Sparse PCA Asymptotics and Analysis of Tree Data

Dan Shen

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2012

Approved by:

J. S. Marron

Haipeng Shen

Andrew Nobel

Michael R. Kosorok

Shankar Bhamidi

ABSTRACT

DAN SHEN: Sparse PCA Asymptotics and Analysis of Tree Data.
(Under the direction of J. S. Marron and Haipeng Shen.)

This research covers two major areas. The first one is asymptotic properties of Principal Component Analysis (PCA) and sparse PCA. The second one is the application of functional data analysis to tree structured data objects.

A general asymptotic framework is developed for studying consistency properties of PCA. Assuming the spike population model, the framework considers increasing sample size, increasing dimension (or the number of variables) and increasing spike sizes (the relative size of the population eigenvalues). Our framework includes several previously studied domains of asymptotics as special cases, and for the first time allows one to investigate interesting connections and transitions among the various domains. This unification provides new theoretical insights.

Sparse PCA methods are efficient tools to reduce the dimension (or number of variables) of complex data. Sparse principal components (PCs) can be easier to interpret than conventional PCs, because most loadings are zero. We study the asymptotic properties of these sparse PC directions for scenarios with fixed sample size and increasing dimension (i.e. High Dimension, Low Sample Size (HDLSS)). We find a large set of sparsity assumptions under which sparse PCA is still consistent even when conventional PCA is strongly inconsistent. The consistency of sparse PCA is characterized along with rates of convergence. The boundaries of the consistent region are clarified using an oracle result.

Functional data analysis has been very successful in the analysis of data lying in standard Euclidean space, such as curve data. However, with recent developments in fields such as medical image analysis, more and more non-Euclidean spaces, such as tree-structured data, present great challenges to statistical analysis. Here, we use the Dyck path approach from probability theory to build a bridge between tree space and curve space to exploit the power

of functional data analysis to analyze data in tree space.

Table of Contents

List of Figures

Chapter 1

Introduction

This thesis contains two major areas. The first one is to study asymptotic properties of Principal Component Analysis (PCA) and sparse PCA. The second one is to apply functional data analysis to tree structured data objects.

High dimensionality has become a common feature of data encountered in many divergent fields, such as genomics, economics and finance. This provides modern challenges for statistical analysis. To cope with the high dimensionality, dimension reduction and sparsity constraints become interesting.

One approach is PCA. This is an important visualization and dimension reduction tool for high dimensional data. In Chapter 2, we develop a general asymptotic framework for studying consistency properties of PCA. Our frame-work 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 between them. We also establish the corresponding convergence rate within each region. Under general spike covariance models, larger dimension (or the number of variables) tends to discourage the consistency of PCA, while larger sample size and/or spike information (the relative size of the population eigenvalues) tends to encourage PCA consistency. Our framework nicely illustrates relationships among these three types of information in terms of dimension, sample size and spike size, and rigorously characterizes how their relationships

affect PCA consistency.

High Dimension, Low Sample Size (HDLSS) asymptotics are based on the limit as the dimension $d \to \infty$ with the sample size n being fixed. It was originally studied by Casella and Hwang (1982) in the context of James-Stein estimation. Ahn et al. (2007) first studied the HDLSS asymptotic properties of PCA. A comprehensive result of this type is Jung and Marron (2009). As shown in Johnstone and Lu (2009), exploitation of sparsity helps to recover consistency, even in contexts where conventional PCA is inconsistent. In Chapter 3, we study the asymptotic properties of sparse PCA in HDLSS settings, as in Shen et al. (2012a). Under the previously studied spike covariance assumption, we show that sparse PCA is consistent under the same large spike condition that was used to gain insight into for conventional PCA. Under a broad range of small spike conditions, we identify a large, new set of sparsity assumptions where sparse PCA is consistent, but conventional PCA is strongly inconsistent. The boundaries of the consistent region are clarified using an oracle result.

