i=k+1}^{d} \lambda_{i,d} = O(d)$.

Then, the first k sample eigenvectors are subspace-consistent with the subspace spanned by the first k population eigenvectors and the others are strongly inconsistent in the sense that

$$Angle(\hat{u}_i, span\{u_1, \ldots, u_k\}) \xrightarrow{p} 0 \text{ as } d \to \infty \ \forall i = 1, \ldots, k,$$

$$Angle(\hat{u}_i, u_i) \xrightarrow{p} \frac{\pi}{2} \text{ as } d \to \infty \ \forall i = k+1, \ldots, n.$$

4.3. *Main theorem.* Propositions 1 - 3 are combined and generalized in the main theorem. Consider p groups of eigenvalues, which grow at the same rate within each group as in Proposition 3. Each group has a finite number of eigenvalues and the number of eigenvalues in all groups, κ, does not exceed n. Also similar to Proposition 2, let the orders of magnitude of the p groups be different to each other. We require that the $\epsilon_{\kappa+1}$-condition holds. The following theorem states that a sample eigenvector of a group converges to the subspace of population eigenvectors of the group.

THEOREM 2 (Main theorem). *For a fixed n, let Σ_d, $X_{(d)}$, and S_d be as before. Assume (a) of Proposition 1. Let $\alpha_1, \ldots, \alpha_p$ be such that $\alpha_1 > \alpha_2 > \cdots > \alpha_p > 1$ for some $p < n$. Let $k_1, \ldots, k_p$ be nonnegative integers such that $\sum_{j=1}^{p} k_j \doteq \kappa < n$. Let $k_0 = 0$ and $k_{p+1} = d - \kappa$. Let $J_1, \ldots, J_{p+1}$ be sets of indices such that*

$$J_l = \left\{ \sum_{j=0}^{l-1} k_j + 1, \sum_{j=0}^{l-1} k_j + 2, \ldots, \sum_{j=0}^{l-1} k_j + k_l \right\}, \quad l = 1, \ldots, p+1.$$

Suppose the following conditions hold:

(b) $\dfrac{\lambda_{i,d}}{d^{\alpha_l}} \longrightarrow c_i$ *for some* $c_i > 0$, $\forall i \in J_l$, $\forall l = 1, \ldots, p$

(c) *The $\epsilon_{\kappa+1}$-condition holds and* $\sum_{i \in J_{p+1}} \lambda_{i,d} = O(d)$.

Then, the sample eigenvectors whose label is in the group J_l, for $l = 1, \dots, p$, are subspace-consistent with the space spanned by the population eigenvectors whose labels are in J_l and the others are strongly inconsistent in the sense that

$$Angle(\hat{u}_i, span\{u_j : j \in J_l\}) \xrightarrow{p} 0 \text{ as } d \to \infty \ ^\forall i \in J_l, \ ^\forall l = 1, \dots p, \tag{4.1}$$

and

$$Angle(\hat{u}_i, u_i) \xrightarrow{p} \frac{\pi}{2} \text{ as } d \to \infty \ ^\forall i = \kappa + 1, \dots, n. \tag{4.2}$$

REMARK. If the cardinality of J_l, k_l, is 1, then (4.1) implies $\hat{u}_i$ is consistent for $i \in J_l$.

Note that the formulation of the theorem is similar to the spiked covariance model but much more general. The uniform assumption on the underlying eigenvalues, i.e. $\lambda_i = 1$ for all $i > \kappa$, is relaxed to the ϵ-condition. We also have catalogued a large collection of specific results according to the various sizes of spikes.

These results are now illustrated for some classes of covariance matrices that are of special interest. These covariance matrices are easily represented in *factor form*, i.e. in terms of $F_d = \Sigma_d^{\frac{1}{2}}$.

EXAMPLE 4.1. Consider a series of covariance matrices $\{\Sigma_d\}_d$. Let $\Sigma_d = F_d F_d'$, where F_d is a $d \times d$ symmetric matrix such that

$$F_d = (1 - \rho_d) I_d + \rho_d J_d = \begin{pmatrix} 1 & \rho_d & \cdots & \rho_d \\ \rho_d & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_d \\ \rho_d & \cdots & \rho_d & 1 \end{pmatrix},$$

where J_d is the $d \times d$ matrix of ones and $\rho_d \in (0, 1)$ depends on d. The eigenvalues of Σ_d are $\lambda_{1,d} = (d\rho_d + 1 - \rho_d)^2, \lambda_{2,d} = \cdots = \lambda_{d,d} = (1 - \rho_d)^2$. The first eigenvector is $u_1 = \frac{1}{\sqrt{d}}(1, 1, \dots, 1)'$, while $\{u_2, \dots, u_d\}$ are any orthogonal sets of direction vectors perpendicular to u_1. Note that $\sum_{i=2}^d \lambda_{i,d} = d(1 - \rho_d)^2 = O(d)$ and the ϵ_2-condition holds. Let $X_d \sim \mathcal{N}_d(0, \Sigma_d)$. By Theorem 2, if $\rho_d \in (0, 1)$ is a fixed constant or decreases to 0 slowly so that $\rho_d \gg d^{-\frac{1}{2}}$, then the first PC direction $\hat{u}_1$ is consistent. Else if ρ_d decreases to 0 so quickly that $\rho_d \ll d^{-\frac{1}{2}}$, then $\hat{u}_1$ is strongly inconsistent. In both cases all the other sample PC directions are strongly inconsistent.

EXAMPLE 4.2. Consider now a $2d \times 2d$ covariance matrix $\Sigma_d = F_d F_d'$, where F_d is a block diagonal matrix such that

$$F_d = \begin{pmatrix} F_{1,d} & O \\ O & F_{2,d} \end{pmatrix},$$

where $F_{1,d} = (1-\rho_{1,d})I_d + \rho_{1,d} J_d$ and $F_{2,d} = (1-\rho_{2,d})I_d + \rho_{2,d} J_d$. Suppose $0 < \rho_{2,d} \le \rho_{1,d} < 1$. Note that $\lambda_{1,d} = (d\rho_{1,d} + 1 - \rho_{1,d})^2$, $\lambda_{2,d} = (d\rho_{2,d} + 1 - \rho_{2,d})^2$ and the ϵ_3-condition holds. Let $X_{2d} \sim \mathcal{N}_{2d}(0, \Sigma_d)$. Application of Theorem 2 for various conditions on $\rho_{1,d}$, $\rho_{2,d}$ is summarized as follows. Denote, for two non-increasing sequences $\mu_d, \nu_d \in (0,1)$, $\mu_d \gg \nu_d$ for $\nu_d = o(\mu_d)$ and $\mu_d \succeq \nu_d$ for $\lim_{d\to\infty} \frac{\mu_d}{\nu_d} = c \in [1, \infty)$.

