Journal of Machine Learning Research 17 (2016) 1-34 Submitted 6/14; Revised 3/16; Published 9/16

# A General Framework for Consistency of Principal Component Analysis

**Dan Shen** DANSHEN@USF.EDU
*Interdisciplinary Data Sciences Consortium*
*Department of Mathematics and Statistics*
*University of South Florida*
*Tampa, FL 33620-5700, USA*

**Haipeng Shen** HAIPENG@HKU.HK
*School of Business*
*University of Hong Kong*
*Pokfulam, Hong Kong*

**J. S. Marron** MARRON@EMAIL.UNC.EDU
*Department of Statistics and Operations Research*
*University of North Carolina at Chapel Hill*
*Chapel Hill, NC 27599-3260, USA*

**Editor:** massimiliano pontil

## Abstract

A general asymptotic framework is developed for studying consistency properties of principal component analysis (PCA). Our framework includes several previously studied domains of asymptotics as special cases and allows one to investigate interesting connections and transitions among the various domains. More importantly, it enables us to investigate asymptotic scenarios that have not been considered before, and gain new insights into the consistency, subspace consistency and strong inconsistency regions of PCA and the boundaries among them. We also establish the corresponding convergence rate within each region. Under general spike covariance models, the dimension (or number of variables) discourages the consistency of PCA, while the sample size and spike information (the relative size of the population eigenvalues) encourage PCA consistency. Our framework nicely illustrates the relationship among these three types of information in terms of dimension, sample size and spike size, and rigorously characterizes how their relationships affect PCA consistency.
**Keywords:** High dimension low sample size, PCA, Random matrix, Spike model

## 1. Introduction

Principal Component Analysis (PCA) is an important visualization and dimension reduction tool which finds orthogonal directions reflecting maximal variation in the data. This allows the low dimensional representation of data, by projecting data onto these directions. PCA is usually obtained by an eigen decomposition of the sample variance-covariance matrix of the data. Properties of the sample eigenvalues and eigenvectors have been analyzed under several domains of asymptotics.

In this paper, we develop a *general asymptotic framework* to explore interesting transitions among the various asymptotic domains. The general framework includes the traditional asymptotic setups as special cases, and furthermore it allows a careful study of the connections among the various setups. More importantly, we investigate a wide range of interesting scenarios that have not been considered before, and offer new insights into the *consistency* (in the sense that the angle between estimated and population eigen directions tends to 0, or the inner product tends to 1) and *strong-inconsistency* (where the angle tends to $\pi/2$, i.e., the inner product tends to 0) properties of PCA, along with some technically challenging convergence rates.

Existing asymptotic studies of PCA roughly fall into four domains:

(a) the **classical** domain of asymptotics, under which the sample size $n \to \infty$ and the dimension $d$ is fixed (hence the ratio $n/d \to \infty$). For example, see Girshick (1939); Lawley (1956); Anderson (1963, 1984); Jackson (1991).

(b) the **random matrix** theory domain, where both the sample size $n$ and the dimension $d$ increase to infinity, with the ratio $n/d \to c$, a constant mostly assumed to be within $(0, \infty)$. Representative work includes Biehl and Mietzner (1994); Watkin and Nadal (1994); Reimann et al. (1996); Hoyle and Rattray (2003) from the statistical physics literature, as well as Johnstone (2001); Baik et al. (2005); Baik and Silverstein (2006); Onatski (2012); Paul (2007); Nadler (2008); Johnstone and Lu (2009); Lee et al. (2010); Benaych-Georges and Nadakuditi (2011) from the statistics literature.

(c) the **high dimension low sample size (HDLSS)** domain of asymptotics, which is based on the limit, as the dimension $d \to \infty$, with the sample size $n$ being fixed (hence the ratio $n/d \to 0$). HDLSS asymptotics was originally studied by Casella and Hwang (1982), and rediscovered by Hall et al. (2005). PCA has been studied using the HDLSS asymptotics by Ahn et al. (2007); Jung and Marron (2009).

(d) the **increasing signal strength** domain of asymptotics, where $n$, $d$ are fixed and the signal strength tends to infinity. Such a setting is studied in Nadler (2008).

PCA consistency and (strong) inconsistency, defined in terms of angles, are important properties that have been studied before. A common technical device is the spike covariance model, initially introduced by Johnstone (2001). This model has been used in this context by, for example, Nadler (2008); Johnstone and Lu (2009); Jung and Marron (2009). Recently, Ma (2013) formulates sparse PCA (Zou et al., 2006) through iterative thresholding and studies its asymptotic properties under the spike model. An interesting, more general, model has been considered by Benaych-Georges and Nadakuditi (2011).

Under the spike model, the first few eigenvalues are much larger than the others. A *major message of the present paper* is that there are three critical features whose relationships drive the consistency properties of PCA, namely

(1) the *sample size*: the sample size $n$ *encourages* the consistency of the sample eigenvectors, meaning that more samples tend towards more frequent consistency;

(2) the *dimension*: the dimension $d$ *discourages* the consistency of the sample eigenvectors, meaning that higher $d$ tends towards less frequent consistency;

(3) the *spike signal*: the relative sizes of the several leading eigenvalues similarly encourage the consistency.

Our general framework considers increasing sample size $n$, increasing dimension $d$, and increasing spike signal. We clearly characterize how their relationships determine the regions of consistency and strong-inconsistency of PCA, along with the boundary in-between.

Note that the classical domain ((a) above) assumes increasing sample size $n$ while fixing dimension $d$; the random matrix domain ((b) above) assumes increasing sample size $n$ and increasing dimension $d$, while fixing the spike signal; the HDLSS domain ((c) above) fixes the sample size, and increases the dimension and the spike signal; the increasing signal strength domain ((d) above) assumes increasing the spike signal, while fixing the sample size and the dimension; thus each of these three domains is a boundary case of our framework. Our theorems, when restricted to these existing domains of asymptotics, are consistent with known results.

In addition, our theorems go beyond these known results to demonstrate the transitions among the existing domains of asymptotics, and for the first time to the best of our knowledge, enable one to understand interesting connections among them. Finally, we also establish novel results on rates of convergence.

Sections 3 and 4 formally state very general theorems for multiple component spike models. For illustration purposes only, in this section we first consider Examples 1 and 2 under some strong assumptions, which provide intuitive insights regarding the much more general theory presented in Sections 3 and 4. In addition, we use Example 3 to show the application of our theoretical study to the factor model considered by Fan et al. (2013).

For Examples 1 and 2, to better demonstrate the connection with existing results, the three types of features (sample size, dimension, and spike signal) and their relationships are mathematically quantified by two indices, namely the *spike index* $\alpha$ and the *sample index* $\gamma$. Within the context of these examples, we point out the significant contributions of our results in comparison with existing results. The comparisons and connections are graphically illustrated in Figure 1 and discussed below.

**Example 1 Single-component Spike Model** *Assume that $X_1, \dots, X_n$ are sample vectors from a $d$-dimensional distribution with zero mean and covariance matrix $\Sigma$, where the entries of $\Sigma^{-\frac{1}{2}} X_i$ are i.i.d. random variables with zero mean, unit variance and finite fourth moment. (A special case: $X_i$ is from the $d$-dimensional normal distribution $N(0, \Sigma)$). In addition, assume that the sample size $n = d^\gamma$ ($\gamma \geq 0$ is defined as the sample index), and the covariance matrix $\Sigma$ has the following eigenvalues:*

$$\lambda_1 = c_1 d^\alpha, \lambda_2 = \cdots = \lambda_d = 1, \alpha \geq 0,$$

*where the constant $\alpha$ is defined as the spike index.*

*Corollary* **B.2** *in the supplementary materials, when applied to this example, shows that the maximal sample eigenvector is consistent when $\alpha + \gamma > 1$ (grey region in Figure 1(A)), and strongly inconsistent when $0 \leq \alpha + \gamma < 1$ (white triangle in Figure 1(A)). These very general new results nicely connect with many existing ones:*

- ***Previous Results I - the classical domain**:*

  *Under the normal assumption, Theorem 1 of Anderson (1963) implied that for fixed dimension $d$ and finite eigenvalues, when the sample size $n \to \infty$ (i.e. $\gamma \to \infty$, the limit on the vertical axis), the maximal sample eigenvector is consistent. This case is the upper left corner of Figure 1(A).*

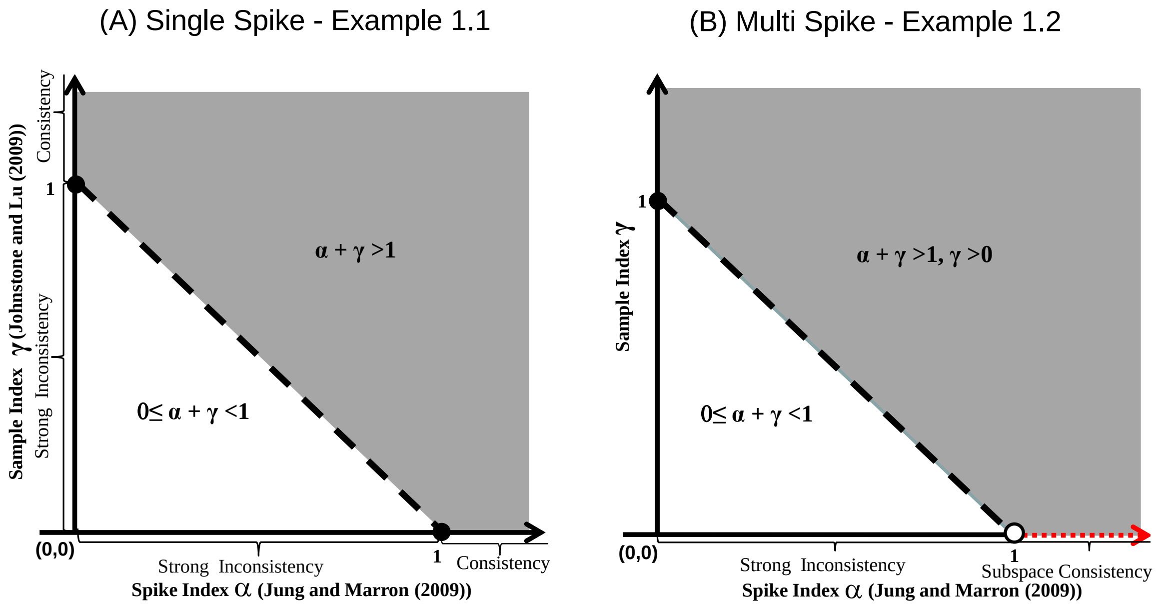


Figure 1: General consistency and strong inconsistency regions for PCA, as a function of the spike index $\alpha$ and the sample index $\gamma$. Panel (A) - single spike model in Example 1: PCA is consistent in the grey region ($\alpha + \gamma > 1$), and strongly inconsistent on the white triangle ($0 \leq \alpha + \gamma < 1$). Panel (B) - multiple spike model in Example 2: the first $m$ sample PCs are consistent in the grey region ($\alpha + \gamma > 1, \gamma > 0$), subspace consistent on the dotted line ($\alpha > 1, \gamma = 0$) on the horizontal axis, and strongly inconsistent on the white triangle ($0 \leq \alpha + \gamma < 1$).

- ***Previous Results II - the random matrix domain:***

  (a) *Assuming normality, the results of Johnstone and Lu (2009) appear on the vertical axis in Panel (A) where the spike index $\alpha = 0$ (as they fix the spike information): the first sample eigenvector is consistent when the sample index $\gamma > 1$ and strongly inconsistent when $\gamma < 1$.*

  (b) *Again, under the normal assumption, Nadler (2008) explored the interesting boundary case of $\alpha = 0, \gamma = 1$ (i.e. $\frac{d}{n} \to c$ for a constant $c$) and showed that $< \hat{u}_1, u_1 >^2 \xrightarrow{\text{a.s}} \frac{((\lambda_1-1)^2-c)_+}{(\lambda_1-1)^2+c(\lambda_1-1)}$, where $\hat{u}_1$ and $u_1$ are the first sample and population eigenvector. This result appears in Panel (A) as the single solid circle $\gamma = 1$ on the vertical axis. Our general framework doesn't cover this boundary case and this boundary result is a complement of our theoretical results.*

- ***Previous Results III - the HDLSS domain:***

  (a) *The theorems of Jung and Marron (2009) are represented on the horizontal axis in Panel (A) when the sample index $\gamma = 0$ (as they fix the sample size): the maximal sample eigenvector is consistent with the first population eigenvector when the spike index $\alpha > 1$ and strongly inconsistent when $\alpha < 1$.*

(b) *Under the normal assumption, Jung et al. (2012) deeply explored limiting behavior at the boundary* $\alpha = 1, \gamma = 0$ *(i.e.* $\frac{d}{\lambda_1} \to c$ *for a constant c) and showed that* $< \hat{u}_1, u_1 >^2 \Rightarrow \frac{A}{A+c}$*, where "* $\Rightarrow$ *" means convergence in distribution and* $A \sim \chi^2_n$*, the chi-squared distribution with n degrees of freedom. This result appears in Panel (A) as the single solid circle* $\alpha = 1$ *on the horizontal axis. This boundary case is again a complement of our general framework.*

- ***Our Results*** *hence nicely connect existing domains of asymptotics, and give a much more complete characterization for the regions of PCA consistency, subspace consistency, and strong inconsistency. We also investigate asymptotic properties of the other sample eigenvectors and all the sample eigenvalues.*

**Example 2 Multiple-component Spike Model** *Assume that the covariance matrix* $\Sigma$ *in Example 1 has the following eigenvalues:*

$$\lambda_j = \begin{cases} c_j d^\alpha & \text{if } j \le m, \\ 1 & \text{if } j > m, \end{cases} \qquad \alpha \ge 0,$$

*where m is a finite positive integer, the constants* $c_j, j = 1, \cdots, m$*, are positive and satisfy that* $c_j > c_{j+1} > 1$*,* $j = 1, \cdots, m-1$*.*

*Corollary* **B.1** *in the supplementary materials, when applied to this example, shows that the first m sample eigenvectors are individually consistent with corresponding population eigenvectors when* $\alpha + \gamma > 1, \gamma > 0$ *(the grey region in Figure 1(B)), instead of being subspace consistent (Jung and Marron, 2009), and strongly inconsistent when* $\alpha + \gamma < 1$ *(the white triangle in Panel (B)). This very general new result connects with many others in the existing literature:*

- ***Previous Results I - the classical domain****:*

  *Assuming normality, Theorem 1 of Anderson (1963) implied that for fixed dimension d and finite eigenvalues, when the sample size* $n \to \infty$ *(i.e.* $\gamma \to \infty$*, the limit on the vertical axis), the first m sample eigenvectors are consistent, while the other sample eigenvectors are subspace consistent. This case is the upper left corner of Figure 1(B).*