The second part of this dissertation studies tree structured data objects. Statistical analysis, including PCA, of data on non-Euclidean spaces, such as tree space, can be challenging, as seen in Wang and Marron (2007). An approach is to build a bridge between trees space (a non-Euclidean space) and curve space (standard Euclidean space). Then, we can exploit the power of functional data analysis to explore statistical properties of tree data sets. The *Dyck path representation*, a tool for asymptotic analysis of point processes, Harris (1952), provides such a bridge. Besides the Dyck path representation, we also develop a novel *branch length representation* to connect tree space and curve space. In addition, we present a *pruned tree* idea to statistically analyze properties of tree structured data at a range of scales. Projection is a fundamental tool in classical Functional Data Analysis (FDA), see Ramsay and Silverman (2002, 2005). We project on PC directions (Jolliffe, 2002) to explore population variation. In addition, partial least squares (PLS) (Wold et al., 1984), and canonical correlation analysis (CCA) directions (Härdle and Simar, 2007), are used here to study population relationship with age, and we use the Distance Weighted Discrimination (DWD) direction (Marron et al., 2007; Qiao et al., 2010), to explore population relationships with gender.

Chapter 2

A General Framework for Consistency of PCA

2.1 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 thesis, 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, which allows careful study of the connections among the various setups, and more importantly it investigates scenarios that have not been considered before, and offers new insights into the *consistency* (in the sense that the angle between estimated and population eigen direction tends to 0, or the inner product tends to 1) and *strong-inconsistency* (where the angle tends to $\frac{\pi}{2}$, 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 three domains:

(a) The first domain of asymptotics is the **classical** one, under which the sample size $n \to \infty$

and the dimension d is fixed (hence the ratio $\frac{n}{d} \to \infty$). For example, see Girshick (1939); Lawley (1956); Anderson (1963, 1984); Jackson (1991).

(b) The second domain considers **random matrix** theory, where both the sample size n and the dimension d increase to infinity, with the ratio $\frac{n}{d} \to c$, 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 (2006); Paul (2007); Nadler (2008); Johnstone and Lu (2009); Lee et al. (2010b); Benaych-Georges and Nadakuditi (2011) from the statistics literature.

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

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). 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 point of the present chapter* is that there are three critical features whose relationships drive the consistency properties of PCA, namely

(1) the *sample information*: the sample size n, which has a *positive* contribution to, i.e. *encourages*, the consistency of the sample eigenvectors.

(2) the *variable information*: the dimension d, which has a *negative* contribution to, i.e. *discourages*, the consistency of the sample eigenvectors.

(3) the *spike information*: the relative sizes of the several leading eigenvalues, which also has a positive contribution to the consistency.

Our general framework considers increasing sample size n, increasing dimension d, and increasing spike information, and clearly characterizes how their relationships determine the consistency and strong-inconsistency regions of PCA, along with the boundary between these two regions. In addition, our theorems demonstrate the transitions among the existing domains of asymptotics, and for the first time to the best of our knowledge, enable one to understand the connections among them. 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 information; the HDLSS domain ((c) above) fixes the sample size, and increases the dimension and the spike information, and thus are all boundary cases of our general framework.