1. $\rho_{1,d} \gg \rho_{2,d} \gg d^{-\frac{1}{2}}$: Both $\hat{u}_1$,$\hat{u}_2$ consistent.
2. $\rho_{1,d} \succeq \rho_{2,d} \gg d^{-\frac{1}{2}}$: Both $\hat{u}_1$,$\hat{u}_2$ subspace-consistent to span$\{u_1, u_2\}$.
3. $\rho_{1,d} \gg d^{-\frac{1}{2}} \gg \rho_{2,d}$: $\hat{u}_1$ consistent, $\hat{u}_2$ strongly inconsistent.
4. $d^{-\frac{1}{2}} \gg \rho_{1,d} \gg \rho_{2,d}$: Both $\hat{u}_1$,$\hat{u}_2$ strongly inconsistent.

4.4. *Corollaries to the main theorem.* The result can be extended for special cases.

First of all, consider constructing $X_{(d)}$ from Z_d by $X_{(d)} \equiv U_d \Lambda_d^{\frac{1}{2}} Z_d$ where Z_d is a truncated set from an infinite sequence of independent random variables with mean zero and variance 1. This assumption makes it possible to have convergence in the almost sure sense. This is mainly because the triangular array $\{Z_{1i,(d)}\}_{i,d}$ becomes the single sequence $\{Z_{1i}\}_i$.

COROLLARY 1. *Suppose all the assumptions in Theorem 2, with the assumption (a) replaced by the following:*

(a′) The components of $Z_{(d)} = \Lambda_d^{-\frac{1}{2}} U_d' X_{(d)}$ have uniformly bounded eighth moments and are independent to each other. Let $Z_{1i,(d)} \equiv Z_{1i}$ for all i, d.

If the strong $\epsilon_{\kappa+1}$-condition (2.3) holds, then the mode of convergence of (4.1) and (4.2) is almost sure.

Second, consider the case that both d, n tend to infinity. Under the setting of Theorem 2, we can separate PC directions better when the eigenvalues are distinct. When $d \to \infty$, we have subspace consistency of $\hat{u}_i$ with the proper subspace, which includes u_i. Now letting $n \to \infty$ makes it possible for $\hat{u}_i$ to be consistent.

COROLLARY 2. *Let Σ_d, $X_{(d)}$, and S_d be as before. Under the assumptions (a), (b) and (c) in Theorem 2, assume further that the first κ eigenvalues are distinct, i.e. $\lambda_{i,d} > \lambda_{i+1,d}$ for $1 \le i < \kappa$, for all d. Then for all $i \le \kappa$,*

$$Angle(\hat{u}_i, u_i) \xrightarrow{p} 0 \text{ as } d \to \infty,\ n \to \infty. \tag{4.3}$$

If the assumption (a) is replaced by the assumption (a′) of Corollary 1, then the mode of convergence of (4.3) is almost sure.

4.5. *Limiting distributions of corresponding eigenvalues.* The study of asymptotic behavior of the sample eigenvalues is an important part in the proof of Theorem 2, and also could be of independent interest. The following lemma states that the large sample eigenvalues increase at the same speed as their population counterpart and the relatively small eigenvalues tend to be of order of d as d tends to infinity.

LEMMA 1. *If the assumptions of Theorem 2 hold, then*

$$\hat{\lambda}_i = \begin{cases} O_p(d^{\alpha_l}), & \text{if } i \in J_l,\ {}^{\forall} l = 1, \ldots, p, \\ O_p(d), & \text{if } i = \kappa + 1, \ldots, n. \end{cases} \tag{4.4}$$

If the data matrix $X_{(d)}$ is Gaussian, then the first κ sample eigenvalues converge in distribution to some quantities, which have known distributions.

COROLLARY 3. *Under all the assumptions of Theorem 2, assume further that $X_{(d)} \sim \mathcal{N}_d(0, \Sigma_d)$ for each d. Then,*

$$\frac{\hat{\lambda}_i}{d^{\alpha_l}} \Longrightarrow \varphi_{i - \sum_{j=1}^{l-1} k_l}(n^{-1}\mathcal{W}_{k_l}(n, C_l)) \text{ as } d \to \infty\ {}^{\forall} i \in J_l,\ {}^{\forall} l = 1, \ldots p,$$

where $\mathcal{W}_{k_l}(n, C_l)$ denotes a $k_l \times k_l$ random matrix distributed as the Wishart distribution with degree of freedom n and covariance C_l, and $\varphi_i(A)$ denotes the ith largest eigenvalue of the square matrix A.

If $k_l = 1$ for some l, then for $i \in J_l$

$$\frac{\hat{\lambda}_i}{\lambda_i} \Longrightarrow \frac{\chi_n^2}{n} \text{ as } d \to \infty,$$

where χ_n^2 denotes a random variable distributed as the χ^2 distribution with degree of freedom n.

This generalizes the results in section 4.2 of Ahn et al. [1].

5. Proofs.

5.1. *Proof of Theorem 1.* First we give the proof of part (1). By (1.1), the mth diagonal entry of nS_D can be expressed as $\sum_{i=1}^{d} \lambda_{i,d} z_{im,d}^2$ where $z_{im,d}$ is the (i, m)th entry of the matrix $Z_{(d)}$. Define the relative eigenvalues $\tilde{\lambda}_{i,d}$ as $\tilde{\lambda}_{i,d} \equiv \frac{\lambda_{i,d}}{\sum_{i=1}^{d} \lambda_{i,d}}$. Let π_d denote the given permutation for each d and let $Y_i = z_{\pi_d(i)m,d}^2 - 1$. Then the Y_i's are ρ-mixing, $\mathrm{E}(Y_i) = 0$ and $\mathrm{E}(Y_i^2) \le B$ for all i for some $B < \infty$. Let $\rho(m) = \sup |\mathrm{corr}(Y_i, Y_{i+m})|$ where the sup is over all i. We shall use the following lemma.

LEMMA 2. *For any permutation π_d^*,*

$$\lim_{d\to\infty} \sum_{i=1}^{d} \tilde{\lambda}_{\pi_d^*(i),d} \rho(i) = 0.$$

PROOF. For any $\delta > 0$, since $\lim_{i\to\infty} \rho(i) = 0$, we can choose N such that $\rho(i) < \frac{\delta}{2}$ for all $i > N$. Since $\lim_{d\to\infty} \sum_{i=1}^{d} \tilde{\lambda}_{\pi_d^*(i),d}^2 = 0$, we get $\lim_{d\to\infty} \sum_{i=1}^{N} \tilde{\lambda}_{\pi_d^*(i),d} = 0$. Thus we can choose d_0 satisfying $\sum_{i=1}^{N} \tilde{\lambda}_{\pi_d^*(i),d} < \frac{\delta}{2}$ for all $d > d_0$. With the fact $\sum_{i=1}^{d} \tilde{\lambda}_{i,d} = 1$ for all d and $\rho(i) < 1$, we get for all $d > d_0$,

$$\sum_{i=1}^{d} \tilde{\lambda}_{\pi_d^*(i),d} \rho(i) = \sum_{i=1}^{N} \tilde{\lambda}_{\pi_d^*(i),d} \rho(i) + \sum_{i=N+1}^{d} \tilde{\lambda}_{\pi_d^*(i),d} \rho(i) < \delta. \qquad \square$$