- ***Previous Results II - the random matrix domain****:*

  *The following results are under the normal assumption. Paul (2007) explored asymptotic properties of the first m eigenvectors and eigenvalues in the interesting boundary case of* $\alpha = 0, \gamma = 1$*, i.e.,* $\frac{d}{n} \to c$ *with* $c \in (0, 1)$ *and showed that* $< \hat{u}_j, u_j >^2 \xrightarrow{\text{a.s}} \frac{((\lambda_j - 1)^2 - c)_+}{(\lambda_j - 1)^2 + c(\lambda_j - 1)}$ *for* $j = 1, \cdots, m$*. This result appears in Panel (B) as the solid circle* $\gamma = 1$ *on the vertical axis. This boundary case is a complement of our results for multiple spike models with distinct eigenvalues (Section* **B.1** *of the supplementary materials). Paul and Johnstone (2012) considered a similar framework but from a minimax risk analysis perspective. Nadler (2008); Johnstone and Lu (2009) did not study multiple spike models.*

- ***Previous Results III - the HDLSS domain:***

  *The theorems of Jung and Marron (2009) are valid on the horizontal axis in Panel (B) where the sample index $\gamma = 0$. In particular, for this example, their results showed that the first $m$ sample eigenvectors are not respectively consistent with the corresponding population eigenvectors when the spike index $\alpha > 1$ (the horizontal dotted red line segment), instead they are subspace consistent with their corresponding population eigenvectors, and are strongly inconsistent when the spike index $\alpha < 1$ (the horizontal solid line segment). They and Jung et al. (2012) did not study the asymptotic behavior on the boundary - the single open circle $(\alpha = 1, \gamma = 0)$ on the horizontal axis.*

- ***Our Results*** *cover the classical domain, and are stronger than what Jung and Marron (2009) obtained: the increasing sample size enables us to separate out the first few leading eigenvectors and characterize individual consistency, while only subspace consistency was obtained by Jung and Marron (2009).*

**Example 3 The Factor Model of Fan et al. (2013)** *Consider the following model:*

$$\mathbf{y}_t = \mathbf{B}\mathbf{f}_t + \mathbf{E}_t,$$

*where $\mathbf{y}_t = (y_{t,1}, \ldots, y_{t,d})^T$ is the $d$-dimensional response vector, $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_d)^T$ is the $d \times m$ ($m$ is fixed) loading matrix, $\mathbf{f}_t$ is the $m \times 1$ vector of common factors, and $\mathbf{E}_t = (e_{t,1}, \ldots, e_{t,d})^T$ is the $d$-dimensional noise vector, $t = 1, \ldots, \mathbb{T}$. The noise vector $\mathbf{E}_t$ is independent of $\mathbf{f}_t$. Then the population covariance matrix of $\mathbf{y}_t$ is*

$$\Sigma = \mathbf{B}\mathrm{cov}(\mathbf{f}_t)\mathbf{B}^T + \Sigma_{\mathrm{E}},$$

*where $\Sigma_{\mathrm{E}}$ is the covariance matrix of $\mathbf{E}_t$. Fan et al. (2013) assumes that the first $m$ eigenvalues of $\mathbf{B}\mathrm{cov}(\mathbf{f}_t)\mathbf{B}^T$ increase with $d$ as $d \to \infty$, whereas all the eigenvalues of $\Sigma_{\mathrm{E}}$ are bounded. It then follows that $\lambda_m(\Sigma) \gg \lambda_{m+1}(\Sigma) \asymp \cdots \asymp \lambda_d(\Sigma) \asymp 1$, as $d \to \infty$. Then our theorems are applicable to this factor model when $\mathbf{f}_1, \ldots, \mathbf{f}_{\mathbb{T}}$ is i.i.d., and $\mathbf{E}_1, \ldots, \mathbf{E}_{\mathbb{T}}$ is i.i.d..*

*Under the above assumptions of the factor model, we have $d/(\mathbb{T}\lambda_m(\Sigma)) \to 0$. Then according to our Theorem 1 (together with the third comment after the theorem), the first $m$ sample eigenvalues and eigenvectors are consistent. On the other hand, Fan et al. (2013) proposed the consistent principal orthogonal complement thresholding (POET) estimator for the covariance matrix $\Sigma$, which is obtained by keeping the first $m$ sample eigenvalues and eigenvectors, and thresholding the residual sample matrix. Hence, our theorem offers another theoretical support on the consistency of their POET estimator.*

The rest of the paper is organized as follows. Section 2 first introduces our notations and relevant consistency concepts. Section 3 studies the PCA asymptotics of spike models with increasing sample size $n$. We state the main results of our paper - Theorem 1 for multiple-component spike models where the dominating eigenvalues are inseparable. Theorem 2 in Section 4 then is about the HDLSS asymptotics of PCA, where the sample size $n$ is fixed, for spike models with inseparable eigenvalues. Section 5 contains some discussions about the asymptotic properties of PCA when some eigenvalues equal to zero and the challenges to obtain non-asymptotic results. Section 7 contains the technical proofs of Theorem 1 and the relevant lemmas. The supplementary materials contain the corresponding corollaries of Theorems 1 and 2, for multiple-spike models with distinct eigenvalues and single spike models, along with the proofs of Theorem 2 and all the corollaries.

## 2. Notations and Concepts

We now introduce some necessary notations, and define consistency concepts relevant for our asymptotic study.

### 2.1 Notation

Let the population covariance matrix be $\Sigma$, whose eigen decomposition is

$$\Sigma = U\Lambda U^T,$$

where $\Lambda$ is the diagonal matrix of population eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$, and $U$ is the matrix of the corresponding eigenvectors $U = [u_1, \ldots, u_d]$.

As in Jung and Marron (2009), assume that $X_1, \ldots, X_n$ are i.i.d. $d$-dimensional random sample vectors and have the following representation

$$X_i = \sum_{j=1}^{d} \lambda_j^{\frac{1}{2}} z_{i,j} u_j, \tag{1}$$

where the $z_{i,j}$'s are i.i.d. random variables with zero mean, unit variance, and finite fourth moment. An important special case is that they follow the standard normal distribution.

**Assumption 1** *$X_1, \ldots, X_n$ are a random sample having the distribution of* (1).

Jung and Marron (2009) assumes that

$$Z_i = (z_{i,1}, \cdots, z_{i,d})^T, \quad i = 1, \cdots, n, \tag{2}$$

are independent and the elements $z_{i,j}$ within $Z_i$ are $\rho$-mixing. This assumption leads to the convergence in probability results under the HDLSS domain in Jung and Marron (2009). Here we assume that the elements $z_{i,j}$ within $Z_i$ are also independent. This helps to get the almost sure convergence results under our general framework, which includes the HDLSS domain. Assumption 1 is necessary to satisfy the conditions of Lemma 1 - the Bai-Yin's law (Bai and Yin, 1993), which is important for our results, for example, Theorem 1.

Denote the sample covariance matrix by $\hat{\Sigma} = n^{-1}XX^T$, where $X = [X_1, \ldots, X_n]$. Note that $\hat{\Sigma}$ can also be decomposed as

$$\hat{\Sigma} = \hat{U}\hat{\Lambda}\hat{U}^T, \tag{3}$$

where $\hat{\Lambda}$ is the diagonal matrix of sample eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_d$, and $\hat{U}$ is the matrix of corresponding sample eigenvectors where $\hat{U} = [\hat{u}_1, \ldots, \hat{u}_d]$.

Below we introduce asymptotic notations that will be used in our theoretical studies. Let $\tau$ stand for either $n$ or $d$, depending on the context. Assume that $\{\xi_\tau : \tau = 1, \ldots, \infty\}$ is a sequence of random variables, and $\{a_\tau : \tau = 1, \ldots, \infty\}$ is a sequence of constant values.

- Denote $\xi_\tau = \mathrm{o}_{\mathrm{a.s}}(a_\tau)$ if $\lim_{\tau\to\infty} \frac{\xi_\tau}{a_\tau} = 0$ almost surely.

- Denote $\xi_\tau = \mathrm{O}_{\mathrm{a.s}}(a_\tau)$ if $\overline{\lim}_{\tau\to\infty} \left|\frac{\xi_\tau}{a_\tau}\right| \leq M$, where $M$ is a positive constant.

- Denote almost surely $\xi_\tau \asymp a_\tau$ if $c_2 \leq \underline{\lim}_{\tau\to\infty} \frac{\xi_\tau}{a_\tau} \leq \overline{\lim}_{\tau\to\infty} \frac{\xi_\tau}{a_\tau} \leq c_1$ almost surely, for two constants $c_1 \geq c_2 > 0$.

In addition, we introduce the following notions to help understand the assumptions on the population eigenvalues in our theorems and corollaries. Assume that $\{a_\tau : \tau = 1, \ldots, \infty\}$ and $\{b_\tau : \tau = 1, \ldots, \infty\}$ are two sequences of real valued numbers.

- Denote $a_\tau \gg b_\tau$ if $\lim_{\tau\to\infty} \frac{b_\tau}{a_\tau} = 0$.

- Denote $a_\tau \asymp b_\tau$ if $c_2 \leq \underline{\lim}_{\tau\to\infty} \frac{a_\tau}{b_\tau} \leq \overline{\lim}_{\tau\to\infty} \frac{a_\tau}{b_\tau} \leq c_1$ for two constants $c_1 \geq c_2 > 0$.

### 2.2 Concepts

We now list several concepts about consistency and strong inconsistency, some of which are modified from the related concepts in Jung and Marron (2009) and Shen et al. (2013).

Let $H$ be an index set, e.g. $H = \{m+1, \cdots, d\}$, and then denote $S = \mathrm{span}\{u_k, k \in H\}$ as the linear span generated by $\{u_k, k \in H\}$. Define angle$(\hat{u}_j, S)$ as the angle between the estimator $\hat{u}_j$ and the subspace $S$, which is the angle between the estimator and its projection onto the subspace (Jung and Marron, 2009). For further clarification, we provide a graphical illustration of the angle in Section B of the supplement (Shen et al., 2015). As pointed out earlier, let $\tau$ stand for either $n$ or $d$, depending on the context.

- If as $\tau \to \infty$, angle$(\hat{u}_j, S) \stackrel{\mathrm{a.s}}{\to} 0$, then $\hat{u}_j$ is **subspace consistent** with $S$. If $H$ only includes one index $j$ such that $S = \mathrm{span}\{u_j\}$, then angle$(\hat{u}_j, S) \stackrel{\mathrm{a.s}}{\to} 0$ is equivalent to $| < \hat{u}_j, u_j > | \stackrel{\mathrm{a.s}}{\to} 1$, and $\hat{u}_j$ is **consistent** with $u_j$.

- If as $\tau \to \infty$, $| < \hat{u}_j, u_j > | \stackrel{\mathrm{a.s}}{\to} 0$, then $\hat{u}_j$ is **strongly inconsistent** with $u_j$.

## 3. Cases with increasing sample size $n$

We study spike models with increasing sample size $n \to \infty$ in this section. As such, the eigenvalues $\lambda_j$ and the dimension $d$ depend on the sample size $n$, and will be denoted as $\lambda_j^{(n)}$ and $d(n)$ throughout this section. They can be viewed as sequences of constant values indexed by $n$. This section considers multiple-component spike models with inseparable eigenvalues and presents the main theorem of our paper. Section **B** of the supplementary materials reports the corollaries for multiple component spike models with distinct eigenvalues and single spike models.

We consider multiple spike models with $m$ (a finite integer) dominating eigenvalues. These $m$ eigenvalues can be grouped into $r$ *tiers*, where the eigenvalues within the same tier have the same limit. To fixed ideas, the first $m$ eigenvalues are grouped into $r$ tiers where there are $q_l (> 0)$ eigenvalues in the $l$th tier with $\sum_{l=1}^{r} q_l = m$. Define $q_0 = 0$,

$q_{r+1} = d(n) - \sum_{l=1}^{r} q_l$, and the index set of the eigenvalues in the $l$th tier as

$$H_l = \left\{ \sum_{k=0}^{l-1} q_k + 1, \sum_{k=0}^{l-1} q_k + 2, \cdots, \sum_{k=0}^{l-1} q_k + q_l \right\}, \quad l = 1, \cdots, r+1. \tag{4}$$

Assume the eigenvalues in the $l$th tier have the same limit $\delta_l^{(n)} (> 0)$, i.e.

**Assumption 2** $\lim_{n\to\infty} \frac{\lambda_j^{(n)}}{\delta_l^{(n)}} = 1,\ j \in H_l, l = 1, \cdots, r.$

According to the above assumption, the eigenvalues that are in the same tier will have the same limit as $n$ goes to infinity. As a result, we can show that the corresponding sample eigenvectors can not be consistently estimated individually. This motives us to consider subspace consistency. In addition, we assume that the first $m$ population eigenvalues from different tiers are asymptotically different, and dominate the additional population eigenvalues beyond the first $r$ tiers that have the same limit $c_\lambda$:

**Assumption 3** *as* $n \to \infty$*,* $\delta_1^{(n)} > \cdots > \delta_r^{(n)} > \lambda_{m+1}^{(n)} \to \cdots \to \lambda_{d(n)}^{(n)} \to c_\lambda > 0.$

For $i < j$, $\delta_i^{(n)} > \delta_j^{(n)}$ means that $\lim_{n\to\infty} \frac{\delta_i^{(n)}}{\delta_j^{(n)}} > 1$. This assumption allows $\delta_i^{(n)} \to \infty$ and $\delta_i^{(n)} \gg \delta_j^{(n)}$, which is not the case in Paul (2007). Regarding the constant $c_\lambda$, the second remark after Theorem 1 discusses what happens when $c_\lambda = 0$.

The above assumptions cover a general class of multiple spike models with tiered eigenvalues. A simple special case is the one where the eigenvalue matrix $\Lambda$ is block diagonal: for $1 \le h \le r$, the $h$-th block of $\Lambda$ is $\lambda_h^{(n)} I_{q_h}$ where $I_{q_h}$ is the $q_h \times q_h$ identity matrix, with

$$\lambda_1^{(n)} > \lambda_2^{(n)} > \ldots > \lambda_r^{(n)}, \qquad q_1 + q_2 + \ldots + q_r = m < d;$$

and the last block of $\Lambda$ is $c_\lambda I_{d(n)-m}$ with $c_\lambda < \lambda_r^{(n)}$.