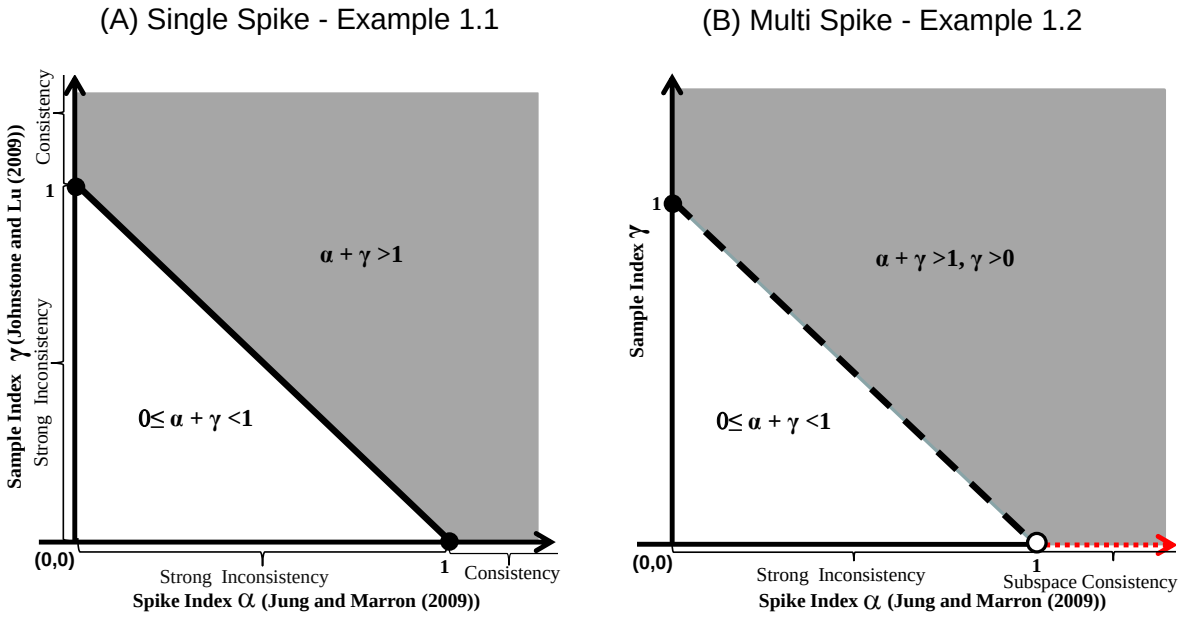

Figure 2.1: General consistency areas for PCA

Our theorems and corollaries are proven for general single and multiple component spike models, which are formally stated in Sections 2.2 and 2.3. For illustration purposes, we now consider two example models: the single-component spike model in Example 2.1.1 and the multiple-component spike model in Example 2.1.2. Within the context of these illustrating examples, we point out below the significant contributions of our results in comparison with

existing results. The comparisons and connections are graphically illustrated in Figure 2.1. Here some strong assumptions are made for the purpose of convenient comparison. For these two models, the three types of information and their relationships can be mathematically quantified by two indices, namely the *spike index* α and the *sample index* γ.

Example 2.1.1. *(Single-component spike model) Assume that $X_1, \ldots, X_n$ are random sample vectors from a d-dimensional normal distribution $N(0, \Sigma)$, where the sample size $n \sim d^\gamma$ ($\gamma \geq 0$ is defined as the sample index) and the covariance matrix Σ has the eigenvalues as*

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

where the constant α is defined as the spike index.

Our Corollary 2.2.1, applied to this example, shows that the maximal sample eigenvector is consistent when $\alpha + \gamma > 1$ (the grey region in Panel (A)), and strongly inconsistent when $0 \leq \alpha + \gamma < 1$ (the white triangle in Panel (A)). Our Theorem 2.3.1 explored behavior on the diagonal boundary $\alpha + \gamma = 1$. These very general new results connect with many existing ones:

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

  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 Panel (A) of Figure 2.1.

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

  (a) *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) *Nadler (2008) explored the interesting boundary case of $\alpha = 0, \gamma = 1$ (i.e. $\frac{d}{n} \to c$ for a constant c). This result appears in Panel (A) as the single solid circle $\gamma = 1$*

on the vertical axis.

- ***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) Jung et al. (2012) deeply explored limiting behavior at the boundary $\alpha = 1, \gamma = 0$. This result appears in Panel (A) as the single solid circle $\alpha = 1$ on the horizontal axis.

- ***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 the consistency of the other sample eigenvectors, and asymptotic properties of all the sample eigenvalues. Furthermore, we provide a new general connection between Previous Results II (a) and III (b), by doing a deeper explanation of assumptions on the boundary case - $\alpha + \gamma = 1$. We also establish technically challenging convergence rates within each region, which have not been studied before.*

Example 2.1.2. *(Multiple-component spike model) Assume that the covariance matrix Σ in Example 2.1.1 has the following eigenvalues*

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

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

Our Corollary 2.2.3, 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 Panel (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****:*

  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 Panel (B) of Figure 2.1.*

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

  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* $c \in (0, 1)$*. This result appears in Panel (B) as the solid circle* $\gamma = 1$ *on the vertical axis. Paul and Johnstone (2007a) considered a similar framework but from 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 separable 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). Convergence rates will also be established.*

The rest of Chapter 2 is organized as follows. Section 2.1.1 first introduces our notations and several relevant consistency concepts, and then provides intuitive explanations of our eigenvalue assumptions and the corresponding results. Section 2.2 presents the main theoretical results of Chapter 2, stating the asymptotic properties of the sample eigenvalues and eigenvectors under our general framework. Section 2.2.1 first considers the most general cases: multiple spike models with tiered eigenvalues. Sections 2.2.2 and 2.2.3 then discuss the implications and insights learned for single-component and multiple-component spike models (without tiered eigenvalues), respectively. Section 2.3 then investigates the property of PCA on the boundary between consistency and strong inconsistency regions under single spike models. Section 2.5 contains the technical proofs of the main theorems. Section 2.4 presents the some future work that we plan to investigate.