Now let π_d^{-1} be the inverse permutation of π_d. Then by Lemma 2 and the ϵ-condition, there exists a permutation π_d^* such that

$$\begin{aligned} \mathrm{E}(\sum_{i=1}^{d} \tilde{\lambda}_{\pi_d^{-1}(i),d} Y_i)^2 &= \sum_{i=1}^{d} \tilde{\lambda}_{\pi_d^{-1}(i),d}^2 \mathrm{E}Y_i^2 + 2 \sum_{i=1}^{d} \tilde{\lambda}_{\pi_d^{-1}(i),d} \sum_{j=i+1}^{d} \tilde{\lambda}_{\pi_d^{-1}(j),d} \mathrm{E}Y_i Y_j \\ &\le \sum_{i=1}^{d} \tilde{\lambda}_{i,d}^2 B + 2 \sum_{i=1}^{d} \tilde{\lambda}_{i,d} \sum_{j=1}^{d} \tilde{\lambda}_{\pi_d^*(j),d} \rho(j) B^2 \to 0, \end{aligned}$$

as $d \to \infty$. Then Chebyshev's inequality gives us, for any $\tau > 0$,

$$P\left[\left| \sum_{i=1}^{d} \tilde{\lambda}_{i,d} z_{ij}^2 - 1 \right| > \tau \right] \le \frac{\mathrm{E}\left(\sum_{i=1}^{d} \tilde{\lambda}_{\pi_d^{-1}(i),d} Y_i \right)^2}{\tau^2} \to 0,$$

as $d \to \infty$. Thus we conclude that the diagonal elements of nS_D converge to 1 in probability.

The off-diagonal elements of nS_D can be expressed as $\sum_{i=1}^{d} \lambda_{i,d} z_{im} z_{il}$. Similar arguments to those used in the diagonal case, together with the fact that z_{im} and z_{il} are independent, gives that

$$\mathrm{E}(\sum_{i=1}^{d} \tilde{\lambda}_{i,d} z_{im} z_{il})^2 \leq \sum_{i=1}^{d} \tilde{\lambda}_{i,d}^2 + 2 \sum_{i=1}^{d} \tilde{\lambda}_{i,d} \sum_{j=i+1}^{d} \tilde{\lambda}_{\pi_d^{-1}(j),d} \rho^2(j-i) \to 0,$$

as $d \to \infty$. Thus by Chebyshev's inequality, the off-diagonal elements of nS_D converge to 0 in probability.

Now, we give the proof for part(2). We begin with the mth diagonal entry of nS_D, $\sum_{i=1}^{d} \lambda_{i,d} z_{ij}^2$. Note that since $\sum_{i=1}^{k-1} \tilde{\lambda}_{i,d} \to 0$ by the ϵ-condition, we assume $k = 1$ in (2.3) without loss of generality.

Let $Y_i = z_{im}^2 - 1$. Note that the Y_i's are independent, $\mathrm{E}(Y_i) = 0$ and $\mathrm{E}(Y_i^4) \leq B$ for all i for some $B < \infty$. Now

$$\mathrm{E}\left(\sum_{i=1}^{d} \tilde{\lambda}_{i,d} Y_i\right)^4 = \mathrm{E} \sum_{i,j,k,l=1}^{d} \tilde{\lambda}_{i,d} \tilde{\lambda}_{j,d} \tilde{\lambda}_{k,d} \tilde{\lambda}_{l,d} Y_i Y_j Y_k Y_l. \tag{5.1}$$

Note that terms in the sum of the form $\mathrm{E}Y_iY_jY_kY_l$, $\mathrm{E}Y_i^2Y_jY_k$, and $\mathrm{E}Y_i^3Y_j$ are 0 if i, j, k, l are distinct. The only terms that do not vanish are those of the form $\mathrm{E}Y_i^4$, $\mathrm{E}Y_i^2Y_j^2$, both of which are bounded by B. Note that by applying the Cauchy-Schwartz inequality repeatedly, we get

$$\sum_{i=1}^{d} \tilde{\lambda}_{i,d}^4 \leq (\sum_{i=1}^{d} \tilde{\lambda}_{i,d}^2)^2$$

Also note that by the strong ϵ-condition, $\sum_{i=1}^{d} \tilde{\lambda}_{i,d}^2 = (d\epsilon)^{-1} = o(d^{-\frac{1}{2}})$. Thus (5.1) is bounded as

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{d} \tilde{\lambda}_{i,d} Y_i\right)^4 &\leq \sum_{i=1}^{d} \tilde{\lambda}_{i,d}^4 B + \sum_{i=j\neq k=l} \tilde{\lambda}_{i,d} \tilde{\lambda}_{k,d} B \\
&\leq (\sum_{i=1}^{d} \tilde{\lambda}_{i,d}^2)^2 B + \binom{4}{2} (\sum_{i=1}^{d} \tilde{\lambda}_{i,d}^2)^2 B \\
&= o(d^{-1})
\end{aligned}$$

Then Chebyshev's inequality gives us, for any $\tau > 0$,

$$P\left[\left|\sum_{i=1}^{d} \tilde{\lambda}_{i,d} z_{ij}^2 - 1\right| > \tau\right] \leq \frac{\mathrm{E}\left(\sum_{i=1}^{d} \tilde{\lambda}_{i,d} Y_i\right)^4}{\tau^4} \leq \frac{o(d^{-1})}{\tau^4}$$

Summing over d gives $\sum_{d=1}^{\infty} P\left[\left|\sum_{i=1}^{d} \tilde{\lambda}_{i,d} z_{ij}^2 - 1\right| > \tau\right] < \infty$ and by Borel-Cantelli Lemma, we conclude that a diagonal element $\sum_{i=1}^{d} \tilde{\lambda}_{i,d} z_{ij}^2$ converges to 1 almost surely.

The off-diagonal elements of nS_D can be expressed as $\sum_{i=1}^{d} \lambda_{i,d} z_{im} z_{il}$. Using similar arguments to those used in the diagonal case, we have

$$P\left[\left|\sum_{i=1}^{d} \tilde{\lambda}_{i,d} z_{im} z_{il}\right| > \tau\right] \le \frac{\mathrm{E}\left(\sum_{i=1}^{d} \tilde{\lambda}_{i,d} z_{im} z_{il}\right)^4}{\tau^4} \le \frac{o(d^{-1})}{\tau^4},$$

and again by the Borel-Cantelli Lemma, the off-diagonal elements converge to 0 almost surely. □

5.2. *Proofs of Lemma 1 and Theorem 2.* The proof of Theorem 2 is divided in two parts. Since eigenvectors are associated to eigenvalues, at first, we focus on asymptotic behavior of sample eigenvalues (section 5.2.1) and then investigate consistency or strong inconsistency of sample eigenvectors (section 5.2.2).