Under the above setup, Theorem 1 shows that the eigenvector estimates are either subspace consistent with the linear space spanned by the population eigenvectors, or strongly inconsistent. As discussed in the Introduction, Theorem 1 considers the delicate balance among the *sample size* $n$, the *spike signal* $\delta_l^{(n)}$, and the *dimension* $d(n)$, and characterize the various PCA consistency and strong-inconsistency regions. The three scenarios of Theorem 1 are arranged in the order of a decreasing amount of signal:

- Theorem 1(a): If the amount of signal dominates the amount of noise up to the $r$th tier, i.e. $\frac{d(n)}{n\delta_r^{(n)}} \to 0$, then the estimates for the eigenvectors in the first $r$ tiers are subspace consistent, and the estimates for the higher order eigenvectors are also subspace consistent (but) at a different rate;
- Theorem 1(b): Otherwise, if the amount of signal dominates the amount of noise only up to the $h$th tier ($1 \le h < r$), i.e. $\frac{d(n)}{n\delta_h^{(n)}} \to 0$ and $\frac{d(n)}{n\delta_{h+1}^{(n)}} \to \infty$, then the estimates for the eigenvectors in the first $h$ tiers are subspace consistent, and the estimates for the other eigenvectors are strongly-inconsistent;

- Theorem 1(c): Finally, if the amount of noise always dominates, i.e. $\frac{d(n)}{n\lambda_1^{(n)}} \to \infty$, then the sample eigenvalues are asymptotically indistinguishable, and the sample eigenvectors are strongly inconsistent.

Before stating Theorem 1, we first introduce several notations. Define the subspace $S_l = \text{span}\{u_k, k \in H_l\}$ for $l = 1, \cdots, r+1$ and denote $\delta_0^{(n)} = \infty$ for every $n$.

**Theorem 1** *Under Assumptions 1, 2 and 3, as $n \to \infty$, the following results hold.*

*(a) If $\frac{d(n)}{n\delta_r^{(n)}} \to 0$, then $\frac{\hat{\lambda}_j}{\lambda_j^{(n)}} \xrightarrow{\text{a.s}} 1$, $j = 1, \cdots, m$, and angle$(\hat{u}_j, S_l) = \mathrm{o}_{\text{a.s}}\left(\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{\delta_{l+1}^{(n)}}{\delta_l^{(n)}}\right\}^{\frac{1}{2}}\right)$, $j \in H_l$, $l = 1, \cdots, r-1$. In addition,*

- *If $\frac{d(n)}{n} \to 0$, then angle$(\hat{u}_j, S_l) = \mathrm{o}_{\text{a.s}}\left(\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{1}{\delta_l^{(n)}}\right\}^{\frac{1}{2}}\right)$, $j \in H_l$ for $l = r$, and $\mathrm{o}_{\text{a.s}}\left(\left\{\frac{1}{\delta_r^{(n)}}\right\}^{\frac{1}{2}}\right)$ for $l = r+1$.*
- *If $\frac{d(n)}{n} \to c$, $0 < c \leq \infty$, then angle$(\hat{u}_j, S_l) = \mathrm{o}_{\text{a.s}}\left(\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}}\right\}^{\frac{1}{2}}\right) \vee \mathrm{O}_{\text{a.s}}\left(\left\{\frac{d(n)}{n\delta_l^{(n)}}\right\}^{\frac{1}{2}}\right)$, $j \in H_l$ for $l = r$, and $\mathrm{O}_{\text{a.s}}\left(\left\{\frac{d(n)}{n\delta_r^{(n)}}\right\}^{\frac{1}{2}}\right)$ for $l = r+1$.*

*(b) If $\frac{d(n)}{n\delta_h^{(n)}} \to 0$ and $\frac{d(n)}{n\delta_{h+1}^{(n)}} \to \infty$, where $1 \leq h < r$, then $\frac{\hat{\lambda}_j}{\lambda_j^{(n)}} \xrightarrow{\text{a.s}} 1$, $j \in H_l, l = 1, \cdots, h$, and the other non-zero $\frac{n\hat{\lambda}_j}{d(n)} \xrightarrow{\text{a.s}} c_\lambda$. In addition, angle$(\hat{u}_j, S_l) = \mathrm{o}_{\text{a.s}}\left(\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{\delta_{l+1}^{(n)}}{\delta_l^{(n)}}\right\}^{\frac{1}{2}}\right)$, $j \in H_l$ for $l = 1, \cdots, h-1$, and $\mathrm{o}_{\text{a.s}}\left(\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\}^{\frac{1}{2}}\right) \vee \mathrm{O}_{\text{a.s}}\left(\left\{\frac{d(n)}{n\delta_h^{(n)}}\right\}^{\frac{1}{2}}\right)$ for $l = h$. Finally, $| < \hat{u}_j, u_j > | = \mathrm{O}_{\text{a.s}}\left(\left\{\frac{n\lambda_j^{(n)}}{d(n)}\right\}^{\frac{1}{2}}\right)$, $j \in H_l, l = h+1, \cdots, r$, and $\mathrm{O}_{\text{a.s}}\left(\left\{\frac{n}{d(n)}\right\}^{\frac{1}{2}}\right)$, $j > m$.*

*(c) If $\frac{d(n)}{n\delta_1^{(n)}} \to \infty$, then the non-zero $\frac{n\hat{\lambda}_j}{d(n)} \xrightarrow{\text{a.s}} c_\lambda$. In addition, $| < \hat{u}_j, u_j > | = \mathrm{O}_{\text{a.s}}\left(\left\{\frac{n\lambda_j^{(n)}}{d(n)}\right\}^{\frac{1}{2}}\right)$, $j = 1, \cdots, m$, and $\mathrm{O}_{\text{a.s}}\left(\left\{\frac{n}{d(n)}\right\}^{\frac{1}{2}}\right)$, $j > m$.*

The following comments can be made for the results of Theorem 1.

- Note that, for $j \in H_1$, the subspace consistency rate for $\hat{u}_j$ is $\left\{\frac{\delta_2^{(n)}}{\delta_1^{(n)}}\right\}^{\frac{1}{2}}$. By defining $\delta_0^{(n)} = \infty$, the consistency rate expression $\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{\delta_{l+1}^{(n)}}{\delta_l^{(n)}}\right\}^{\frac{1}{2}}$ remains valid for $l = 1$.

- If $c_\lambda = 0$ in Assumption 3, then that assumption can be rewritten as

$$\delta{*}_1^{(n)} > \cdots > \delta{*}_r^{(n)} > \lambda{*}_{m+1}^{(n)} \to \cdots \to \lambda{*}_{d(n)}^{(n)} = 1,$$

  where $\delta{*}_j^{(n)} = \frac{\delta_j^{(n)}}{\lambda_{d(n)}^{(n)}}$ and $\lambda{*}_j^{(n)} = \frac{\lambda_j^{(n)}}{\lambda_{d(n)}^{(n)}}$. We comment that the asymptotic properties of $\hat{u}_j$ then depend on the rescaled eigenvalues $\lambda{*}_j^{(n)}$, instead of the raw eigenvalues $\lambda_j^{(n)}$. In particular, with $c_\lambda = 0$, Theorem 1 can be slightly modified by replacing $\delta_j^{(n)}$ with $\delta{*}_j^{(n)}$, "$\frac{n\hat{\lambda}_j}{d(n)} \xrightarrow{\text{a.s}} c_\lambda$" with "$\frac{n\hat{\lambda}_j}{d(n)\lambda_{d(n)}^{(n)}} \xrightarrow{\text{a.s}} 1$", and the strongly inconsistency rate $\left\{\frac{n\lambda_j^{(n)}}{d(n)}\right\}^{\frac{1}{2}}$ with $\left\{\frac{n\lambda{*}_j^{(n)}}{d(n)}\right\}^{\frac{1}{2}}$, respectively.

- In Assumption 3, if there is a big gap between $\delta_r^{(n)}$ and $\lambda_{m+1}^{(n)}$ such that $\delta_r^{(n)} \gg \lambda_{m+1}^{(n)}$, then $\lambda_{m+1}^{(n)} \to \cdots \to \lambda_{d(n)}^{(n)} \to c_\lambda$ can be weakened to $\lambda_{m+1}^{(n)} \asymp \cdots \asymp \lambda_{d(n)}^{(n)} \asymp 1$. It follows that the consistency results of the first $r$ tiers of sample eigenvalues in Scenario (a) or the first $h$ tiers in Scenario (b) remain the same, while all other results of the form "$\xrightarrow{\text{a.s}}$" for the sample eigenvalues should be replaced by almost surely "$\asymp$". The results for the sample eigenvectors remain the same.

- One needs $\lambda_{m+1}^{(n)} \to \cdots \to \lambda_{d(n)}^{(n)} \to c_\lambda$, or $\lambda_{m+1}^{(n)} \asymp \cdots \lambda_{d(n)}^{(n)} \asymp 1$, to obtain general convergence results for the non-spike sample eigenvalues $\hat{\lambda}_j$, $j > m$, under the wide range of scenarios: $\frac{d(n)}{n} \to 0$, $\frac{d(n)}{n} \to \infty$, or $\lim_{n\to\infty} \frac{d(n)}{n} = c$ $(0 < c < \infty)$. When one focusses only on the spike eigenvalues, a weaker assumption, such as the slowly decaying non-spike eigenvalues assumed by Bai and Yao (2012), is sufficient. Then, the spike condition $\delta_r^{(n)} \gg \lambda_{m+1}^{(n)}$ is enough to generate the consistency properties of $\hat{\lambda}_j$ and $\hat{u}_j$, $j \le m$ in Scenario (a). In that case, the behaviors of the other sample eigenvalues and eigenvectors are scenario specific, depending on whether $\frac{d(n)}{n} \to 0$, $\frac{d(n)}{n} \to \infty$, or $\lim_{n\to\infty} \frac{d(n)}{n} = c$ $(0 < c < \infty)$.

- The cases covered by Theorem 1 are not studied in Paul (2007), where the eigenvalues are considered to be individually estimable.

- In Theorem 1, the dimension $d$ can be fixed. In addition, suppose $\infty > \delta_1^{(n)} > \cdots > \delta_r^{(n)} > \lambda_{m+1}^{(n)} \to \cdots \to \lambda_d^{(n)} \to c_\lambda$, and the eigenvalues satisfy Assumption 2. Then, the results of Theorem 1(a) are consistent with the classical asymptotic subspace consistency results implied by Theorem 1 of Anderson (1963).

## 4. Cases with fixed $n$

This section studies spike models when the sample size $n$ is fixed. Now the eigenvalues are denoted as $\lambda_j^{(d)}$, a sequence indexed by the dimension $d$. We first report here the theoretical results for spike models with inseparable eigenvalues. The corresponding results for models with distinct eigenvalues are presented in Section **C** of the supplementary materials.

Theorem 2 summarizes the results for spike models with tiered eigenvalues. In comparison with Jung and Marron (2009), we make more general assumptions on the population eigenvalues, and obtain the convergence rate results; furthermore, we obtain almost sure convergence, instead of convergence in probability.

Assume that as $d \to \infty$, the first $m$ eigenvalues fall into $r$ tiers, where the eigenvalues in the same tier are asymptotically equivalent, as stated in the following assumption:

**Assumption 4** *for fixed $n$, as $d \to \infty$, $\lambda_j^{(d)} \asymp \delta_l^{(d)}$, $j \in H_l$, $l = 1, \cdots, r$.*

Different from Assumption 2 for diverging sample size $n$, now with a fixed $n$, the eigenvalues within the same tier are assumed to be of the same order, rather than of the same limit when $n$ increases to $\infty$. As we will see below in Theorem 2, one can no longer separately estimate the eigenvalues of the same order when $n$ is fixed, which is feasible with an increasing $n$ as long as they do not have the same limit as shown in Theorem 1.

In addition, we assume that the population eigenvalues from different tiers are of different orders and dominate the higher-order eigenvalues which are asymptotically equivalent:

**Assumption 5** *for fixed $n$, as $d \to \infty$, $\delta_1^{(d)} \gg \cdots \gg \delta_r^{(d)} \gg \lambda_{m+1}^{(d)} \asymp \cdots \asymp \lambda_d^{(d)} \asymp 1$.*

Note that for fixed $n$ and $d \to \infty$, the assumption $\delta_l^{(d)} > \delta_{l+1}^{(d)}$ can not guarantee asymptotic separation of the corresponding sample eigenvalues $\hat{\lambda}_j$ for $j \in H_l$ and $j \in H_{l+1}$. Thus, we need to replace Assumption 3 with Assumption 5 in order to asymptotically separate the first $r$ subgroups of sample eigenvalues.

Before formally stating Theorem 2, we first introduce several notations. Denote $\delta_0^{(d)} = \infty$ for every $d$, which is used to describe the subspace consistent rates. Consider the $z_{i,j}$ in (1), and let

$$\widetilde{Z}_j = (z_{1,j}, \cdots, z_{n,j})^T, \quad j = 1, \cdots, d. \tag{5}$$

Define

$$K = \lim_{d\to\infty} \frac{\sum_{j=m+1}^{d} \lambda_j^{(d)}}{d} \quad \text{and} \quad A_l^* = \frac{1}{n} \sum_{k \in H_l} \widetilde{Z}_k \widetilde{Z}_k^T, \quad l = 1, \cdots, r, \tag{6}$$

which are used to describe the asymptotic properties of the sample eigenvalues.

**Theorem 2** *Under Assumptions 1, 4 and 5, for fixed $n$, as $d \to \infty$, the following results hold.*

*(a) If $\frac{d}{\delta_h^{(d)}} \to 0$ and $\frac{d}{\delta_{h+1}^{(d)}} \to \infty$, where $1 \le h \le r$, then for $j \in H_l$, $l = 1, \cdots, h$, almost surely*

$$\lambda_{\min}(A_l^*) \times \min_{k \in H_l} \lambda_k^{(d)} \le \hat{\lambda}_j \le \lambda_{\max}(A_l^*) \times \max_{k \in H_l} \lambda_k^{(d)}, \tag{7}$$

*and the other non-zero $\hat{\lambda}_j$ satisfy $\frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{a.s}} K$. In addition, angle$(\hat{u}_j, S_l) = \mathrm{o}_{\text{a.s}}\left(\left\{\frac{\delta_l^{(d)}}{\delta_{l-1}^{(d)}} \vee \frac{\delta_{l+1}^{(d)}}{\delta_l^{(d)}}\right\}^{\frac{1}{2}}\right)$, $j \in H_l$ for $l = 1, \cdots, h-1$, and $\mathrm{o}_{\text{a.s}}\left(\left\{\frac{\delta_h^{(d)}}{\delta_{h-1}^{(d)}}\right\}^{\frac{1}{2}}\right) \vee \mathrm{O}_{\text{a.s}}\left(\left\{\frac{d}{\delta_h^{(d)}}\right\}^{\frac{1}{2}}\right)$ for $l = h$. Finally, $| < \hat{u}_j, u_j > | = \mathrm{O}_{\text{a.s}}\left(\left\{\frac{\lambda_j^{(d)}}{d}\right\}^{\frac{1}{2}}\right)$, $j \in H_l$, $l = h+1, \cdots, r$, and $\mathrm{O}_{\text{a.s}}\left(\left\{\frac{1}{d}\right\}^{\frac{1}{2}}\right)$, $j > m$.*

*(b) If* $\frac{d}{\delta_1^{(d)}} \to \infty$*, then the non-zero* $\frac{n\hat{\lambda}_j}{d} \stackrel{\text{a.s}}{\to} K$*. In addition,* $| < \hat{u}_j, u_j > | = \mathrm{O}_{\text{a.s}}\left(\left\{\frac{\lambda_j^{(d)}}{d}\right\}^{\frac{1}{2}}\right)$*,* $j = 1, \cdots, m$*, and* $\mathrm{O}_{\text{a.s}}\left(\left\{\frac{1}{d}\right\}^{\frac{1}{2}}\right)$*,* $j > m$*.*

The following comments can be made about the results of Theorem 2.