2.1.1 Notation

Let the population covariance matrix be Σ, whose eigen decomposition is

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

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

First, we make the following assumption about our sample:

Assumption 2.1.1. *$X_1, \ldots, X_n$ are a random sample from a d-dimensional normal distribution $N(0, \Sigma)$.*

Denote the data matrix by $X = [X_1, \ldots, X_n]_{d\times n}$ and the sample covariance matrix by $\hat{\Sigma} = n^{-1}XX^T$. Then, $\hat{\Sigma}$ can be similarly decomposed as

$$\hat{\Sigma} = \hat{U}\hat{\Lambda}\hat{U}^T, \tag{2.1}$$

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. Assume that $\{\xi_n : n = 1, \ldots, \infty\}$ is a sequence of random variables, and $\{a_n : n = 1, \ldots, \infty\}$ is a sequence of constant values.

- Denote $\xi_n = \mathrm{o}_{\mathrm{a.s}}(a_n)$ if $\lim_{n\to\infty} \frac{\xi_n}{a_n} = 0$ almost surely.
- Denote $\xi_n = \mathrm{O}_{\mathrm{a.s}}(a_n)$ if $\overline{\lim}_{n\to\infty} \left|\frac{\xi_n}{a_n}\right| \leq c$ almost surely, for some constant $c > 0$.
- Denote $\xi_n \overset{\mathrm{a.s}}{\asymp} a_n$ if $c_2 \leq \underline{\lim}_{n\to\infty} \frac{\xi_n}{a_n} \leq \overline{\lim}_{n\to\infty} \frac{\xi_n}{a_n} \leq c_1$ almost surely, for two constants $c_1 \geq c_2 > 0$.

2.1.2 Concepts

Below we list six important concepts relevant for consistency and strong inconsistency, some of which are modified from the related concepts given by Jung and Marron (2009); Shen et al. (2012a).

Let $\bar{u}_j$ be any sample based estimator of u_j for $j = 1, \ldots, n \wedge d$. For example, $\bar{u}_j = \hat{u}_j$, the jth sample eigenvector.

- **Consistency**: $\bar{u}_j$ is consistent with its population counterpart u_j if $|< \bar{u}_j, u_j >| = 1 + \mathrm{o}_{a.s}(1)$, i.e the angle between $\bar{u}_j$ and u_j tends to 0.
- **Consistency with convergence rate** a_n: $\bar{u}_j$ is consistent with u_j with the convergence rate a_n if $|< \bar{u}_j, u_j >| = 1 + \mathrm{O}_{\mathrm{a.s}}(a_n)$. For example, $a_n = \left(\frac{n\lambda_1}{d}\right)^{\frac{1}{2}}$.
- **Strong inconsistency**: $\bar{u}_j$ is strongly inconsistent with u_j if $|< \bar{u}_j, u_j >| = \mathrm{o}_{\mathrm{a.s}}(1)$, i.e the angle between $\bar{u}_j$ and u_j tends to $\frac{\pi}{2}$.
- **Strong inconsistency with convergence rate** a_n: $\bar{u}_j$ is strongly inconsistent with the convergence rate a_n if $|< \bar{u}_j, u_j >| = \mathrm{O}_{\mathrm{a.s}}(a_n)$.

Let H be an index set, e.g. $H = \{m + 1, \cdots, d\}$. Suppose $j \in H$.

- **Subspace consistency**: $\bar{u}_j$ is subspace consistent with u_j if

$$\mathrm{angle}\left(\bar{u}_j, \mathrm{span}\{u_k, k \in H\}\right) = \mathrm{o}_{a.s}(1),$$

where $\text{span}\{u_k, k \in H\}$ is the linear span generated by $\{u_k, k \in H\}$.