5.2.1. *Proof of Lemma 1.* We introduce a few definitions and lemmas that are useful to prove this lemma. Let $\mathcal{S}_m$ be the set of all $m \times m$ symmetric matrices. Let $\varphi(A)$ be a vector of eigenvalues of A for $A \in \mathcal{S}_m$ arranged in non-increasing order and let $\varphi_i(A)$ be the ith largest eigenvalue of A. Let $\| \bullet \|_2$ be the usual 2-norm of vectors, and $\| \bullet \|_F$ be the Frobenius norm of matrices defined by $\|A\|_F = (\sum_{i,j} A_{ij}^2)^{1/2}$.

LEMMA 3 (Wielandt-Hoffman inequality). *If A, $B \in \mathcal{S}_m$, then*

$$\|\varphi(A+B) - \varphi(A)\|_2 \le \|\varphi(B)\|_2 = \|B\|_F.$$

This inequality is known as Wielandt-Hoffman inequality. See Wilkinson [15] for detailed discussion and proof.

COROLLARY 4 (Continuity of eigenvalues). *The mapping of eigenvalues $\varphi : \mathcal{S}_m \longmapsto \mathbf{R}_m$ is uniformly continuous.*

PROOF. By Lemma 3, $^\forall \epsilon > 0$, $^\forall A, B \in \mathcal{S}_m$, $^\exists \delta = \epsilon$ such that $\|A - B\|_F \le \delta$, then

$$\|\varphi(A) - \varphi(B)\|_2 \le \|\varphi(A-B)\|_2 \le \delta = \epsilon.$$ □

The proof relies heavily on the following lemma.

LEMMA 4 (Weyl's inequality). *If A, $B \in \mathcal{S}_m$, then* $^\forall k = 1, \dots, m$,

$$\left.\begin{array}{c} \varphi_k(A) + \varphi_m(B) \\ \varphi_{k+1}(A) + \varphi_{m-1}(B) \\ \vdots \\ \varphi_m(A) + \varphi_k(B) \end{array}\right\} \leq \varphi_k(A+B) \leq \left\{\begin{array}{c} \varphi_k(A) + \varphi_1(B) \\ \varphi_{k-1}(A) + \varphi_2(B) \\ \vdots \\ \varphi_1(A) + \varphi_k(B) \end{array}\right.$$

This inequality is discussed in Rao [14] and its use on asymptotic studies of eigenvalues of a random matrix appeared in Eaton and Tyler [5].

Since the dimension of the sample covariance matrix S grows, it is not easy to deal with eigenvalues of S directly. One of the main ideas of the proof is working with S_D, dual of S. By our decomposition (1.1),

$$nS_D = (Z'\Lambda^{\frac{1}{2}}U')(U\Lambda^{\frac{1}{2}}Z) = Z'\Lambda Z.$$

We also write Z and Λ as block matrices such that

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{p+1} \end{pmatrix},$$

where Z_l is a $k_l \times n$ matrix for each $l = 1, \dots, p+1$ and

$$\Lambda = \begin{pmatrix} \Lambda_1 & O & \cdots & O \\ O & \Lambda_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \Lambda_{p+1} \end{pmatrix},$$

where $\Lambda_l (\equiv \Lambda_{l,d})$ is a $k_l \times k_l$ diagonal matrix for each $l = 1, \dots, p+1$ and O denotes a matrix where all elements are zeros. Now we can write

$$nS_D = Z'\Lambda Z = \sum_{l=1}^{p+1} Z_l'\Lambda_l Z_l. \tag{5.2}$$

Note that the $k_l \times k_l$ random matrix Z_l depends on d. We will, however, simplify notation Z_l for representing for all $d = 1, \dots, \infty$.

Note that Theorem 1 implies that when the last term in equation (5.2) is divided by d, it converges to an identity matrix, namely,

$$d^{-1} Z_{p+1}'\Lambda_{p+1} Z_{p+1} \xrightarrow{p} K \cdot I_n, \tag{5.3}$$

where $K \in (0,\infty)$ is such that $d^{-1}\sum_{j \in J_{p+1}} \lambda_{i,d} \to K$. Moreover dividing by d^{α_1} gives us

$$nd^{-\alpha_1} S_D = d^{-\alpha_1} Z_1' \Lambda_1 Z_1 + d^{-\alpha_1} \sum_{l=2}^{p} Z_l' \Lambda_l Z_l + d^{1-\alpha_1} d^{-1} Z_{p+1}' \Lambda_{p+1} Z_{p+1}.$$

By the assumption (b), the first term on the right hand side converges to $Z_1' C_1 Z_1$ where C_1 is the $k_1 \times k_1$ diagonal matrix such that $C_1 = \mathrm{diag}\{c_j; j \in J_1\}$ and the other terms tend to a zero matrix. Thus, we get

$$nd^{-\alpha_1} S_D \Longrightarrow Z_1' C_1 Z_1 \text{ as } d \to \infty.$$

Note that the non-zero eigenvalues of $Z_1' C_1 Z_1$ are the same as the nonzero eigenvalues of $C_1^{\frac{1}{2}} Z_1 Z_1' C_1^{\frac{1}{2}}$ which is a $k_1 \times k_1$ random matrix with full rank almost surely. By Corollary 4, we have for $i \in J_1$,

$$\begin{aligned} \varphi_i(nd^{-\alpha_1} S_D) &\Longrightarrow \varphi_i(Z_1' C_1 Z_1) \text{ as } d \to \infty \\ &= \varphi_i(C_1^{\frac{1}{2}} Z_1 Z_1' C_1^{\frac{1}{2}}). \end{aligned}$$

Thus, we conclude that for the sample eigenvalues in the group J_1, $\hat{\lambda}_i = \varphi_i(S_D) = O_p(d^{\alpha_1})$ for $i \in J_1$.