- Even if the non-spike eigenvalues $\lambda_j^{(d)}$, $j > m$, decay slowly, the condition $\delta_1^{(d)} \gg \cdots \gg \delta_r^{(d)} \gg \lambda_{m+1}^{(d)}$ can still guarantee the same properties for $\hat{\lambda}_j$ and $\hat{u}_j$, $j \in H_l$, $l \leq h$, in Scenario (a).
- Assumption 1 assumes that the $z_{i,j}$'s are i.i.d. rather than $\rho$-mixing as in Jung and Marron (2009). Thus, convergence in probability in Jung and Marron (2009) is strengthened to almost sure convergence here.

## 5. Discussions

Throughout the paper, we assume that the small eigenvalues have the same limit or are of the same order as 1, i.e. $\lambda_{m+1}^{(n)} \to \cdots \to \lambda_{d(n)}^{(n)} \to c_\lambda$ or $\lambda_{m+1}^{(n)} \asymp \cdots \asymp \lambda_{d(n)}^{(n)} \asymp 1$. In fact, this is a convenient choice. Our results remain valid when these small eigenvalues are not of the same order, and even when some of them are 0. For example, suppose $\lambda_{d_1+1} = \cdots = \lambda_d = 0$ for $m + 1 < d_1 < d$. As shown in Section **E** of the supplementary material (Shen et al., 2015), the asymptotic properties of PCA are independent of the basis choice for the $d$-dimensional space. If the population eigenvectors $u_j$, $j = 1, \ldots, d$, are chosen as the basis of the $d$-dimensional space, the population covariance matrix becomes

$$\Sigma = \Lambda = \begin{pmatrix} \Lambda_1 & 0_{d_1 \times (d-d_1)} \\ 0_{(d-d_1)\times d_1} & 0_{(d-d_1)\times(d-d_1)} \end{pmatrix}, \quad \text{where} \quad \Lambda_1 = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{d_1} \end{pmatrix},$$

and $0_{k\times l}$ is the $k$-by-$l$ zero matrix. Then, the asymptotic properties of PCA under the population covariance matrix $\Sigma$ is the same as those under the covariance matrix $\Lambda_1$. Therefore, we only need to replace the dimension $d$ by the *effective* dimension $d_1$, and all the earlier results remain valid.

It would be interesting to explore non-asymptotic results under our general framework. There have been interesting relevant progresses made recently. Koltchinskii and Lounici (2016, 2015) consider a general framework that encompasses the spike model with fixed spike sizes, and establish theorems about non-asymptotic properties of sample eigenvalues/eigenvectors under either Gaussian or centered subgaussian assumption. These results pave the way to study non-asymptotic properties under our framework where the spike sizes are allowed to grow and we only assume finite fourth moment.

## 6. Acknowledgements

We acknowledge support from Startup Fund of University of South Florida, US National Science Foundation Grants DMS-1106912 and DMS-1407655, and National Institutes of

Health Challenge Grant 1 RC1 DA029425-01. We thank the action actor and the four referees for their helpful comments and suggestions.

## 7. Proofs

We now provide the detailed proof for Theorem 1. To save space, the proofs for Theorem 2 and the corresponding corollaries of the two theorems (which are often similar, and simpler) are provided in the supplement (Shen et al., 2015). We first provide some overview in Section 7.1 and list four lemmas in Section 7.2, and then derive the asymptotic properties of the sample eigenvalues and the sample eigenvectors in Sections 7.3 and 7.4, respectively.

We study the consistency and strong inconsistency of PCA through the angle or the inner product between a sample eigenvector and the corresponding population eigenvector. We first note that this angle has a nice invariance property: it doesn't depend on the specific choice of the basis for the $d$-dimensional space, as discussed in details in the supplement (Shen et al., 2015). Given this invariance property, for the rest of the paper, we choose to use the population eigenvectors $u_j$, $j = 1, \ldots, d(n)$, as the basis of the $d$-dimensional space, which is equivalent to assuming that $X_i$, $i = 1, \ldots, n$, is a $d$-dimensional random vector with mean zero and a diagonal covariance matrix as $\Sigma = \Lambda = \mathrm{diag}\{\lambda_1^{(n)}, \ldots, \lambda_{d(n)}^{(n)}\}$. This will simplify our mathematical analysis, see for example (32) and (33).

Define $j_l$ to be the largest index in $H_l$ and then $j_l = \sum_{k=0}^{l} q_k$, $l = 1, \cdots, r$. Note that $j_r = \sum_{k=0}^{r} q_k = m$. Since the first $m$ eigenvalues are grouped into $r$ tiers in Assumption 2, then Assumption 2 can be rewritten as

$$\frac{\lambda_1^{(n)}}{\delta_1^{(n)}} \to \cdots \to \frac{\lambda_{j_1}^{(n)}}{\delta_1^{(n)}} \to 1, \quad \cdots, \quad \frac{\lambda_{j_{r-1}+1}^{(n)}}{\delta_r^{(n)}} \to \cdots \to \frac{\lambda_{j_r}^{(n)}}{\delta_r^{(n)}} \to 1. \tag{8}$$

### 7.1 Overview

Our proof makes use of the connection between the sample covariance matrix $\hat{\Sigma}$ and its dual matrix $\hat{\Sigma}_D$, which share the same nonzero eigenvalues. Since $\Sigma = \Lambda = \mathrm{diag}\{\lambda_1^{(n)}, \ldots, \lambda_{d(n)}^{(n)}\}$, then it follows from (1) and (5) that the dual matrix can be expressed as

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

where $\widetilde{Z}_j$ is the $n$-dimensional random vector and its elements are i.i.d random variables with zero mean, unit variance, and finite fourth moment. Furthermore, the dual matrix can be rewritten as the sum of two matrices as follows:

$$\hat{\Sigma}_D = A + B, \quad \text{with} \quad A = \frac{1}{n} \sum_{j=1}^{m} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T, \quad B = \frac{1}{n} \sum_{j=m+1}^{d(n)} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T. \tag{9}$$

First, we study the asymptotic properties of the eigenvalues of $A$ and $B$ in Lemmas 3 and 4, respectively. Then, the *Weyl Inequality and dual Weyl Inequality* (Tao, 2010), now restated as Lemma 2, enable us to establish the asymptotic properties of the eigenvalues of the dual matrix $\hat{\Sigma}_D$ in Section 7.3. Finally, we derive the asymptotic properties of the sample eigenvectors of $\hat{\Sigma}$ in Section 7.4. Some intuitive ideas are provided in the supplement (Shen et al., 2015) to help understanding the proof.

### 7.2 Lemmas

We list four lemmas that are used in our proof. Lemma 1 studies asymptotic properties of the largest and smallest non-zero eigenvalues of a random matrix.

**Lemma 1** *Suppose $B = \frac{1}{q}VV^T$ where $V$ is an $p \times q$ random matrix composed of i.i.d. random variables with zero mean, unit variance and finite fourth moment. As $q \to \infty$ and $\frac{p}{q} \to c \in [0, \infty)$, the largest and smallest non-zero eigenvalues of $B$ converge almost surely to $(1+\sqrt{c})^2$ and $(1-\sqrt{c})^2$, respectively.*

**Remark 1** *Lemma 1 is known as the Bai-Yin's law (Bai and Yin, 1993). As in Remark 1 of Bai and Yin (1993), the smallest non-zero eigenvalue is the $p-q+1$ smallest eigenvalue of $B$ for $c > 1$.*

Lemma 2 is about the Weyl Inequality and the dual Weyl Inequality (Tao, 2010), which appear below as the right-hand-side inequality and the left-hand-side inequality, respectively.

**Lemma 2** *If $A, B$ are $p \times p$ real symmetric matrices, then for all $j = 1, \dots, p$,*

$$\left\{ \begin{array}{lcl} \lambda_j(A) & + & \lambda_p(B) \\ \lambda_{j+1}(A) & + & \lambda_{p-1}(B) \\ & \vdots & \\ \lambda_p(A) & + & \lambda_j(B) \end{array} \right\} \leq \lambda_j(A+B) \leq \left\{ \begin{array}{lcl} \lambda_j(A) & + & \lambda_1(B) \\ \lambda_{j-1}(A) & + & \lambda_2(B) \\ & \vdots & \\ \lambda_1(A) & + & \lambda_j(B) \end{array} \right\},$$

*where $\lambda_j(\cdot)$ is the $j$-th largest eigenvalue of the matrix.*

**Lemma 3** *As $n \to \infty$, the eigenvalues of the matrix $A$ in (9) satisfy*

$$\frac{\lambda_j(A)}{\lambda_j^{(n)}} \xrightarrow{\text{a.s}} 1, \quad for \quad j = 1, \cdots, m.$$

**Proof** Define the $m$-dimensional random vectors $X_i^* = \left[I_m, 0_{m\times(d-m)}\right] X_i$, $i = 1, \cdots, n$. Then, $X_i^*$ has mean zero and the following covariance matrix $\Sigma^*$:

$$\Sigma^* = \begin{pmatrix} \lambda_1^{(n)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m^{(n)} \end{pmatrix}.$$

Let $A^*$ be the dual matrix of the matrix $A$. The sample covariance matrix of $X_i^*$ is

$$\begin{aligned} A^* &= \frac{1}{n}\sum_{i=1}^{n} X_i^* X_i^{*T} \\ &= \lambda_1^{(n)} \times \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n z_{i,1}^2 & \cdots & \left\{\frac{\lambda_m^{(n)}}{\lambda_1^{(n)}}\right\}^{\frac{1}{2}} \frac{1}{n}\sum_{i=1}^n z_{i,1}z_{i,m} \\ \vdots & \ddots & \vdots \\ \left\{\frac{\lambda_m^{(n)}}{\lambda_1^{(n)}}\right\}^{\frac{1}{2}} \frac{1}{n}\sum_{i=1}^n z_{i,1}z_{i,m} & \cdots & \frac{\lambda_m^{(n)}}{\lambda_1^{(n)}} \frac{1}{n}\sum_{i=1}^n z_{i,m}^2 \end{pmatrix}, \end{aligned} \tag{10}$$

where the $z_{i,j}$'s are defined in (1).

Since $A^*$ is the dual matrix of $A$, then $A$ and $A^*$ share the same non-zero eigenvalues. Below we study the eigenvalues of $A$ through the dual matrix $A^*$.

The i.i.d. and unit variance properties of the $z_{i,j}$'s yield that as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} z_{i,k}z_{i,l} \xrightarrow{\text{a.s}} \begin{cases} 1 & 1 \le k = l \le m \\ 0 & 1 \le k \ne l \le m \end{cases}. \tag{11}$$

Denote $b_k = \lim_{n\to\infty} \frac{\lambda_k^{(n)}}{\lambda_1^{(n)}} \le 1$, $k = 1, \cdots, m$. Then it follows from (10) and (11) that as $n \to \infty$,

$$\frac{1}{\lambda_1^{(n)}} A^* \xrightarrow{\text{a.s}} \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_m \end{pmatrix},$$

which further yields

$$\frac{\lambda_1(A)}{\lambda_1^{(n)}} = \frac{\lambda_1(A^*)}{\lambda_1^{(n)}} \xrightarrow{\text{a.s}} 1. \tag{12}$$

Similarly, for $k = 2, \cdots, m$, we have that as $n \to \infty$,

$$\frac{\lambda_1(\frac{1}{n}\sum_{j=k}^{m} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T)}{\lambda_k^{(n)}} \xrightarrow{\text{a.s}} 1. \tag{13}$$

Next we derive the upper and lower bounds for $\lambda_k(A)$, $k = 2, \cdots, m$. According to Lemma 2, we have the following inequality :

$$\lambda_k(A) = \lambda_k(\frac{1}{n}\sum_{j=1}^{m} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T) \le \lambda_1(\frac{1}{n}\sum_{j=k}^{m} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T) + \lambda_k(\frac{1}{n}\sum_{j=1}^{k-1} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T).$$

Since the rank of $\frac{1}{n}\sum_{j=1}^{k-1} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T$ is at most $k-1$, then $\lambda_k(\frac{1}{n}\sum_{j=1}^{k-1} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T) = 0$, which together with (13), yields that

$$\frac{\lambda_k(A)}{\lambda_k^{(n)}} \le \frac{1}{\lambda_k^{(n)}} \times \lambda_1(\frac{1}{n}\sum_{j=k}^{m} \lambda_j^{(n)} \widetilde{Z}_j \widetilde{Z}_j^T). \tag{14}$$

For the lower bound, it follows from Equation (5.9) in Jung and Marron (2009) that

$$\lambda_1(\frac{\lambda_k^{(n)}}{n}\widetilde{Z}_k\widetilde{Z}_k^T) + \lambda_n(\frac{1}{n}\sum_{j=k+1}^{m}\lambda_j^{(n)}\widetilde{Z}_j\widetilde{Z}_j^T) \le \lambda_k(A). \tag{15}$$

Given that the rank of $\frac{1}{n}\sum_{j=k+1}^{m}\lambda_j^{(n)}\widetilde{Z}_j\widetilde{Z}_j^T$ is at most $m$ with $m < n$, then $\lambda_n(\frac{1}{n}\sum_{j=k+1}^{m}\lambda_j^{(n)}\widetilde{Z}_j\widetilde{Z}_j^T) = 0$, which together with (15), yields that

$$\frac{\lambda_k(A)}{\lambda_k^{(n)}} \ge \frac{1}{\lambda_k^{(n)}} \times \lambda_1(\frac{\lambda_k^{(n)}}{n}\widetilde{Z}_k\widetilde{Z}_k^T). \tag{16}$$

Note that as $n \to \infty$,

$$\frac{1}{\lambda_k^{(n)}} \times \lambda_1(\frac{\lambda_k^{(n)}}{n}\widetilde{Z}_k\widetilde{Z}_k^T) = \frac{1}{n}\widetilde{Z}_k^T\widetilde{Z}_k \xrightarrow{\text{a.s}} 1. \tag{17}$$

It follows from (13), (14), (16) and (17) that, for $k = 2, \cdots, m$,

$$\frac{\lambda_k(A)}{\lambda_k^{(n)}} \xrightarrow{\text{a.s}} 1, \quad \text{as} \quad n \to \infty. \tag{18}$$

The combination of (12) and (18) proves Lemma 6.1.