- **Subspace consistency with convergence rate** a_n: $\bar{u}_j$, $j \in H$, is subspace consistent with u_j with convergence rate a_n if

$$\text{angle}\,(\bar{u}_j, \text{span}\{u_k, k \in H\}) = \text{O}_{a.s}(a_n).$$

2.1.3 Assumptions

Our main theorems in Section 2.2.1 are very general, but also quite complicated. Insights and connections to previous work come from various special cases, which are carefully studied as corollaries in Sections 2.2.2 and 2.2.3. For these corollaries and main theorems, it is useful to develop a sequence of eigenvalue assumptions of increasing complexity.

Single Component Spike Models

Here we assume the maximal eigenvalue λ_1 dominates the other eigenvalues. The other eigenvalues are assumed to be asymptotically equivalent. For simplicity of notation, we assume they are asymptotically equivalent to 1. More specifically,

Assumption 2.1.2. *As $n \to \infty$, the population eigenvalues satisfy*

- $\frac{\lambda_2}{\lambda_1} \to 0, \qquad \lambda_2 \sim \cdots \sim \lambda_d \sim 1.$

As discussed in the Introduction, we consider the delicate balance among the positive *sample information* n, the positive *spike information* λ_1, and the negative *variable information* d, and characterize the various PCA consistency and strong-inconsistency regions.

Corollary 2.2.1 suggests that the asymptotic properties of the sample eigenvalues and eigenvectors depend on the relative strength of the positive information and the negative information, as measured by two ratios. First, $\frac{d}{n\lambda_1}$, corresponding to $d^{1-(\alpha+\gamma)}$ in Example 2.1.1, determines whether the maximal sample eigenvalue is separable from the other eigenvalues, and further determines the consistency of the maximal sample eigenvector. Second, $\frac{d}{n}$ determines the strong inconsistency of the higher order sample eigenvectors.

The following discussion and the scenarios in Corollary 2.2.1 are arranged according to a decreasing amount of positive information:

- Corollary 2.2.1(a): If the amount of positive information dominates the amount of negative information up to the maximal eigenvalue, i.e. $\frac{d}{n\lambda_1} \to 0$, the maximal sample eigenvector is consistent, and the other sample eigenvectors are subspace consistent.

- Corollary 2.2.1(b): In addition, if the amount of negative information dominates the amount of positive information for the eigenvalues whose indices are greater than 1, i.e. $\frac{d}{n} \to \infty$, then the corresponding sample eigenvectors are strongly-inconsistent.

- Corollary 2.2.1(c): On the other hand, if the amount of negative information always dominates, i.e. $\frac{d}{n\lambda_1} \to \infty$, then the sample eigenvalues are asymptotically indistinguishable, and the sample eigenvectors are strongly inconsistent.

Corollary 2.2.1 considers the cases where $n \to \infty$. Parallel results can be obtained for the fixed n cases (i.e. the HDLSS domain) as summarized in Corollary 2.2.2. In comparison with Jung and Marron (2009), we make more general assumptions on the population eigenvalues, and obtain the corresponding convergence rate results, which were not considered in Jung and Marron (2009).

Under the HDLSS domain, Assumption 2.1.2 on the eigenvalues becomes

Assumption 2.1.3. *As $d \to \infty$, the population eigenvalues satisfy*

- $\frac{\lambda_2}{\lambda_1} \to 0, \qquad \lambda_2 \sim \cdots \sim \lambda_d \sim 1.$

Multiple Component Spike Models

We now state a parallel series of assumptions for multiple spike models with m dominating spikes where $m \in [1, n \wedge d]$, and each of the first m eigenvalues is uniquely identifiable.

Assumption 2.1.4. *As $n \to \infty$, the population eigenvalues satisfy*

- $\overline{\lim}_{n\to\infty} \frac{\lambda_j}{\lambda_i} < 1,\ 1 \le i < j \le m,$

- $\frac{\lambda_{m+1}}{\lambda_m} \to 0$,
- $\lambda_{m+1} \sim \cdots \sim \lambda_d \sim 1$.