Let us focus on eigenvalues whose indices are in the group $J_2, \ldots, J_p$. Suppose we have $\hat{\lambda}_i = O_p(d^{\alpha_j})$ for all $i \in J_j$, for $j = 1, \ldots, l-1$. Pick any $i \in J_l$. We will provide upper and lower bounds on $\hat{\lambda}_i$ by Weyl's inequality (Lemma 4). Dividing both sides of (5.2) by d^{α_l}, we get

$$nd^{-\alpha_l} S_D = d^{-\alpha_l} \sum_{j=1}^{l-1} Z_j' \Lambda_j Z_j + d^{-\alpha_l} \sum_{j=l}^{p+1} Z_j' \Lambda_j Z_j$$

and apply Weyl's inequality for the upper bound,

$$\begin{aligned} \varphi_i(nd^{-\alpha_l} S_D) &\le \varphi_{1+\sum_{j=1}^{l-1} k_j}(d^{-\alpha_l} \sum_{j=1}^{l-1} Z_j' \Lambda_j Z_j) + \varphi_{i-\sum_{j=1}^{l-1} k_j}(d^{-\alpha_l} \sum_{j=l}^{p+1} Z_j' \Lambda_j Z_j) \\ &= \varphi_{i-\sum_{j=1}^{l-1} k_j}(d^{-\alpha_l} \sum_{j=l}^{p+1} Z_j' \Lambda_j Z_j). \end{aligned} \tag{5.4}$$

Note that the first term vanishes since the rank of $d^{-\alpha_l} \sum_{j=1}^{l-1} Z_j' \Lambda_j Z_j$ is at most $\sum_{j=1}^{l-1} k_j$. Also note that the matrix in the upper bound (5.4) converges to a simple form

$$\begin{aligned} d^{-\alpha_l} \sum_{j=l}^{p+1} Z_j' \Lambda_j Z_j &= d^{-\alpha_l} Z_l' \Lambda_l Z_l + d^{-\alpha_l} \sum_{j=l+1}^{p+1} Z_j' \Lambda_j Z_j \\ &\Longrightarrow Z_l' C_l Z_l \text{ as } d \to \infty, \end{aligned}$$

where C_l is the $k_l \times k_l$ diagonal matrix such that $C_l = \mathrm{diag}\{c_j; j \in J_l\}$.

In order to have a lower bound of $\hat{\lambda}_i$, Weyl's inequality is applied to the expression

$$d^{-\alpha_l} \sum_{j=1}^{l} Z_j' \Lambda_j Z_j + d^{-\alpha_l} \sum_{j=l+1}^{p+1} Z_j' \Lambda_j Z_j = n d^{-\alpha_l} S_D,$$

so that

$$\varphi_i(d^{-\alpha_l} \sum_{j=1}^{l} Z_j' \Lambda_j Z_j) + \varphi_n(d^{-\alpha_l} \sum_{j=l+1}^{p+1} Z_j' \Lambda_j Z_j) \leq \varphi_i(n d^{-\alpha_l} S_D). \tag{5.5}$$

It turns out that the first term of the left hand side is not easy to manage, so we again use Weyl's inequality to get

(5.6)

$$\varphi_{\sum_{j=1}^{l} k_j}(d^{-\alpha_l} \sum_{j=1}^{l-1} Z_j' \Lambda_j Z_j) \leq \varphi_i(d^{-\alpha_l} \sum_{j=1}^{l} Z_j' \Lambda_j Z_j) + \varphi_{\sum_{j=1}^{l} k_j - i + 1}(-d^{-\alpha_l} Z_l' \Lambda_l Z_l),$$

where the left hand side is 0 since the rank of the matrix inside is at most $\sum_{j=1}^{l-1} k_j$. Note that since $d^{-\alpha_l} Z_l' \Lambda_l Z_l$ and $d^{-\alpha_l} \Lambda_l^{\frac{1}{2}} Z_l Z_l' \Lambda_l^{\frac{1}{2}}$ share non-zero eigenvalues, we get

$$\begin{aligned}
\varphi_{\sum_{j=1}^{l} k_j - i + 1}(-d^{-\alpha_l} Z_l' \Lambda_l Z_l) &= \varphi_{\sum_{j=1}^{l} k_j - i + 1}(-d^{-\alpha_l} \Lambda_l^{\frac{1}{2}} Z_l Z_l' \Lambda_l^{\frac{1}{2}}) \\
&= \varphi_{k_l - i + \sum_{j=1}^{l-1} k_j + 1}(-d^{-\alpha_l} \Lambda_l^{\frac{1}{2}} Z_l Z_l' \Lambda_l^{\frac{1}{2}}) \\
&= -\varphi_{i - \sum_{j=1}^{l-1} k_j}(d^{-\alpha_l} \Lambda_l^{\frac{1}{2}} Z_l Z_l' \Lambda_l^{\frac{1}{2}}) \\
&= -\varphi_{i - \sum_{j=1}^{l-1} k_j}(d^{-\alpha_l} Z_l' \Lambda_l Z_l). \qquad (5.7)
\end{aligned}$$

Here we use the fact that for any $m \times m$ positive definite symmetric matrix A, $\varphi_i(A) = -\varphi_{m-i+1}(-A)$ for all $i = 1, \ldots, m$.

Combining (5.5), (5.6), and (5.7) gives the lower bound

$$\varphi_{i - \sum_{j=1}^{l-1} k_j}(d^{-\alpha_l} Z_l' \Lambda_l Z_l) + \varphi_n(d^{-\alpha_l} \sum_{j=l+1}^{p+1} Z_j' \Lambda_j Z_j) \leq \varphi_i(n d^{-\alpha_l} S_D). \tag{5.8}$$

Note that the matrix inside of the first term of the lower bound (5.8) converges to $Z_l' C_l Z_l$ in distribution. The second term converges to 0 since the matrix inside converges to a zero matrix.

The difference between the upper and lower bounds of $\varphi_i(nd^{-\alpha_l}S_D)$ converges to 0 since

$$\varphi_{i-\sum_{j=1}^{l-1}k_j}(d^{-\alpha_l}\sum_{j=l}^{p+1}Z_j'\Lambda_jZ_j) - \varphi_{i-\sum_{j=1}^{l-1}k_j}(d^{-\alpha_l}Z_l'\Lambda_lZ_l) \to 0,$$

as $d \to \infty$. This is because φ is a continuous function and the difference between the two matrices converges to zero matrix. Therefore $\varphi_i(nd^{-\alpha_l}S_D)$ converges to the upper or lower bound as $d \to \infty$.

Now since both upper and lower bound of $\varphi_i(nd^{-\alpha_l}S_D)$ converge in distribution to same quantity, we have

$$\begin{aligned} \varphi_i(nd^{-\alpha_l}S_D) &\Longrightarrow \varphi_{i-\sum_{j=1}^{l-1}k_j}(Z_l'C_lZ_l) \text{ as } d \to \infty. \\ &= \varphi_{i-\sum_{j=1}^{l-1}k_j}(C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}}). \end{aligned} \tag{5.9}$$

Thus, by induction, we have for the ith sample eigenvalue $\hat{\lambda}_i = \varphi_i(S_D) = O_p(d^{\alpha_l})$ for $i \in J_l$, $l = 1, \ldots, p$.