■

**Lemma 4** *Assume that* $\lim_{n\to\infty}\frac{d(n)}{n} = c$, *where* $0 \le c \le \infty$, *and let* $\lambda_{\max}(\cdot)$ *and* $\lambda_{\min}(\cdot)$ *be the largest and smallest non-zero eigenvalues of the matrix, respectively. As* $n \to \infty$, $\lambda_{\max}(B)$ *and* $\lambda_{\min}(B)$, *where* $B$ *in* (9), *satisfy*

$$\lambda_{\max}(B) \quad \text{and} \quad \lambda_{\min}(B) \quad \stackrel{\text{a.s}}{\to} c_\lambda, \quad \text{for} \quad c = 0, \tag{19}$$

$$\frac{n}{d(n)}\lambda_{\max}(B) \quad \text{and} \quad \frac{n}{d(n)}\lambda_{\min}(B) \quad \stackrel{\text{a.s}}{\to} c_\lambda, \quad \text{for} \quad c = \infty, \tag{20}$$

*and*

$$\lambda_{\max}(B) \stackrel{\text{a.s}}{\to} c_\lambda(1+\sqrt{c})^2 \quad \text{and} \quad \lambda_{\min}(B) \stackrel{\text{a.s}}{\to} c_\lambda(1-\sqrt{c})^2, \quad \text{for} \quad 0 < c < \infty. \tag{21}$$

**Remark 2** *If* $\lambda_{m+1}^{(n)} \to \cdots \to \lambda_{d(n)}^{(n)}$ *is relaxed to* $\lambda_{m+1}^{(n)} \asymp \cdots \asymp \lambda_{d(n)}^{(n)}$, *then* "$\stackrel{\text{a.s}}{\to}$" *is replaced by almost surely* "$\asymp$".

**Proof** Define $B^* = \frac{1}{n}\sum_{j=m+1}^{d(n)}\widetilde{Z}_j\widetilde{Z}_j^T$. The proof uses the following inequalities for $k \ge 1$:

$$\lambda_{d(n)}^{(n)} \times \lambda_k(B^*) \le \lambda_k(B) \le \lambda_{m+1}^{(n)} \times \lambda_k(B^*). \tag{22}$$

We first prove the right inequality of (22). Note that $\lambda_{m+1}^{(n)} B^*$ can be rewritten as $\lambda_{m+1}^{(n)} B^* = B + B_R^*$, where $B_R^* = \frac{1}{n}\sum_{j=m+1}^{d(n)}(\lambda_m^{(n)} - \lambda_j^{(n)})\widetilde{Z}_j \widetilde{Z}_j^T$ and is a non-negative matrix. It then follows from Lemma 2 that for $k \geq 1$,

$$\lambda_{m+1}^{(n)} \times \lambda_k(B^*) = \lambda_k(\lambda_{m+1}^{(n)} B^*) \geq \lambda_k(B) + \lambda_n(B_R^*) \geq \lambda_k(B),$$

which yields the right inequality of (22).

For the left inequality in (22), note that $B = \lambda_{d(n)}^{(n)} B^* + B_L^*$, where $B_L^* = \frac{1}{n}\sum_{j=m+1}^{d(n)}(\lambda_j^{(n)} - \lambda_{d(n)}^{(n)})\widetilde{Z}_j \widetilde{Z}_j^T$ and is a non-negative matrix. Lemma 2 implies that for $k \geq 1$,

$$\lambda_k(B) \geq \lambda_k(\lambda_{d(n)}^{(n)} B^*) + \lambda_n(B_L^*) \geq \lambda_k(\lambda_{d(n)}^{(n)} B^*) = \lambda_{d(n)}^{(n)} \times \lambda_k(B^*),$$

which yields the left inequality of (22).

Note that $B^*$ can be rewritten as $B^* = \frac{1}{n} V V^T$, where $V = [\widetilde{Z}_{m+1}, \cdots, \widetilde{Z}_{d(n)}]$ is an $n \times (d(n) - m)$ matrix. If $\lim_{n\to\infty} \frac{d(n)}{n} = \lim_{n\to\infty} \frac{d(n)-m}{n} = \infty$, then according to Lemma 1, we have that

$$\frac{1}{d(n)-m}\lambda_{\max}(VV^T) \quad \text{and} \quad \frac{1}{d(n)-m}\lambda_{\min}(VV^T) \overset{\text{a.s}}{\to} 1.$$

It then follows that $\frac{n}{d(n)}\lambda_{\max}(B^*)$ and $\frac{n}{d(n)}\lambda_{\min}(B^*) \overset{\text{a.s}}{\to} 1$, which, together with (22) and $\lambda_{m+1}^{(n)} \to \lambda_{d(n)}^{(n)} \to c_r$, yields (20).

Now consider the case $\lim_{n\to\infty} \frac{d(n)}{n} = \lim_{n\to\infty} \frac{d(n)-m}{n} = c < \infty$. Since $B^* = \frac{1}{n}VV^T$ and $\frac{1}{n}V^T V$ share the non-zero eigenvalues, then we study the eigenvalues of $B^*$ through $\frac{1}{n}V^T V$. Applying Lemma 1 to $\frac{1}{n}V^T V$ yields that

$$\lambda_{\max}(\frac{1}{n}V^T V) \overset{\text{a.s}}{\to} (1+\sqrt{c})^2 \quad \text{and} \quad \lambda_{\min}(\frac{1}{n}V^T V) \overset{\text{a.s}}{\to} (1-\sqrt{c})^2.$$

It then follows that $\lambda_{\max}(B^*) \overset{\text{a.s}}{\to} (1+\sqrt{c})^2$ and $\lambda_{\min}(B^*) \overset{\text{a.s}}{\to} (1-\sqrt{c})^2$. In addition, given that $\lambda_{m+1}^{(n)} \to \lambda_{d(n)}^{(n)} \to c_r$ and (22), then we have $\lambda_{\max}(B) \overset{\text{a.s}}{\to} c_r(1+\sqrt{c})^2$ and $\lambda_{\min}(B) \overset{\text{a.s}}{\to} c_r(1-\sqrt{c})^2$ for $0 \leq c < \infty$, which yields (19) ($c = 0$) and (21) ($0 < c < \infty$) .

■

### 7.3 Asymptotic properties of the sample eigenvalues

We now study the asymptotic properties of the sample eigenvalues $\hat{\lambda}_j$ for $j = 1, \cdots, [n \wedge d(n)]$, which are the same as those of the dual matrix $\hat{\Sigma}_D$, denoted as $\lambda_j(\hat{\Sigma}_D) = \lambda_j(A + B)$.

#### 7.3.1 Scenario (a) in Theorem 1

Scenario (a) contains three different cases: $\lim_{n\to\infty} \frac{d(n)}{n} = 0$, $\infty$, or $c$ ($0 < c < \infty$). The proofs are different for each case and are provided separately below.

Consider the first case: $\lim_{n\to\infty} \frac{d(n)}{n} = 0$. According to Lemma 2, we have that

$$\frac{\lambda_j(A)}{\lambda_j^{(n)}} \leq \frac{\hat{\lambda}_j}{\lambda_j^{(n)}} \leq \frac{\lambda_j(A)}{\lambda_j^{(n)}} + \frac{\lambda_1(B)}{\lambda_j^{(n)}}. \tag{23}$$

If $\lambda_m^{(n)} \to \infty$, it follows from (19) that $\frac{\lambda_1(B)}{\lambda_j^{(n)}} \xrightarrow{\text{a.s}} 0$ for $j = 1, \cdots, m$. Then the combination of Lemma 3 and (23) proves that, as $n \to \infty$,

$$\frac{\hat{\lambda}_j}{\lambda_j^{(n)}} \xrightarrow{\text{a.s}} 1, \quad j = 1, \cdots, m. \tag{24}$$

If $\lambda_m^{(n)} < \infty$, according to Theorem 1 ($c = 0$) of Baik and Silverstein (2006), we still have (24). In addition, according to Lemma 2, we have

$$\lambda_j(B) \leq \hat{\lambda}_j \leq \lambda_j(A) + \lambda_1(B). \tag{25}$$

Since the rank of $A$ is at most $m$, then $\lambda_j(A) = 0$ for $j \geq m+1$, which, together with (25), yields that for $j = m+1, \cdots, [n \wedge (d(n) - m)]$,

$$\lambda_{\min}(B) \leq \hat{\lambda}_j \leq \lambda_{\max}(B). \tag{26}$$

Thus it follows from (19) and (26) that as $n \to \infty$,

$$\hat{\lambda}_j \xrightarrow{\text{a.s}} c_r, \quad j = m+1, \cdots, [n \wedge (d(n) - m)].$$

Now consider the second case: $\lim_{n\to\infty} \frac{d(n)}{n} = \infty$. Since $\frac{d(n)}{n\lambda_m^{(n)}} \to 0$, then $\lambda_m^{(n)} \to \infty$, which, together with (20), (23) and Lemma 3, yields (24). Since $\lim_{n\to\infty} \frac{d(n)}{n} = \infty$, then $[n \wedge (d(n) - m)] = [n \wedge d(n)] = n$ as $n \to \infty$. It follows from (20) and (26) that

$$\frac{n}{d(n)} \hat{\lambda}_j \xrightarrow{\text{a.s}} c_r, \quad j = m+1, \cdots, [n \wedge d(n)].$$

Finally, consider the third case: $\lim_{n\to\infty} \frac{d(n)}{n} = c$ ($0 < c < \infty$). Similarly, it follows from $\frac{d(n)}{n\lambda_m^{(n)}} \to 0$ that $\lambda_m^{(n)} \to \infty$, which, jointly with (21), (23) and Lemma 3, yields (24). In addition, note that (21) and (26), and then almost surely we have

$$c_r(1 - \sqrt{c})^2 \leq \underline{\lim}_{n\to\infty} \hat{\lambda}_j \leq \overline{\lim}_{n\to\infty} \hat{\lambda}_j \leq c_r(1 + \sqrt{c})^2, \quad j = m+1, \cdots, [n \wedge (d(n) - m)].$$

All together, we have proven the consistency of the first $m$ sample eigenvalues under Scenario (a), as stated in (24).

#### 7.3.2 Scenario (b) in Theorem 1

Given $\frac{d(n)}{n\delta_{h+1}^{(n)}} \to \infty$ and (8), then $\frac{d(n)}{n} \to \infty$ and $\frac{d(n)}{n\lambda_j^{(n)}} \to 0$ for $j \in H_l, l = 1, \cdots, h$. Thus, according to (20), we have that $\frac{\lambda_1(B)}{\lambda_j^{(n)}} = \left[\frac{n}{d(n)}\lambda_1(B)\right]\left[\frac{d(n)}{n\lambda_j^{(n)}}\right] \xrightarrow{\text{a.s}} 0$ for $j \in H_l, l = 1, \cdots, h$. Furthermore, it follows from Lemma 3 and (23) that as $n \to \infty$,

$$\frac{\hat{\lambda}_j}{\lambda_j^{(n)}} \xrightarrow{\text{a.s}} 1, \quad j \in H_l, l = 1, \cdots, h. \tag{27}$$

Note that (25) can be rewritten as

$$\frac{n}{d(n)}\lambda_j(B) \leq \frac{n}{d(n)}\hat{\lambda}_j \leq \frac{n}{d(n)}\lambda_j(A) + \frac{n}{d(n)}\lambda_1(B), \tag{28}$$

which yields that for $j = j_h + 1, \cdots, [n \wedge (d(n) - m)]$,

$$\frac{n}{d(n)}\lambda_{\min}(B) \leq \frac{n}{d(n)}\hat{\lambda}_j \leq \frac{n}{d(n)}\lambda_j(A) + \frac{n}{d(n)}\lambda_{\max}(B). \tag{29}$$

Note that for $j = j_h + 1, \cdots, [n \wedge (d(n) - m)]$, we have

$$\frac{n}{d(n)}\lambda_j(A) \leq \frac{n}{d(n)}\lambda^{(n)}_{j_{h+1}}(A) = \left\{\frac{n\delta^{(n)}_{h+1}}{d(n)}\right\}\left\{\frac{\lambda^{(n)}_{j_{h+1}}(A)}{\delta^{(n)}_{h+1}}\right\}.$$

It then follows from $\frac{d(n)}{n\delta^{(n)}_{h+1}} \to \infty$ and Lemma 3 that $\frac{n}{d(n)}\lambda_j(A) \overset{\text{a.s}}{\to} 0$. Since $\frac{d(n)}{n} \to \infty$, then $[n \wedge (d(n) - m)] = [n \wedge d(n)] = n$, as $n \to \infty$. Then it follows from (20) and (29) that as $n \to \infty$

$$\frac{n}{d(n)}\hat{\lambda}_j \overset{\text{a.s}}{\to} c_\lambda, \quad j = j_h + 1, \cdots, [n \wedge d(n)]. \tag{30}$$

The combination of (27) and (30) yields the asymptotic properties of the non-zero sample eigenvalues in Scenario (b).

#### 7.3.3 Scenario (c) in Theorem 1

Since $\frac{d(n)}{n\delta^{(n)}_1} \to \infty$, then $\frac{d(n)}{n} \to \infty$. According to (28), we have that for $j = 1, \cdots, [n \wedge (d(n) - m)]$,

$$\frac{n}{d(n)}\lambda_{\min}(B) \leq \frac{n}{d(n)}\hat{\lambda}_j \leq \frac{n}{d(n)}\lambda_1(A) + \frac{n}{d(n)}\lambda_{\max}(B). \tag{31}$$

Since $\frac{d(n)}{n\delta^{(n)}_1} \to \infty$, it follows from (8) and Lemma 3 that

$$\frac{n}{d(n)}\lambda_1(A) = \left[\frac{n\delta^{(n)}_1}{d(n)}\right] \times \left[\frac{\lambda^{(n)}_1}{\delta^{(n)}_1}\right] \times \left[\frac{\lambda_1(A)}{\lambda^{(n)}_1}\right] \overset{\text{a.s}}{\to} 0.$$

Again note that $[n \wedge (d(n) - m)] = [n \wedge d(n)] = n$, as $n \to \infty$. Then it follows from (20) and (31) that

$$\frac{n}{d(n)}\hat{\lambda}_j \overset{\text{a.s}}{\to} c_\lambda, \quad j = 1, \cdots, [n \wedge d(n)].$$

### 7.4 Asymptotic properties of the sample eigenvectors

We first state two results that simplify the proof. As aforementioned, in light of the invariance property of the angle, we choose the population eigenvectors $u_j$, $j = 1, \ldots, d(n)$, as the basis of the $d$-dimensional space. It then follows that $u_j = e_j$ where the $j$th component of $e_j$ equals to 1 and all the other components equal to zero. This suggests that

$$|< \hat{u}_j, u_j >|^2 = |< \hat{u}_j, e_j >|^2 = \hat{u}^2_{j,j}, \tag{32}$$

and for any index set $H$,

$$\cos\left[\text{angle}\left(\hat{u}_j, \text{span}\{u_k, k \in H\}\right)\right] = \sum_{k \in H} \hat{u}_{k,j}^2. \tag{33}$$

As a reminder, the population eigenvalues are grouped into $r+1$ tiers and the index set of the eigenvalues in the $l$th tier $H_l$ is defined in (4). Define

$$\hat{U}_{k,l} = (\hat{u}_{i,j})_{i \in H_k, j \in H_l}, \quad 1 \le k, l \le r+1.$$

Then, the sample eigenvector matrix $\hat{U}$ can be rewritten as the following:

$$\hat{U} = [\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_{d(n)}] = \begin{pmatrix} \hat{U}_{1,1} & \hat{U}_{1,2} & \cdots & \hat{U}_{1,r+1} \\ \hat{U}_{2,1} & \hat{U}_{2,2} & \cdots & \hat{U}_{2,r+1} \\ \vdots & \vdots & & \vdots \\ \hat{U}_{r+1,1} & \hat{U}_{r+1,2} & \cdots & \hat{U}_{r+1,r+1} \end{pmatrix}.$$

To derive the asymptotic properties of the sample eigenvectors $\hat{u}_j$, we consider the three scenarios of Theorem 1 separately.