In the same spirit as Corollary 2.2.1, Corollary 2.2.3 states asymptotic properties of the sample eigenvalues and eigenvectors in a trichotomous manner, separated by the size of $\frac{d}{n\lambda_j}$, which again measures the relative strength of positive and negative information. We arrange the scenarios below and in Corollary 2.2.3 in the order of decreasing amount of positive information:

- Corollary 2.2.3(a): If the amount of positive information dominates the amount of negative information up to the mth spike, i.e. $\frac{d}{n\lambda_m} \to 0$, then each of the first m sample eigenvectors is consistent, and the additional ones are subspace consistent.
- Corollary 2.2.3(b): Otherwise, if the amount of positive information dominates the amount of negative information only up to the hth spike ($h \in [1, m]$), i.e. $\frac{d}{n\lambda_h} \to 0$ and $\frac{d}{n\lambda_{h+1}} \to \infty$, then each of the first h sample eigenvector is consistent, and each of the remaining higher-order sample eigenvectors is strongly-inconsistent.
- Corollary 2.2.3(c): Finally, if the amount of negative information always dominates, i.e. $\frac{d}{n\lambda_1} \to \infty$, then the sample eigenvalues are asymptotically indistinguishable, and the sample eigenvectors are strongly inconsistent.

Corollary 2.2.3 considers the cases where $n \to \infty$. Corresponding results can be obtained for the fixed n cases in Corollary 2.2.4. Assumption 2.1.4 then becomes

Assumption 2.1.5. *As $d \to \infty$, the population eigenvalues satisfy*

- $\lim_{d\to\infty} \frac{\lambda_j}{\lambda_i} = 0,\ 1 \le i < j \le m+1$,
- $\lambda_{m+1} \sim \cdots \sim \lambda_d \sim 1$.

Multiple Component Spike Models with Tiered Eigenvalues

Finally, we consider models where the m dominating eigenvalues can be grouped into r tiers. Within tiers, eigenvalues are either of the same limit, or in some cases have the same order.

Below we first consider cases with an increasing sample size n. To fix ideas, assume that there are q_l eigenvalues in the lth tier, and the positive integers q_l satisfy $\sum_{l=1}^{r} q_l = m$. Define $q_0 = 0$, $q_{r+1} = d - \sum_{l=1}^{r} q_l$, and the index set of the eigenvalues in the lth tier as

$$H_l = \left\{ \sum_{k=0}^{l-1} q_k + 1, \sum_{k=0}^{l-1} q_k + 2, \cdots, \sum_{k=0}^{l-1} q_k + q_l \right\}, \quad l = 1, \cdots, r+1. \tag{2.2}$$

In addition, we assume the eigenvalues in the lth tier have the same asymptotic behavior, represented in terms of a sequence $\delta_l (> 0)$, in the sense that

$$\begin{cases} \lim_{n\to\infty} \frac{\lambda_1}{\delta_1} = \cdots = \lim_{n\to\infty} \frac{\lambda_{q_1}}{\delta_1} = 1, \\ \lim_{n\to\infty} \frac{\lambda_{q_1+1}}{\delta_2} = \cdots = \lim_{n\to\infty} \frac{\lambda_{q_1+q_2}}{\delta_2} = 1, \\ \qquad \vdots \\ \lim_{n\to\infty} \frac{\lambda_{q_1+\cdots q_{r-1}+1}}{\delta_r} = \cdots = \lim_{n\to\infty} \frac{\lambda_{q_1+\cdots+q_r}}{\delta_r} = 1. \end{cases} \tag{2.3}$$

We impose the following assumption on the eigenvalues:

Assumption 2.1.6. *As $n \to \infty$, the population eigenvalues satisfy* (2.3) *and the following properties:*

- $\overline{\lim}_{n\to\infty} \frac{\delta_j}{\delta_i} < 1,\ 1 \le i < j \le r,$
- $\frac{\lambda_{m+1}}{\lambda_m} \to 0,$
- $\lambda_{m+1} \sim \cdots \sim \lambda_d \sim 1.$

Under the above setup, our main Theorem 2.2.1 suggests that the eigenvalues with the same limiting behavior can not be individually estimated consistently; their estimates are either subspace consistent with the linear space spanned by them, or the estimates are strongly inconsistent. Convergence rates depend on the eigenvalue ratios among the tiers, defined as $a_l = \max_{1\le k\le l} \frac{\delta_{k+1}}{\delta_k}$, for $l = 1, \cdots, r$, where $\delta_{r+1} = 1$.