Now let us focus on the rest of the sample eigenvalues $\hat{\lambda}_i$, $i = \kappa+1, \ldots, n$. For any i, again by Weyl's upper bound inequality we get

$$\begin{aligned} \varphi_i(nd^{-1}S_D) &\leq \varphi_{i-\kappa}(d^{-1}Z_{p+1}'\Lambda_{p+1}Z_{p+1}) + \varphi_{\kappa+1}(d^{-1}\sum_{j=1}^{p}Z_j'\Lambda_jZ_j) \\ &= \varphi_{i-\kappa}(d^{-1}Z_{p+1}'\Lambda_{p+1}Z_{p+1}), \end{aligned}$$

where the second term on the right hand side vanishes since the matrix inside is of rank at most κ. Also for lower bound, we have

$$\begin{aligned} \varphi_i(nd^{-1}S_D) &\geq \varphi_i(d^{-1}Z_{p+1}'\Lambda_{p+1}Z_{p+1}) + \varphi_n(d^{-1}\sum_{j=1}^{p}Z_j'\Lambda_jZ_j) \\ &= \varphi_i(d^{-1}Z_{p+1}'\Lambda_{p+1}Z_{p+1}), \end{aligned}$$

where the second term vanishes since $\kappa < n$. Thus we have complete bounds for $\varphi_i(nd^{-1}S_D)$ such that

$$\varphi_i(d^{-1}Z_{p+1}'\Lambda_{p+1}Z_{p+1}) \leq \varphi_i(nd^{-1}S_D) \leq \varphi_{i-\kappa}(d^{-1}Z_{p+1}'\Lambda_{p+1}Z_{p+1}),$$

for all $i = \kappa+1, \ldots, n$. However, by (5.3), the matrix in both bounds converges to $K \cdot I_n$ in probability. Thus lower and upper bounds of $\varphi_i(nd^{-1}S_D)$ converge to K in probability which implies that $\hat{\lambda}_i = \varphi_i(S_D) = O_p(d)$ for all $i = \kappa+1, \ldots, n$.

Therefore, we have a complete proof of (4.4). □

5.2.2. *Proof of Theorem 2.* We begin by defining a standardized version of the sample covariance matrix as

$$\begin{aligned} \tilde{S} &= \Lambda^{-\frac{1}{2}} U' S U \Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}} U' (\hat{U}\hat{\Lambda}\hat{U}') U \Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}} P \hat{\Lambda} P' \Lambda^{-\frac{1}{2}}, \end{aligned} \tag{5.10}$$

where $P = U'\hat{U} = \{u_i'\hat{u}_j\}_{ij} \equiv \{p_{ij}\}_{ij}$. Note that elements of P are inner products between population eigenvectors and sample eigenvectors. Since we also have $S = n^{-1}XX'$ and $X = U\Lambda^{\frac{1}{2}}Z$,

$$\begin{aligned} \tilde{S} &= n^{-1}\Lambda^{-\frac{1}{2}} U' X X' U \Lambda^{-\frac{1}{2}} \\ &= n^{-1}\Lambda^{-\frac{1}{2}} U' U \Lambda^{\frac{1}{2}} Z Z' \Lambda^{\frac{1}{2}} U' U \Lambda^{-\frac{1}{2}} \\ &= n^{-1} Z Z'. \end{aligned} \tag{5.11}$$

Note that the angle between the two directions can be formulated as an inner product of the two direction vectors. Thus we will investigate the behavior of the inner product matrix P as $d \to \infty$, by showing that

$$\sum_{j \in J_l} p_{ji}^2 \xrightarrow{p} 1 \text{ as } d \to \infty, \tag{5.12}$$

for all $i \in J_l$, $l = 1, \ldots, p$, and

$$p_{ii}^2 \xrightarrow{p} 0 \text{ as } d \to \infty, \tag{5.13}$$

for all $i = \kappa + 1, \ldots, n$.

Suppose for now we have the result of (5.12) and (5.13). Then for any $i \in J_l$, $l = 1, \ldots, p$,

$$\begin{aligned} \text{Angle}(\hat{u}_i, \text{span}\{u_j : j \in J_l\}) &= \arccos\left(\frac{\hat{u}_i'[\text{Proj}_{\text{span}\{u_j : j \in J_l\}} \hat{u}_i]}{\|\hat{u}_i\|_2 \cdot \|[\text{Proj}_{\text{span}\{u_j : j \in J_l\}} \hat{u}_i]\|_2}\right) \\ &= \arccos\left(\frac{\hat{u}_i'(\sum_{j \in J_l} (u_j'\hat{u}_i) u_j)}{\|\hat{u}_i\|_2 \cdot \|\sum_{j \in J_l} (u_j'\hat{u}_i) u_j\|_2}\right) \\ &= \arccos\left(\frac{\sum_{j \in J_l} (u_j'\hat{u}_i)^2}{1 \cdot (\sum_{j \in J_l} (u_j'\hat{u}_i)^2)^{\frac{1}{2}}}\right) \\ &= \arccos\left(\left(\sum_{j \in J_l} p_{ji}^2\right)^{\frac{1}{2}}\right) \\ &\xrightarrow{p} 0 \text{ as } d \to \infty, \end{aligned}$$

by (5.12) and for $i = \kappa + 1, \ldots, n$,

$$\begin{aligned} \text{Angle}(\hat{u}_i, u_i) &= \arccos(|u_i' \hat{u}_i|) \\ &= \arccos(|p_{ii}|) \\ &\xrightarrow{p} \frac{\pi}{2} \text{ as } d \to \infty, \end{aligned}$$

by (5.13), as desired.

Therefore, it is enough to show (5.12) and (5.13). We begin with taking jth diagonal entry of $\tilde{S}$, $\tilde{s}_{jj}$, from (5.10) and (5.11),

$$\tilde{s}_{jj} = \lambda_j^{-1} \sum_{i=1}^{n} \hat{\lambda}_i p_{ji}^2 = n^{-1} z_j z_j',$$

where z_j denotes the jth row vector of Z. Since

$$\lambda_j^{-1} \hat{\lambda}_i p_{ji}^2 \le n^{-1} z_j z_j', \tag{5.14}$$

we have at most

$$p_{ji}^2 = O_p(\frac{\lambda_j}{\hat{\lambda}_i}),$$

for all $i = 1, \ldots, n$, $j = 1, \ldots, d$. Note that by Lemma 1, we have for $i \in J_{l_1}$, $j \in J_{l_2}$ where $1 \le l_1 < l_2 \le p+1$,

$$p_{ji}^2 = O_p(\frac{\lambda_j}{\hat{\lambda}_i}) = \begin{cases} O_p(d^{\alpha_{l_2} - \alpha_{l_1}}), & \text{if } l_1 = 1, \ldots, p, \\ O_p(d^{1-\alpha_{l_1}}), & \text{if } l_2 = p+2, \end{cases} \tag{5.15}$$

so that $p_{ji}^2 \xrightarrow{p} 0$ as $d \to \infty$ in both cases.

Note that the inner product matrix P is also a unitary matrix. The norm of the ith column vector of P must be 1 for all d, i.e. $\sum_{j=1}^{d} p_{ji}^2 = 1$. Thus (5.12) is equivalent to $\sum_{j \in \{1,\ldots,d\}/J_l} p_{ji}^2 \xrightarrow{p} 0$ as $d \to \infty$.