### 7.4.1 Scenario (b) in Theorem 1

Under Scenario (b), there exists a constant $h \in [1, r]$, such that $\frac{d(n)}{n\delta_h^{(n)}} \to 0$ and $\frac{d(n)}{n\delta_{h+1}^{(n)}} \to \infty$. In order to obtain the the subspace consistency properties in Scenario (b), according to (33), we only need to show that as $n \to \infty$,

$$\sum_{k \in H_l} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{\delta_{l+1}^{(n)}}{\delta_l^{(n)}}\right\}, \quad j \in H_l, \quad l = 1, \cdots, h-1, \tag{34}$$

$$\sum_{k \in H_h} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\} \vee \mathrm{O}_{\mathrm{a.s}}\left\{\frac{d(n)}{n\delta_h^{(n)}}\right\}, \quad j \in H_h, \tag{35}$$

which are respectively equivalent to

$$\sum_{j \in H_l} \sum_{k \in H_l} \hat{u}_{k,j}^2 = |H_l| + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{\delta_{l+1}^{(n)}}{\delta_l^{(n)}}\right\}, \quad l = 1, \cdots, h-1, \tag{36}$$

$$\sum_{j \in H_h} \sum_{k \in H_h} \hat{u}_{k,j}^2 = |H_h| + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\} \vee \mathrm{O}_{\mathrm{a.s}}\left\{\frac{d(n)}{n\delta_h^{(n)}}\right\}, \tag{37}$$

where $|H_l|$ is the number of elements in $H_l$ and less than $m$. Since $\sum_{j \in H_l} \sum_{k \in H_l} = \sum_{k \in H_l} \sum_{j \in H_l}$, then in order to obtain (36) and (37), we just need to prove that as $n \to \infty$,

$$\sum_{j \in H_l} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{\delta_{l+1}^{(n)}}{\delta_l^{(n)}}\right\}, \quad k \in H_l, \quad l = 1, \cdots, h-1, \tag{38}$$

$$\sum_{j \in H_h} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\} \vee \mathrm{O}_{\mathrm{a.s}}\left\{\frac{d(n)}{n\delta_h^{(n)}}\right\}, \quad k \in H_h. \tag{39}$$

Therefore the proof of the subspace consistency contains two steps (38) and (39). Here we first prove (39) and then (38).

The third step is to show the strong inconsistency in Scenario (b). Since $\hat{\lambda}_j = 0$ for $j > [n \wedge d(n)]$, then we only need to show the strong inconsistency of $\hat{u}_j$, $j < [n \wedge d(n)]$. Here we will prove that as $n \to \infty$,

$$\max_{j_h+1 \leq j \leq [n \wedge d(n)]} \left\{ \frac{d(n)}{n\lambda_j^{(n)}} \hat{u}_{j,j}^2 \right\} = \mathrm{O}_{\mathrm{a.s}}(1). \tag{40}$$

**The First Step: Proof of** (39)**.** Since

$$\sum_{j \in H_h} \hat{u}_{k,j}^2 = 1 - \sum_{l=1}^{h-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 - \sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2,$$

then in order to obtain (39), we just need to show that as $n \to \infty$,

$$\sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}} \left\{ \frac{d(n)}{n\delta_h^{(n)}} \right\}, \quad k \in H_h, \tag{41}$$

$$\sum_{l=1}^{h-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 = \mathrm{o}_{\mathrm{a.s}} \left\{ \frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}} \right\}, \quad k \in H_h. \tag{42}$$

We first prove (41). Since $\sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2 \leq \sum_{k=1}^{j_h} \sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2$ for $k \in H_h$, then in order to generate (41), we need to show that as $n \to \infty$,

$$\sum_{k=1}^{j_h} \sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}} \left\{ \frac{d(n)}{n\delta_h^{(n)}} \right\}. \tag{43}$$

Since $\sum_{k=1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{j=1}^{d(n)} \hat{u}_{k,j}^2 = 1$, then we have

$$d(n) - j_l = \sum_{j=j_l+1}^{d(n)} \sum_{k=1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{k=1}^{j_l} \sum_{j=j_l+1}^{d(n)} \hat{u}_{k,j}^2 + \sum_{k=j_l+1}^{d(n)} \sum_{j=j_l+1}^{d(n)} \hat{u}_{k,j}^2,$$

$$d(n) - j_l = \sum_{k=j_l+1}^{d(n)} \sum_{j=1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{k=j_l+1}^{d(n)} \sum_{j=1}^{j_l} \hat{u}_{k,j}^2 + \sum_{k=j_l+1}^{d(n)} \sum_{j=j_l+1}^{d(n)} \hat{u}_{k,j}^2,$$

which yields

$$\sum_{k=1}^{j_l} \sum_{j=j_l+1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{k=j_l+1}^{d(n)} \sum_{j=1}^{j_l} \hat{u}_{k,j}^2. \tag{44}$$

Let $l = h$ in (44) and then (43) can be obtained through showing

$$\sum_{k=j_h+1}^{d(n)} \sum_{j=1}^{j_h} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}} \left\{ \frac{d(n)}{n\delta_h^{(n)}} \right\}. \tag{45}$$

Therefore, in order to show (41), we need to prove (45).

Before proving (45), we need some preparation. Denote $S = \Lambda^{-\frac{1}{2}}\hat{U}\hat{\Lambda}^{\frac{1}{2}}$ where $\hat{U}$ is the sample eigenvector matrix and $\hat{\Lambda}$ is the sample eigenvalue matrix defined in (3). Define

$$Z = (Z_1, \cdots, Z_n), \tag{46}$$

where $Z_i$ is in (2). It follows from (1), (2) and (3) that $SS^T = \frac{1}{n}ZZ^T$. Since $s_{k,j} = \lambda_k^{(n)^{-\frac{1}{2}}}\hat{\lambda}_j^{\frac{1}{2}}\hat{u}_{k,j}$, then considering the $k$-th diagonal entry of the matrices $SS^T = \frac{1}{n}ZZ^T$ on the two sides leads to

$$\frac{1}{\lambda_k^{(n)}}\sum_{j=1}^{d}\hat{\lambda}_j\hat{u}_{k,j}^2 = \sum_{j=1}^{d} s_{k,j}^2 = \frac{1}{n}\sum_{i=1}^{n} z_{i,k}^2, \quad k = 1, \cdots, d(n). \tag{47}$$

In addition, the $j$-th diagonal entry of $S^TS$ is less than or equal to its largest eigenvalue, i.e. $\lambda_{\max}(SS^T) = \lambda_{\max}(\frac{1}{n}ZZ^T) = \lambda_{\max}(\frac{1}{n}Z^TZ)$, which yields

$$\hat{\lambda}_j\sum_{k=1}^{d(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2 = \sum_{k=1}^{d(n)} s_{k,j}^2 \le \lambda_{\max}(\frac{1}{n}Z^TZ), \quad j = 1, \cdots, d(n). \tag{48}$$

According to (48), we have that for $l = 1, \cdots, h$,

$$\begin{aligned}
\hat{\lambda}_{j_l} \times \frac{1}{\lambda_{m+1}^{(n)}} \times \sum_{j=1}^{j_l}\sum_{k=m+1}^{d(n)}\hat{u}_{k,j}^2 &\le \sum_{j=1}^{j_l}\hat{\lambda}_j\sum_{k=m+1}^{d(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2 \\
&\le \sum_{j=1}^{j_l}\hat{\lambda}_j\sum_{k=1}^{d(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2 \le j_l \times \lambda_{\max}(\frac{1}{n}Z^TZ),
\end{aligned}$$

which yields

$$\sum_{k=m+1}^{d(n)}\sum_{j=1}^{j_l}\hat{u}_{k,j}^2 = \sum_{j=1}^{j_l}\sum_{k=m+1}^{d(n)}\hat{u}_{k,j}^2 \le j_l\lambda_{m+1}^{(n)} \times \frac{\delta_l^{(n)}}{\hat{\lambda}_{j_l}} \times \lambda_{\max}(\frac{1}{d(n)}Z^TZ) \times \frac{d(n)}{n\delta_l^{(n)}}. \tag{49}$$

Since $\frac{d(n)}{n} = \delta_{h+1}^{(n)} \times \frac{d(n)}{n\delta_{h+1}^{(n)}} \to \infty$, it follows from Lemma 3 that $\lambda_{\max}(\frac{1}{d(n)}Z^TZ) \xrightarrow{\text{a.s}} 1$. According to (8) and (27), $\frac{\delta_l^{(n)}}{\hat{\lambda}_{j_l}} = \frac{\delta_l^{(n)}}{\lambda_{j_l}^{(n)}} \times \frac{\lambda_{j_l}^{(n)}}{\hat{\lambda}_{j_l}} \xrightarrow{\text{a.s}} 1$, $l = 1, \cdots, h$. In addition, note that $j_l(< m)$ is finite and $\lambda_{m+1}^{(n)} \to c_\lambda$. Thus it follows from (49) that as $n \to \infty$,

$$\sum_{k=m+1}^{d(n)}\sum_{j=1}^{j_l}\hat{u}_{k,j}^2 = \mathrm{O}_{\text{a.s}}\left\{\frac{d(n)}{n\delta_l^{(n)}}\right\}. \tag{50}$$

From (47), we have that for $l = 1, \cdots, h$,

$$\begin{aligned}
\frac{1}{\lambda_{j_h+1}^{(n)}} \times \hat{\lambda}_{j_l}\sum_{k=j_h+1}^{m}\sum_{j=1}^{j_l}\hat{u}_{k,j}^2 &\le \sum_{k=j_h+1}^{m}\frac{1}{\lambda_k^{(n)}}\sum_{j=1}^{j_l}\hat{\lambda}_j\hat{u}_{k,j}^2 \\
&\le \sum_{k=j_h+1}^{m}\frac{1}{\lambda_k^{(n)}}\sum_{j=1}^{d(n)}\hat{\lambda}_j\hat{u}_{k,j}^2 = \sum_{k=j_h+1}^{m}\frac{1}{n}\sum_{i=1}^{n} z_{i,k}^2,
\end{aligned}$$

which yields

$$\sum_{k=j_h+1}^{m} \sum_{j=1}^{j_l} \hat{u}_{k,j}^2 \le \frac{\lambda_{j_h+1}^{(n)}}{\delta_l^{(n)}} \times \frac{\delta_l^{(n)}}{\hat{\lambda}_{j_l}} \times \sum_{k=j_h+1}^{m} \frac{1}{n} \sum_{i=1}^{n} z_{i,k}^2. \tag{51}$$

Since $\frac{\lambda_{j_h+1}^{(n)}}{\delta_{h+1}^{(n)}} \to 1$, $\frac{\delta_l^{(n)}}{\hat{\lambda}_{j_l}} \xrightarrow{\text{a.s}} 1$ and $\sum_{k=j_h+1}^{m} \frac{1}{n} \sum_{i=1}^{n} z_{i,k}^2 \xrightarrow{\text{a.s}} m - j_h$, it follows from (51) that as $n \to \infty$,

$$\sum_{k=j_h+1}^{m} \sum_{j=1}^{j_l} \hat{u}_{k,j}^2 = \mathrm{O}_{\text{a.s}} \left\{ \frac{\delta_{h+1}^{(n)}}{\delta_l^{(n)}} \right\}. \tag{52}$$

Since $\delta_{h+1}^{(n)} << \frac{d(n)}{n}$, it follows from (50) and (52) that as $n \to \infty$,

$$\sum_{k=j_h+1}^{d(n)} \sum_{j=1}^{j_l} \hat{u}_{k,j}^2 = \mathrm{O}_{\text{a.s}} \left\{ \frac{d(n)}{n\delta_l^{(n)}} \right\}, \quad l = 1, \cdots, h. \tag{53}$$

Letting $l = h$ in (53) results in (45).

Until now we have proven (41). In order to finish the first step proof, we need to show (42). Since $\frac{1}{n} \sum_{i=1}^{n} z_{i,k}^2 \xrightarrow{\text{a.s}} 1$, it follows from (47) that for $k \in H_h$,

$$\frac{1}{\lambda_k^{(n)}} \sum_{l=1}^{h-1} \sum_{j \in H_l} \hat{\lambda}_j \hat{u}_{k,j}^2 + \frac{1}{\lambda_k^{(n)}} \sum_{j \in H_h} \hat{\lambda}_j \hat{u}_{k,j}^2 + \frac{1}{\lambda_k^{(n)}} \sum_{j=j_h+1}^{d(n)} \hat{\lambda}_j \hat{u}_{k,j}^2 = \frac{1}{\lambda_k^{(n)}} \sum_{j=1}^{d(n)} \hat{\lambda}_j \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 1. \tag{54}$$

Since

$$\frac{\hat{\lambda}_j}{\lambda_k^{(n)}} \xrightarrow{\text{a.s}} \frac{\delta_l^{(n)}}{\delta_h^{(n)}}, \quad k \in H_h, j \in H_l, \tag{55}$$

and

$$\frac{1}{\lambda_k^{(n)}} \sum_{j=j_h+1}^{d(n)} \hat{\lambda}_j \hat{u}_{k,j}^2 \le \frac{\hat{\lambda}_{j_h+1}}{\lambda_k^{(n)}} \xrightarrow{\text{a.s}} \frac{\delta_{h+1}^{(n)}}{\delta_h^{(n)}} \to 0,$$

it follows from (54) that for $k \in H_h$,

$$\sum_{l=1}^{h-1} \frac{\delta_l^{(n)}}{\delta_h^{(n)}} \sum_{j \in H_l} \hat{u}_{k,j}^2 + \sum_{j \in H_h} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 1. \tag{56}$$

According to (43), we have $\sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2 \le \sum_{k=1}^{j_h} \sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 0$, which together with

$$\sum_{l=1}^{h-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 + \sum_{j \in H_h} \hat{u}_{k,j}^2 + \sum_{j=j_h+1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{j=1}^{d} \hat{u}_{k,j}^2 = 1,$$

yields that for $k \in H_h$,

$$\sum_{l=1}^{h-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 + \sum_{j \in H_h} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 1. \tag{57}$$

Since $\lim_{n\to\infty}\frac{\delta_l^{(n)}}{\delta_h^{(n)}} > 1$ for $l < h$, it follows from (56) and (57) that $\sum_{j\in H_h}\hat{u}_{k,j}^2 \overset{\text{a.s}}{\to} 1$ for $k \in H_h$, which together with (56), yields that for $k \in H_h$,

$$\sum_{l=1}^{h-1}\frac{\delta_l^{(n)}}{\delta_h^{(n)}}\sum_{j\in H_l}\hat{u}_{k,j}^2 \overset{\text{a.s}}{\to} 0. \tag{58}$$

Since $\lim_{n\to\infty}\frac{\delta_l^{(n)}}{\delta_h^{(n)}} \geq \lim_{n\to\infty}\frac{\delta_{h-1}^{(n)}}{\delta_h^{(n)}}$ for $l \leq h-1$, it follows from (58) that as $n\to\infty$,

$$\sum_{l=1}^{h-1}\sum_{j\in H_l}\hat{u}_{k,j}^2 = \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\}, \quad k\in H_h,$$

which is (42).