Now for the fixed n cases, we assume that as $d \to \infty$, the first m eigenvalues fall into r tiers according to the following assumption (2.4):

$$\begin{cases} \lambda_1 \sim \cdots \sim \lambda_{q_1} \sim \delta_1, \\ \lambda_{q_1+1} \sim \cdots \sim \lambda_{q_1+q_2} \sim \delta_2, \\ \qquad \vdots \\ \lambda_{q_1+\cdots q_{r-1}+1} \sim \cdots \sim \lambda_m \sim \delta_r. \end{cases} \tag{2.4}$$

We then make the following eigenvalue assumption:

Assumption 2.1.7. *As $d \to \infty$, the population eigenvalues satisfy* (2.4) *and the following properties:*

- $\lim_{d\to\infty} \frac{\delta_j}{\delta_i} = 0,\ 1 \le i < j \le r,$
- $\frac{\lambda_{m+1}}{\lambda_m} \to 0,$
- $\lambda_{m+1} \sim \cdots \sim \lambda_d \sim 1.$

Different from (2.3) of Theorem 2.2.1, now with a fixed sample size n, the eigenvalues in the same tier are assumed to be of the same order, rather than of the same limit as assumed when n increases to ∞ (Theorem 2.2.1). Theorem 2.2.2 shows that 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 stated in Theorem 2.2.1.

2.2 Asymptotic Properties of PCA

We state main theorems for multiple-component spike models with tiered eigenvalues in Section 2.2.1, and corollaries for simpler single and multiple spike models in Sections 2.2.2 and 2.2.3, respectively. Technical proofs will be provided in Section 2.5.

2.2.1 Multiple Component Spike Models with Tiered Eigenvalues

Theorems 2.2.1 and 2.2.2 in this section consider the most general models: the m dominating eigenvalues are grouped into r tiers, where the eigenvalues within the same tier are either the

same or have the same limit or are of the same order, and the eigenvalues within different tiers have either different limits or are of different orders. The theorems show that the eigenvalues in the same tier can not be individually estimated consistently.

Theorem 2.2.1. *Under Assumptions 2.1.1 and 2.1.6, the results below hold.*

(a) If $\frac{d}{n\delta_r} \to 0$*, then the sample eigenvalues satisfy*

$$\begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1, & 1 \le j \le m, \\ \hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}, & m+1 \le j \le [n \wedge (d-m)], \\ \hat{\lambda}_j = \mathrm{O}_{\text{a.s}}\left(\frac{d}{n}\right), & [n \wedge (d-m)+1] \le j \le n \wedge d; \end{cases} \tag{2.5}$$

(There is no need to consider the last scenario above if $[n \wedge (d-m)+1] > n \wedge d$*.) In addition, the sample eigenvectors satisfy*

$$\begin{cases} angle\left(\hat{u}_j, span\{u_k, k \in H_l\}\right) = \left[\mathrm{o}_{a.s}\left(a_l\right) \vee \mathrm{O}_{\text{a.s}}\left(\frac{d}{n\delta_l}\right)\right]^{\frac{1}{2}}, j \in H_l, 1 \le l \le r, \\ angle\left(\hat{u}_j, span\{u_k, k \in H_{r+1}\}\right) = \left[\mathrm{o}_{a.s}\left(a_r\right) \vee \mathrm{O}_{\text{a.s}}\left(\frac{d}{n\delta_r}\right)\right]^{\frac{1}{2}}, m+1 \le j \le n \wedge d, \end{cases} \tag{2.6}$$

which shows that the sample eigenvector whose index is in H_l*,* $l \in [1, r]$*, is subspace consistent with the linear space spanned by the population eigenvectors whose labels are in* H_l*, with convergence rate* $\left(a_l \vee \frac{d}{n\delta_l}\right)^{\frac{1}{2}}$*; each of the rest sample eigenvectors is subspace consistent with the linear space spanned by* H_{r+1}*, with convergence rate* $\left(a_r \vee \frac{d}{n\delta_r}\right)^{\frac{1}{2}}$*.*