Now for any $i \in J_1$,

$$\sum_{j \in \{1,\ldots,d\}/J_1} p_{ji}^2 = \sum_{j \in J_2 \cup \cdots \cup J_p} p_{ji}^2 + \sum_{j \in J_{p+1}} p_{ji}^2.$$

Since the first term on the right hand side is a finite sum of quantities converging to 0, it converges to 0 almost surely as d tends to infinity. By (5.14), we have an upper bound for the second term,

$$\begin{aligned} \sum_{j \in J_{p+1}} p_{ji}^2 &= \sum_{j \in J_{p+1}} \lambda_j^{-1} \hat{\lambda}_i p_{ji}^2 \frac{\lambda_j}{\hat{\lambda}_i} \\ &\le \frac{\sum_{j \in J_{p+1}} n^{-1} z_j z_j'}{d} \cdot \frac{\lambda_{\kappa+1}}{d^{-\alpha_1} \hat{\lambda}_i} \cdot d^{1-\alpha_1} \\ &\le \frac{\sum_{k=1}^{n} \sum_{j=\kappa+1}^{d} z_{j,k}^2}{nd} \cdot \frac{\lambda_{\kappa+1}}{d^{-\alpha_1} \hat{\lambda}_i} \cdot d^{1-\alpha_1}, \end{aligned}$$

where the $z_{j,k}$'s are the entries of a row random vector z_j. Note that by applying Theorem 1 with $\Sigma_d = I_d$, we have $\frac{\sum_{j=\kappa+1}^{d} z_{j,k}^2}{d} \xrightarrow{p} 1$ as $d \to \infty$. Also because $d^{-\alpha_1}\hat{\lambda}_i = O_p(1)$ and $d^{1-\alpha_1} \to 0$, the upper bound converges to 0 in probability. Thus we get

$$\sum_{j \in \{1,\ldots,d\}/J_1} p_{ji}^2 \xrightarrow{p} 0 \text{ as } d \to \infty,$$

which is equivalent to

$$\sum_{j \in J_1} p_{ji}^2 \xrightarrow{p} 1 \text{ as } d \to \infty. \tag{5.16}$$

Let us focus on the group $J_2, \ldots, J_p$. For any $l = 2, \ldots, p$, suppose we have $\sum_{j \in J_m} p_{ji}^2 \xrightarrow{p} 1$ as $d \to \infty$ for all $i \in J_m$, $m = 1, \ldots, l-1$. Note that it implies that for any $j \in J_m$,, $m = 1, \ldots, l-1$,

$$\sum_{i \in \{1,\ldots,d\}/J_m} p_{ji}^2 \xrightarrow{p} 0 \text{ as } d \to \infty, \tag{5.17}$$

since

$$\sum_{j \in J_m} \sum_{i \in \{1,\ldots,d\}/J_m} p_{ji}^2 = \sum_{j \in J_m} \sum_{i=1}^{d} p_{ji}^2 - \sum_{j \in J_m} \sum_{i \in J_m} p_{ji}^2 \xrightarrow{p} \sum_{j \in J_m} 1 - \sum_{i \in J_m} 1 = 0,$$

as $d \to \infty$.

Now pick $i \in J_l$. We have

$$\sum_{j \in \{1,\ldots,d\}/J_l} p_{ji}^2 = \sum_{j \in J_1 \cup \cdots \cup J_{l-1}} p_{ji}^2 + \sum_{j \in J_{l+1} \cup \cdots \cup J_p} p_{ji}^2 + \sum_{j \in J_{p+1}} p_{ji}^2.$$

Note that the first term is bounded as

$$\sum_{j \in J_1 \cup \cdots \cup J_{l-1}} p_{ji}^2 \le \sum_{i \in J_l} \sum_{j \in J_1 \cup \cdots \cup J_{l-1}} p_{ji}^2 \le \sum_{m=1}^{l-1} \sum_{j \in J_m} \left(\sum_{i \in \{1,\ldots,d\}/J_m} p_{ji}^2 \right) \xrightarrow{p} 0$$

by (5.17). The second term also converges to 0 by (5.15). The last term is also bounded as

$$\begin{aligned} \sum_{j \in J_{p+1}} p_{ji}^2 &= \sum_{j \in J_{p+1}} \lambda_j^{-1} \hat{\lambda}_i p_{ji}^2 \frac{\lambda_j}{\hat{\lambda}_i} \\ &\le \frac{\sum_{j \in J_{p+1}} n^{-1} z_j z_j'}{d} \cdot \frac{\lambda_{m+1}}{d^{-\alpha_l} \hat{\lambda}_i} \cdot d^{1-\alpha_l}, \end{aligned}$$

so that it also converges to 0 in probability. Thus, we have $\sum_{j\in\{1,\dots,d\}/J_l} p_{ji}^2 \xrightarrow{p} 0$ as $d\to\infty$ which implies that

$$\sum_{j\in J_l} p_{ji}^2 \xrightarrow{p} 1 \text{ as } d\to\infty.$$

Thus, by induction, (5.12) is proved.

For $i=\kappa+1,\dots,n$, We have $\lambda_i^{-1}\hat{\lambda}_i p_{ii}^2 \le n^{-1}z_i z_i'$, and so

$$p_{ii}^2 \le \hat{\lambda}_i^{-1}\lambda_i n^{-1} z_i z_i' = O_p(\hat{\lambda}_i^{-1}\lambda_i) = \frac{o_p(d)}{O_p(d)},$$

which implies (5.13) and the proof is completed. □

5.3. *Proof of Corollary 1.* The proof follows the same lines as the proof of Theorem 2, with convergence in probability replaced by almost sure convergence. □

5.4. *Proof of Corollary 2.* From the proof of Theorem 2, write the inner product matrix P of (5.10) as a block matrix such that

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1p} & P_{1,p+1} \\ \vdots & \ddots & \vdots & \vdots \\ P_{p1} & \cdots & P_{pp} & P_{p,p+1} \\ P_{p+1,1} & \cdots & P_{p+1,p} & P_{p+1,p+1} \end{pmatrix},$$

where each P_{ij} is a $k_i\times k_j$ random matrix. In the proof of theorem 2 we have shown that P_{ii}, $i=1,\dots,p$, tends to be a unitary matrix and P_{ij}, $i\neq j$, tends to be a zero matrix as $d\to\infty$. Likewise, Λ and $\hat{\Lambda}$ can be blocked similarly as $\Lambda = \mathrm{diag}\{\Lambda_i : i=1,\dots,p+1\}$ and $\hat{\Lambda} = \mathrm{diag}\{\hat{\Lambda}_i : i = 1,\dots,p+1\}$.