**The Second Step: Proof of** (38)**.** Below we illustrate how one can use (39) to prove (38) for $l = h-1$. Then through a similar procedure, the result for $l = h-1$ in (38) can be used to prove that (38) holds for $l = h-2$, which is then iterated until finishing the proof of (38).

Since

$$\sum_{j\in H_{h-1}}\hat{u}_{k,j}^2 = 1 - \sum_{l=1}^{h-2}\sum_{j\in H_l}\hat{u}_{k,j}^2 - \sum_{j=j_{h-1}+1}^{d(n)}\hat{u}_{k,j}^2,$$

then in order to obtain (38) for $l = h-1$, we need to prove that as $n\to\infty$,

$$\sum_{j=j_{h-1}+1}^{d(n)}\hat{u}_{k,j}^2 = \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\}, \quad k\in H_{h-1}, \tag{59}$$

$$\sum_{l=1}^{h-2}\sum_{j\in H_l}\hat{u}_{k,j}^2 = \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_{h-1}^{(n)}}{\delta_{h-2}^{(n)}}\right\}, \quad k\in H_{h-1}. \tag{60}$$

Now we show the proof of (59). Since $j_h < m$ is finite and $\sum_{j=1}^{j_{h-1}} = \sum_{l=1}^{h-1}\sum_{j\in H_l}$, it follows from (42) that as $n\to\infty$,

$$\sum_{k=j_{h-1}+1}^{j_h}\sum_{j=1}^{j_{h-1}}\hat{u}_{k,j}^2 = \sum_{k=j_{h-1}+1}^{j_h}\left(\sum_{l=1}^{h-1}\sum_{j\in H_l}\hat{u}_{k,j}^2\right) = \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\}. \tag{61}$$

Let $l = h-1$ in (53) to obtain that $\sum_{k=j_h+1}^{d(n)}\sum_{j=1}^{j_{h-1}}\hat{u}_{k,j}^2 = \mathrm{O}_{\text{a.s}}\left\{\frac{d(n)}{n\delta_{h-1}^{(n)}}\right\}$. Since $\delta_h^{(n)} >> \frac{d(n)}{n}$, it follows from (61) that as $n\to\infty$,

$$\begin{aligned}\sum_{k=j_{h-1}+1}^{d(n)}\sum_{j=1}^{j_{h-1}}\hat{u}_{k,j}^2 &= \sum_{k=j_{h-1}+1}^{j_h}\sum_{j=1}^{j_{h-1}}\hat{u}_{k,j}^2 + \sum_{k=j_h+1}^{d(n)}\sum_{j=1}^{j_{h-1}}\hat{u}_{k,j}^2\\ &= \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\} + \mathrm{O}_{\text{a.s}}\left\{\frac{d(n)}{n\delta_{h-1}^{(n)}}\right\} = \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}}\right\}.\end{aligned} \tag{62}$$

Let $l = h-1$ in (44), which together with (62), proves that as $n \to \infty$,

$$\sum_{k=1}^{j_{h-1}} \sum_{j=j_{h-1}+1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{k=j_{h-1}+1}^{d(n)} \sum_{j=1}^{j_{h-1}} \hat{u}_{k,j}^2 = \mathrm{o}_{\mathrm{a.s}} \left\{ \frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}} \right\}. \tag{63}$$

Since $\sum_{j=j_{h-1}+1}^{d(n)} \hat{u}_{k,j}^2 \le \sum_{k=1}^{j_{h-1}} \sum_{j=j_{h-1}+1}^{d(n)} \hat{u}_{k,j}^2$ for $k \in H_{h-1}$, then (59) follows from (63).

Now we show the proof of (60) to finish the second step. Since $\frac{1}{n}\sum_{i=1}^n z_{i,k}^2 \overset{\mathrm{a.s}}{\to} 1$, it follows from (47) that for $k \in H_{h-1}$,

$$\frac{1}{\lambda_k^{(n)}} \sum_{l=1}^{h-2} \sum_{j\in H_l} \hat{\lambda}_j \hat{u}_{k,j}^2 + \frac{1}{\lambda_k^{(n)}} \sum_{j\in H_{h-1}} \hat{\lambda}_j \hat{u}_{k,j}^2 + \frac{1}{\lambda_k^{(n)}} \sum_{j=j_{h-1}+1}^{d(n)} \hat{\lambda}_j \hat{u}_{k,j}^2 = \frac{1}{\lambda_k^{(n)}} \sum_{j=1}^{d(n)} \hat{\lambda}_j \hat{u}_{k,j}^2 \overset{\mathrm{a.s}}{\to} 1. \tag{64}$$

Since $\frac{1}{\lambda_k^{(n)}} \sum_{j=j_{h-1}+1}^{d(n)} \hat{\lambda}_j \hat{u}_{k,j}^2 \le \frac{1}{\lambda_k^{(n)}} \hat{\lambda}_{j_{h-1}+1} \sum_{j=j_{h-1}+1}^{d(n)} \hat{u}_{k,j}^2$ and $\frac{\hat{\lambda}_{j_{h-1}+1}}{\lambda_k^{(n)}} \overset{\mathrm{a.s}}{\to} \lim_{n\to\infty} \frac{\delta_h^{(n)}}{\delta_{h-1}^{(n)}} < 1$ for $k \in H_{h-1}$, it follows from (59) that

$$\frac{1}{\lambda_k^{(n)}} \sum_{j=j_{h-1}+1}^{d(n)} \hat{\lambda}_j \hat{u}_{k,j}^2 \overset{\mathrm{a.s}}{\to} 0,$$

which together with (55) and (64), yields

$$\sum_{l=1}^{h-2} \frac{\delta_l^{(n)}}{\delta_{h-1}^{(n)}} \sum_{j\in H_l} \hat{u}_{k,j}^2 + \sum_{j\in H_{h-1}} \hat{u}_{k,j}^2 \overset{\mathrm{a.s}}{\to} 1, \quad k \in H_{h-1}. \tag{65}$$

In addition, since

$$\sum_{l=1}^{h-2} \sum_{j\in H_l} \hat{u}_{k,j}^2 + \sum_{j\in H_{h-1}} \hat{u}_{k,j}^2 + \sum_{j=j_{h-1}+1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{j=1}^{d(n)} \hat{u}_{k,j}^2 = 1,$$

it follows from (59) that

$$\sum_{l=1}^{h-2} \sum_{j\in H_l} \hat{u}_{k,j}^2 + \sum_{j\in H_{h-1}} \hat{u}_{k,j}^2 \overset{\mathrm{a.s}}{\to} 1, \quad k \in H_{h-1}. \tag{66}$$

Note that $\lim_{n\to\infty} \frac{\delta_l^{(n)}}{\delta_{h-1}^{(n)}} > 1$ for $l < h-1$. Then the combination of (65) and (66) gives $\sum_{j\in H_{h-1}} \hat{u}_{k,j}^2 \overset{\mathrm{a.s}}{\to} 1$ for $k \in H_{h-1}$, which together with (65), yields

$$\sum_{l=1}^{h-2} \frac{\delta_l^{(n)}}{\delta_{h-1}^{(n)}} \sum_{j\in H_l} \hat{u}_{k,j}^2 \overset{\mathrm{a.s}}{\to} 0, \quad k \in H_{h-1}. \tag{67}$$

Since $\lim_{n\to\infty} \frac{\delta_l^{(n)}}{\delta_{h-1}^{(n)}} \ge \lim_{n\to\infty} \frac{\delta_{h-2}^{(n)}}{\delta_{h-1}^{(n)}}$ for $l \le h-2$, it follows from (67) that as $n \to \infty$,

$$\sum_{l=1}^{h-2} \sum_{j\in H_l} \hat{u}_{k,j}^2 = \mathrm{o}_{\mathrm{a.s}} \left\{ \frac{\delta_{h-1}^{(n)}}{\delta_{h-2}^{(n)}} \right\}, \quad k \in H_h,$$

which is (60).

**The Third Step: Proof of** (40)**.** According to (47), we have $\frac{1}{\lambda_j^{(n)}}\hat{\lambda}_j\hat{u}_{j,j}^2 \le \frac{1}{n}\sum_{i=1}^n z_{i,j}^2$ for $j = 1, \cdots, d(n)$, which yields

$$\hat{\lambda}_{j_{[n\wedge d(n)]}} \times \max_{j_h+1\le j\le [n\wedge d(n)]}\left\{\frac{1}{\lambda_j^{(n)}}\hat{u}_{j,j}^2\right\} \le \max_{j_h+1\le j\le [n\wedge d(n)]}\left\{\frac{\hat{\lambda}_j}{\lambda_j^{(n)}}\hat{u}_{j,j}^2\right\}$$
$$\le \max_{1\le j\le [n\wedge d(n)]}\left\{\frac{\hat{\lambda}_j}{\lambda_j^{(n)}}\hat{u}_{j,j}^2\right\} \le \max_{1\le j\le [n\wedge d(n)]}\left\{\frac{1}{n}\sum_{i=1}^n z_{i,j}^2\right\}. \tag{68}$$

Select the first $[n \wedge d(n)]$ rows of $Z$ in (46) and denote the resulting random matrix as $Z^*$. Since $\frac{1}{n}\sum_{i=1}^n z_{i,j}^2$ is the $j$-th diagonal entry of $\frac{1}{n}Z^*Z^{*T}$ for $1 \le j \le [n \wedge d(n)]$, it follows that

$$\frac{1}{n}\sum_{i=1}^n z_{i,j}^2 \le \lambda_{\max}(\frac{1}{n}Z^*Z^{*T}), \quad 1 \le j \le [n\wedge d(n)],$$

which yields

$$\max_{1\le j\le [n\wedge d(n)]}\left\{\frac{1}{n}\sum_{i=1}^n z_{i,j}^2\right\} \le \lambda_{\max}(\frac{1}{n}Z^*Z^{*T}).$$

Then from (68),

$$\left\{\frac{n}{d(n)}\hat{\lambda}_{j_{[n\wedge d(n)]}}\right\} \times \max_{j_h+1\le j\le [n\wedge d(n)]}\left\{\frac{d(n)}{n\lambda_j^{(n)}}\hat{u}_{j,j}^2\right\} \le \lambda_{\max}(\frac{1}{n}Z^*Z^{*T}). \tag{69}$$

Since $\frac{d(n)}{n} \to \infty$ here, $[n \wedge d(n)] = n$. According to Lemma 1, we have $\lambda_{\max}(\frac{1}{n}Z^*Z^{*T}) \overset{\text{a.s}}{\to} 4$, which together with (30) and (69), yields (40).

### 7.4.2 Scenario (a) in Theorem 1

Scenario (a) contains three different cases: $\lim_{n\to\infty}\frac{d(n)}{n} = 0$, $\infty$, or $c$ $(0 < c < \infty)$. The proofs are slightly different for each case and are provided separately below.

Consider the case $\lim_{n\to\infty}\frac{d(n)}{n} = \infty$. Since $\lambda_j^{(n)} \to c_\lambda$ for $j \in H_{r+1}$, then $\frac{d(n)}{n\delta_r^{(n)}} \to 0$ and $\frac{d(n)}{n\lambda_j^{(n)}} \to \infty$ for $j \in H_{r+1}$. Thus $h$ in (34) and (35) becomes $r$ such that as $n \to \infty$,

$$\sum_{k\in H_l}\hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_l^{(n)}}{\delta_{l-1}^{(n)}} \vee \frac{\delta_{l+1}^{(n)}}{\delta_l^{(n)}}\right\}, \quad j \in H_l, \quad l = 1, \cdots, r-1, \tag{70}$$

$$\sum_{k\in H_r}\hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\text{a.s}}\left\{\frac{\delta_r^{(n)}}{\delta_{r-1}^{(n)}}\right\} \vee \mathrm{O}_{\text{a.s}}\left\{\frac{d(n)}{n\delta_r^{(n)}}\right\}, \quad j \in H_r. \tag{71}$$

Since $j_r = m$, then (50) becomes that as $n \to \infty$,

$$\sum_{k=m+1}^{d(n)}\sum_{j=1}^m \hat{u}_{k,j}^2 = \mathrm{O}_{\text{a.s}}\left\{\frac{d(n)}{n\delta_r^{(n)}}\right\},$$

which together with (44), yields that

$$\sum_{k=1}^{m}\sum_{j=m+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}}\left\{\frac{d(n)}{n\delta_r^{(n)}}\right\}. \tag{72}$$

Since

$$1 \geq \sum_{k\in H_{r+1}} \hat{u}_{k,j}^2 = 1 - \sum_{k=1}^{m} \hat{u}_{k,j}^2 \geq 1 - \sum_{k=1}^{m}\sum_{j=m+1}^{d(n)} \hat{u}_{k,j}^2, \quad j > m, \tag{73}$$

it follows from (72) that

$$\sum_{k\in H_{r+1}} \hat{u}_{k,j}^2 = 1 + \mathrm{O}_{\mathrm{a.s}}\left\{\frac{d(n)}{n\delta_r}\right\}, \quad j = m+1, \cdots, [n \wedge d(n)]. \tag{74}$$

Now consider the second case $\lim_{n\to\infty}\frac{d(n)}{n} = c$ $(0 < c < \infty)$. Note that the subspace consistency of the sample eigenvectors in (70) only depends on the the asymptotic properties of the sample eigenvalues $\hat{\lambda}_j$, $j = 1, \cdots, m$. According to Section 7.3.1, the asymptotic properties of $\hat{\lambda}_j$, $j = 1, \cdots, m$, only depends $\frac{d(n)}{n\delta_r^{(n)}} \to 0$, and is the same as in the first case $\lim_{n\to\infty}\frac{d(n)}{n} = \infty$. Thus (70) remains valid here.

However, the subspace consistency of the other eigenvectors also depends on $\hat{\lambda}_j$, $j > m$, whose properties are different from the first case. In fact (71) and (74) respectively become that as $n \to \infty$,

$$\sum_{k\in H_r} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_r^{(n)}}{\delta_{r-1}^{(n)}}\right\} \vee \mathrm{O}_{\mathrm{a.s}}\left\{\frac{1}{\delta_r^{(n)}}\right\}, \quad j \in H_r, \tag{75}$$

$$\sum_{k\in H_{r+1}} \hat{u}_{k,j}^2 = 1 + \mathrm{O}_{\mathrm{a.s}}\left\{\frac{1}{\delta_r^{(n)}}\right\}, \quad j = m+1, \cdots, [n \wedge d(n)]. \tag{76}$$

In order to obtain (75), following the first step proof procedure in Section 7.4.1, we only need to show that as $n \to \infty$,

$$\sum_{k=1}^{m}\sum_{j=m+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}}\left\{\frac{1}{\delta_r^{(n)}}\right\}. \tag{77}$$

Since $\lim_{n\to\infty}\frac{d(n)}{n} = c$ $(0 < c < \infty)$, then (72) becomes (77). In addition, it follows from (73) and (77) that (76) is established.

Note that we can combine the first and the second cases together as follows. If $\frac{d(n)}{n} \to c$, $0 < c \leq \infty$, then the combination of (71) and (75) provides

$$\sum_{k\in H_r} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_r^{(n)}}{\delta_{r-1}^{(n)}}\right\} \vee \mathrm{O}_{\mathrm{a.s}}\left\{\frac{d(n)}{n\delta_r^{(n)}}\right\}, \quad j \in H_r,$$

and the combination of (74) and (77) yields

$$\sum_{k\in H_{r+1}} \hat{u}_{k,j}^2 = 1 + \mathrm{O}_{\mathrm{a.s}}\left\{\frac{d(n)}{n\delta_r}\right\}, \quad j = m+1,\cdots,[n\wedge d(n)].$$

In addition, (70) remains valid for both cases. Thus it follows from (33) that we have finished the proof of the second bullet point in Scenario (a).

Finally, consider the last case $\lim_{n\to\infty}\frac{d(n)}{n} = 0$. It is clear that (70) still holds. According to (33), in order to finish the proof of the first bullet point in Scenario (a), we only need to show that as $n\to\infty$,

$$\sum_{k\in H_r} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{\delta_r^{(n)}}{\delta_{r-1}^{(n)}} \vee \frac{1}{\delta_r^{(n)}}\right\}, \quad j\in H_r, \tag{78}$$

$$\sum_{k\in H_{r+1}} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{\mathrm{a.s}}\left\{\frac{1}{\delta_r^{(n)}}\right\}, \quad j = m+1,\cdots,[n\wedge d(n)]. \tag{79}$$

In fact, in order to prove (78) and (79), we need to replace (77) by that as $n\to\infty$,

$$\sum_{k=1}^{m}\sum_{j=m+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{o}_{\mathrm{a.s}}\left\{\frac{1}{\delta_r^{(n)}}\right\}. \tag{80}$$

It follows from (44) that

$$\sum_{k=1}^{m}\sum_{j=m+1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{j=1}^{m}\sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2.$$

We also have

$$\sum_{j=1}^{m}\sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2 = \sum_{l=1}^{r}\sum_{j\in H_l}\sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2.$$

Then in oder to obtain (80), we only need to prove that as $n\to\infty$,

$$\sum_{j\in H_r}\sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{o}_{\mathrm{a.s}}\left\{\frac{1}{\delta_r^{(n)}}\right\}, \tag{81}$$

$$\sum_{l=1}^{r-1}\sum_{j\in H_l}\sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{o}_{\mathrm{a.s}}\left\{\frac{1}{\delta_r^{(n)}}\right\}. \tag{82}$$

We now prove (81). Since the $j$-th diagonal entry of $S^TS$ is between its largest and smallest eigenvalue, then (48) becomes

$$\lambda_{\min}(\frac{1}{n}ZZ^T) \le \hat{\lambda}_j \sum_{k=1}^{d(n)} \frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2 = \sum_{k=1}^{d(n)} s_{k,j}^2 \le \lambda_{\max}(\frac{1}{n}ZZ^T), \quad j = 1,\cdots,d(n). \tag{83}$$

Since $\lim_{n\to\infty}\frac{d(n)}{n}=0$, it follows from Lemma 1 that $\lambda_{\min}(\frac{1}{n}ZZ^T)$ and $\lambda_{\max}(\frac{1}{n}ZZ^T)\overset{\text{a.s}}{\to}1$. In addition, since $\frac{\hat{\lambda}_j}{\lambda_j^{(n)}}\overset{\text{a.s}}{\to}1$ for $j=1,\cdots,m$ (Section 7.3.1), it follows from (83) that

$$\sum_{k=1}^{d(n)}\lambda_j^{(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2\overset{\text{a.s}}{\to}1,\quad j=1,\cdots,m. \tag{84}$$

Note that

$$\sum_{k=1}^{d(n)}\lambda_j^{(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2=\sum_{l=1}^{r-1}\sum_{k\in H_l}\lambda_j^{(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2+\sum_{k\in H_r}\lambda_j^{(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2+\sum_{k=m+1}^{d(n)}\lambda_j^{(n)}\frac{1}{\lambda_k^{(n)}}k\hat{u}_{k,j}^2. \tag{85}$$

According to (70), we have that

$$\sum_{j\in H_l}\sum_{k\in H_l}\hat{u}_{k,j}^2\overset{\text{a.s}}{\to}|H_l|,\quad l=1,\cdots,r-1,$$

which leads to

$$\sum_{l=1}^{r-1}|H_l|=\sum_{j=1}^{d(n)}\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\geq\sum_{l^*=1}^{r-1}\sum_{j\in H_{l^*}}\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\geq\sum_{l=1}^{r-1}\sum_{j\in H_l}\sum_{k\in H_l}\hat{u}_{k,j}^2\overset{\text{a.s}}{\to}\sum_{l=1}^{r-1}|H_l|.$$

Then it follows that

$$\sum_{j\in H_r}\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\leq\sum_{j=1}^{d(n)}\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2-\sum_{l^*=1}^{r-1}\sum_{j\in H_{l^*}}\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\overset{\text{a.s}}{\to}0. \tag{86}$$

According to (86), we have that

$$\sum_{l=1}^{r-1}\sum_{k\in H_l}\lambda_j^{(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2\leq\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\leq\sum_{j\in H_r}\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\overset{\text{a.s}}{\to}0,\quad j\in H_r. \tag{87}$$

Since $\frac{\lambda_j^{(n)}}{\lambda_k^{(n)}}\to 1$ for $k,j\in H_r$, it follows from (85) and (87) that

$$\sum_{k\in H_r}\hat{u}_{k,j}^2+\sum_{k=m+1}^{d(n)}\lambda_j^{(n)}\frac{1}{\lambda_k^{(n)}}\hat{u}_{k,j}^2\overset{\text{a.s}}{\to}1,\quad j\in H_r. \tag{88}$$

According to (86), we have that

$$\begin{aligned}1\geq\sum_{k\in H_r}\hat{u}_{k,j}^2+\sum_{k=m+1}^{d(n)}\hat{u}_{k,j}^2 &= 1-\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\\ &\geq 1-\sum_{j\in H_r}\sum_{l=1}^{r-1}\sum_{k\in H_l}\hat{u}_{k,j}^2\overset{\text{a.s}}{\to}1,\quad j\in H_r.\end{aligned}$$

Then it follows that

$$\sum_{k\in H_r} \hat{u}_{k,j}^2 + \sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 1, \quad j \in H_r. \tag{89}$$

Since $\lim_{n\to\infty} \frac{\lambda_j^{(n)}}{\lambda_k^{(n)}} > 1$ for $j \in H_r$ and $k \geq m+1$, then combining (88) and (89) gives

$$\sum_{k\in H_r} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 1, \quad j \in H_r,$$

which together with (88), yields

$$\sum_{k=m+1}^{d(n)} \lambda_j^{(n)} \frac{1}{\lambda_k^{(n)}} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 0, \quad j \in H_r. \tag{90}$$

Since $\lambda_k^{(n)} \to c_\lambda$ for $k \geq m+1$ and $\frac{\lambda_j^{(n)}}{\delta_r^{(n)}} \to 1$ for $j \in H_r$, it follows from (90) that as $n \to \infty$,

$$\sum_{j\in H_r} \sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{o}_{\text{a.s}} \left\{ \frac{1}{\delta_r^{(n)}} \right\},$$

which is (81).

We now show the proof of (82). According to (84), we have that for $j \in H_l$, $l = 1, \cdots, r-1$,

$$\sum_{k\in H_l} \lambda_j^{(n)} \frac{1}{\lambda_k^{(n)}} \hat{u}_{k,j}^2 + \sum_{k=m+1}^{d(n)} \lambda_j^{(n)} \frac{1}{\lambda_k^{(n)}} \hat{u}_{k,j}^2 \leq \sum_{k=1}^{d(n)} \lambda_j^{(n)} \frac{1}{\lambda_k^{(n)}} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 1. \tag{91}$$

Since $\lambda_j^{(n)} \frac{1}{\lambda_k^{(n)}} \to 1$ for $k, j \in H_l$, it follows from (70) that

$$\sum_{k\in H_l} \lambda_j^{(n)} \frac{1}{\lambda_k^{(n)}} \hat{u}_{k,j}^2 \to \sum_{k\in H_l} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 1, \quad k, j \in H_l,$$

which together with (91), yields

$$\sum_{k=m+1}^{d(n)} \lambda_j^{(n)} \frac{1}{\lambda_k^{(n)}} \hat{u}_{k,j}^2 \xrightarrow{\text{a.s}} 0, \quad j \in H_l,\ l = 1, \cdots, r-1. \tag{92}$$

Since $\lambda_k^{(n)} \to c_\lambda$ for $k \geq m+1$ and $\frac{\lambda_j^{(n)}}{\delta_l^{(n)}} \to 1$ for $j \in H_l$, it follows from (92) that as $n \to \infty$,

$$\sum_{j\in H_l} \sum_{k=m+1}^{d(n)} \hat{u}_{k,j}^2 = \mathrm{o}_{\text{a.s}} \left\{ \frac{1}{\delta_l^{(n)}} \right\}, \quad l = 1, \cdots, r-1. \tag{93}$$

Since $\delta_l^{(n)} \leq \delta_r^{(n)}$ for $l = 1, \cdots, r-1$, (82) then follows from (93).

#### 7.4.3 Scenario (c) in Theorem 1

Finally, for Scenario (c) where $\frac{d(n)}{n\delta_1^{(n)}} \to \infty$, $h$ in (40) equals to 0. Since $j_0 = 0$, then (40) becomes that as $n \to \infty$,

$$\max_{1 \le j \le [n \wedge d(n)]} \left\{ \frac{d(n)}{n\lambda_j} \hat{u}_{j,j}^2 \right\} = \mathrm{O}_{\mathrm{a.s}}(1),$$

which yields the strong inconsistency of the sample eigenvectors in Scenario (c).

## References

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

T. W. Anderson. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34(1):122–148, 1963.

T. W. Anderson. *An introduction to multivariate statistical analysis.* John Willey & Sons, New York, 1984.

Z. D. Bai and J. F. Yao. On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177, 2012.

Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, pages 1275–1294, 1993.

J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

M. Biehl and A. Mietzner. Statistical mechanics of unsupervised structure recognition. *Journal of Physics A: Mathematical and General*, 27:1885–1897, 1994.

G. Casella and J. T. Hwang. Limit expressions for the risk of James-Stein estimators. *Canadian Journal of Statistics*, 10(4):305–309, 1982.

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):603–680, 2013.

M. A. Girshick. On the sampling theory of roots of determinantal equations. *The Annals of Mathematical Statistics*, 10(3):203–224, 1939.

P. Hall, J. S. Marron, and A. Neeman. Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B*, 67(3):427–444, 2005.

D. C. Hoyle and M. Rattray. PCA learning for sparse high-dimensional data. *Europhysics Letters*, 62(1):117–123, 2003.

J.E. Jackson. *A user's guide to principal components*. John Willey & Sons, New York, 1991. ISBN 0471622672.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327, 2001.

I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009.

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

S. Jung, A. Sen, and J. S. Marron. Boundary behavior in high dimension, low sample size asymptotics of pca. *Journal of Multivariate Analysis*, 109:190–203, 2012.

V. Koltchinskii and K. Lounici. Normal approximation and concentration of spectral projectors of sample covariance. *arXiv:1504.07333*, 2015.

V. Koltchinskii and K. Lounici. Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Annals of Institute Henri Poincare, to appear*, 2016.

D. N. Lawley. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 43(1):128–136, 1956.

S. Lee, F. Zou, and F. A. Wright. Convergence and prediction of principal component scores in high-dimensional settings. *The Annals of Statistics*, 38(6):3605–3629, 2010.

Z. M. Ma. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 41(2):772–801, 2013.

B. Nadler. Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics*, 36(6):2791–2817, 2008. ISSN 0090-5364.

A. Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2012.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

D. Paul and I. Johnstone. Augmented sparse principal component analysis for high dimensional data. *Technical Report, UC Davis*, 2012.

P. Reimann, C. Broeck, and G. J. Bex. A Gaussian scenario for unsupervised learning. *Journal of Physics A: Mathematical and General*, 29:3521–3535, 1996.

D. Shen, H. P. Shen, and J. S. Marron. Consistency of sparse PCA in high dimension, low sample size contexts. *Journal of Multivariate Analysis*, 115:317–333, 2013.

D. Shen, H. P. Shen, and J. S. Marron. A general framework for consistency of principal component analysis: Supplement materials. *Available online at http://www.unc.edu/∼dshen/PCA/PCASupplement.pdf*, 2015.

T. Tao. 254A, Notes 3a: Eigenvalues and sums of hermitian matrices. *Available online at https://terrytao.wordpress.com/2010/01/12/254a-notes-3a-eigenvalues-and-sums-of-hermitian-matrices/*, 2010.

TLH Watkin and J-P Nadal. Optimal unsupervised learning. *Journal of Physics A: Mathematical and General*, 27(6):1899–1915, 1994.

H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006. ISSN 1061-8600.