Now pick $l\in\{1,\dots,p\}$. The lth block diagonal of $\tilde{S}$, $\tilde{S}_{ll}$, is expressed as $\tilde{S}_{ll} = \sum_{j=1}^{p+1}\Lambda_l^{-\frac{1}{2}}P_{lj}\hat{\Lambda}_l P_{lj}'\Lambda_l^{-\frac{1}{2}}$. Since $P_{ij}\to 0$, $i\neq j$, we get

$$\|\tilde{S}_{ll} - \Lambda_l^{-\frac{1}{2}}P_{ll}\hat{\Lambda}_l P_{ll}'\Lambda_l^{-\frac{1}{2}}\|_F \xrightarrow{p} 0$$

as $d\to\infty$.

Note that by (5.11), $\tilde{S}_{ll}$ can be replaced by $n^{-1}Z_lZ_l'$. We also have $d^{-\alpha_l}\Lambda_l\to C_l$ by the assumption (b) and $d^{-\alpha_l}\hat{\Lambda}_l \xrightarrow{p} \mathrm{diag}\{\varphi(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}})\}$ by (5.9). Thus we get

$$\|n^{-1}Z_lZ_l' - C_l^{-\frac{1}{2}}P_{ll}\mathrm{diag}\{\varphi(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}})P_{ll}'C_l^{-\frac{1}{2}}\|_F \xrightarrow{p} 0$$

as $d \to \infty$.

Also note that since $n^{-1}Z_lZ_l' \to I_{k_l}$ almost surely as $n \to \infty$, we get $n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}} \to C_l$ and $\mathrm{diag}\{\varphi(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}})\} \to C_l$ almost surely as $n \to \infty$. Using the fact that the Frobenius norm is unitarily invariant and $\|AB\|_F \le \|A\|_F\|B\|_F$ for any square matrices A and B, we get

(5.18)

$$\begin{aligned}
\|P_{ll}'C_lP_{ll} - C_l\|_F &\le \|P_{ll}'C_lP_{ll} - \mathrm{diag}\{\varphi(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}})\|_F + o_p(1) \\
&= \|C_l - P_{ll}\mathrm{diag}\{\varphi(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}})P_{ll}'\|_F + o_p(1) \\
&\le \|n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}} - P_{ll}\mathrm{diag}\{\varphi(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}})P_{ll}'\|_F + o_p(1) \\
&\le \|C_l^{\frac{1}{2}}\|_F^2\|n^{-1}Z_lZ_l' - C_l^{-\frac{1}{2}}P_{ll}\mathrm{diag}\{\varphi(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}})P_{ll}'C_l^{-\frac{1}{2}}\|_F + o_p(1) \\
&\xrightarrow{p} 0 \text{ as } d, n \to \infty.
\end{aligned}$$

Note that in order to have (5.18), P_{ll} must converge to $\mathrm{diag}\{\pm 1, \pm 1, \ldots, \pm 1\}$ since diagonal entries of C_l are distinct and a spectral decomposition is unique up to sign changes. Let $l = 1$ for simplicity. Now suppose for any $\delta > 0$, $\lim_{d,n} P(p_{m1}^2 > \delta) > 0$ for $m = 2, \ldots, k_1$. Then for any $m = 2, \ldots, k_1$,

$$\|P_{11}'C_1P_{11} - C_1\|_F \ge \sum_{j=1}^{k_1}(c_1 - c_j)p_{j1}^2 \ge (c_1 - c_m)p_{m1}^2,$$

which contradicts (5.18) since $c_1 - c_m > 0$. Thus $p_{m1}^2 \xrightarrow{p} 0$ for all $m = 2, \ldots, k_l$ which implies $p_{11}^2 \xrightarrow{p} 1$ as $d, n \to \infty$. Now by induction, $p_{ii}^2 \xrightarrow{p} 1$ for all $i \in J_l, l = 1, \ldots, p$. Therefore $\mathrm{Angle}(\hat{u}_i, u_i) = \arccos(|p_{ii}|) \xrightarrow{p} 0$ as $d, n \to \infty$.

If the assumptions of Corollary 1 also hold, then every convergence in the proof is replaced by almost sure convergence, which completes the proof. □

5.5. *Proof of Corollary 3.* For any $i \in J_l, l = 1, \ldots, p$, (5.9) gives

$$\frac{\hat{\lambda}_i}{d^{\alpha_l}} \Longrightarrow \varphi_{i-\sum_{j=1}^{l-1}k_l}(n^{-1}C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}}) \text{ as } d \to \infty.$$

Noticing $C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}} \sim \mathcal{W}_{k_l}(n, C_l)$ gives the result. When $k_l = 1$, The assumption (b) and that $C_l^{\frac{1}{2}}Z_lZ_l'C_l^{\frac{1}{2}} \sim c_i\chi_n^2$ imply that

$$\frac{\hat{\lambda}_i}{\lambda_i} = \frac{\hat{\lambda}_i}{c_id^{\alpha_l}} \cdot \frac{c_id^{\alpha_l}}{\lambda_i} \Longrightarrow \frac{\chi_n^2}{n} \text{ as } d \to \infty. \quad \square$$

Acknowledgements. The authors are very grateful to John T. Kent (University of Leeds, UK) for the insightful example 3.1.

References.

[1] Ahn, J., Marron, J. S., Muller, K. M., and Chi, Y.-Y. (2007). The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrika* **94**, 3, 760–766.

[2] Baik, J., Ben Arous, G., and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33**, 5, 1643–1697. MR2165575

[3] Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97**, 6, 1382–1408. MR2279680

[4] Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv. 2*, 107–144 (electronic). Update of, and a supplement to, the 1986 original. MR2178042

[5] Eaton, M. L. and Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.* **19**, 1, 260–271. MR1091849

[6] Gaydos, T. L. (2008). Data representation and basis selection to understand variation of function valued traits. Ph.D. thesis, University of North Carolina at Chapel Hill. http://www.unc.edu/ tgaydos/Dissertation.

[7] Hall, P., Marron, J. S., and Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 3, 427–444. MR2155347

[8] John, S. (1971). Some optimal multivariate tests. *Biometrika 58*, 123–127.

[9] John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika 59*, 169–173.

[10] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 2, 295–327. MR1863961

[11] Johnstone, I. M. and Lu, A. Y. (2004). Sparse principal component analysis. Manuscript.

[12] Kolmogorov, A. N. and Rozanov, Y. A. (1960). On strong mixing conditions for stationary gaussian processes. *Theory Probab. Appl.* **5**, 2, 204–208.

[13] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica 17*, 1617–1642.

[14] Rao, C. R. (1973). *Linear statistical inference and its applications*, Second ed. John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics. MR0346957

[15] Wilkinson, J. H. (1988). *The algebraic eigenvalue problem*. Monographs on Numerical Analysis. The Clarendon Press Oxford University Press, New York. Oxford Science Publications. MR950175

Department of Statistics
and Operations Research
University of North Carolina
Chapel Hill, NC 27599
E-mail: sungkyu@email.unc.edu
marron@email.unc.edu