(b) If there exists a constant h*,* $1 \le h \le r$*, such that* $\frac{d}{n\delta_h} \to 0$ *and* $\frac{d}{n\delta_{h+1}} \to \infty$*, then the sample eigenvalues satisfy*

$$\begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1, & j \in H_l, 1 \le l \le h, \\ \hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}, & \sum_{l=1}^{h} q_l < j \le n \wedge d; \end{cases} \tag{2.7}$$

in addition, the sample eigenvectors satisfy

$$\begin{cases} angle\left(\hat{u}_j, span\{u_k, k \in H_l\}\right) = \left[\mathrm{o}_{a.s}\left(a_l\right) \vee \mathrm{O}_{\mathrm{a.s}}\left(\frac{d}{n\delta_l}\right)\right]^{\frac{1}{2}}, & j \in H_l, 1 \leq l \leq h, \\ |< \hat{u}_j, u_j >|^2 = \mathrm{O}_{\mathrm{a.s}}\left(\frac{n\lambda_j}{d}\right), & \sum_{l=1}^{h} q_l < j \leq n \wedge d, \end{cases} \tag{2.8}$$

which shows that the sample eigenvector whose index is in H_l, $1 \leq l \leq h$, is subspace consistent with the space spanned by the population eigenvectors whose labels are also in H_l, with the convergence rate $\left(a_l \vee \frac{d}{n\delta_l}\right)^{\frac{1}{2}}$; each of the rest of the sample eigenvectors is strongly inconsistent, with the convergence rate $\left(\frac{n\lambda_j}{d}\right)^{\frac{1}{2}}$.

(c) If $\frac{d}{n\delta_1} \to \infty$, then the sample eigenvalues satisfy

$$\hat{\lambda}_j \overset{\mathrm{a.s}}{\sim} \frac{d}{n},\ j = 1, \cdots, n \wedge d; \tag{2.9}$$

in addition, the sample eigenvectors satisfy

$$|< \hat{u}_j, u_j >|^2 = \mathrm{O}_{\mathrm{a.s}}\left(\frac{n\lambda_j}{d}\right), \quad j = 1, \cdots, n \wedge d, \tag{2.10}$$

which shows that the sample eigenvector $\hat{u}_j$ is strongly inconsistent with u_j, with the convergence rate $(\frac{n\lambda_j}{d})^{\frac{1}{2}}$, $j = 1, \cdots, n \wedge d$, respectively.

The following comments can be made for the results of Theorem 2.2.1.

- The cases covered by Theorem 2.2.1 were not studied by Paul (2007), which required the eigenvalues to be individually estimable.
- The asymptotic properties of the sample eigenvectors in Theorem 2.2.1 will not change under more general assumptions on the population eigenvalues. For example, if $\lim_{n\to\infty}\lambda_{m+1} = \cdots = \lim_{n\to\infty}\lambda_d = c$ with c being a constant, the condition $\frac{\lambda_{m+1}}{\lambda_m} \to 0$ can be relaxed to assuming only that λ_m is at least a constant away from λ_{m+1} such that $\underline{\lim}_{n\to\infty}\lambda_m > c$; in addition, if $\frac{d}{n} \to 0$ or ∞, the result "$\hat{\lambda}_j \overset{\mathrm{a.s}}{\sim} \frac{d}{n}$" can be strengthened to "$\hat{\lambda}_j \xrightarrow{\mathrm{a.s}} c\frac{d}{n}$".
- In Theorem 2.2.1, the dimension d can be fixed. In addition, consider $\overline{\lim}_{n\to\infty}\lambda_1 < \infty$

and the eigenvalues satisfying (2.3). Then, the results of Theorem 2.2.1(a) are consistent with the classical asymptotic subspace consistency results implied by Theorem 1 of Anderson (1963).

The above Theorem 2.2.1 considers the cases where $n \to \infty$. We can obtain parallel results for the fixed n cases (i.e. the HDLSS domain) as summarized below in Theorem 2.2.2. Note that we make more general assumptions on the population eigenvalues than Jung and Marron (2009), and obtain the corresponding convergence rate results, which were not considered before.

It is first worth pointing out that with n being fixed, we will consider convergence in probability, instead of almost surely. Consequently, we need to modify the convergence notations introduced in Section 2.1.1 to the following:

- Denote $\xi_d = \mathrm{o_p}(a_d)$ if $\lim_{d\to\infty} \frac{\xi_d}{a_d} = 0$ in probability.
- Denote $\xi_d = \mathrm{O_p}(a_d)$ if $\overline{\lim}_{d\to\infty} \left|\frac{\xi_d}{a_d}\right| \leq z$ in probability, where the random variable z satisfies $\mathrm{P}(0 < z < \infty) = 1$.
- Denote $\xi_d \overset{\mathrm{p}}{\sim} a_d$ if $z_2 \leq \underline{\lim}_{d\to\infty} \frac{\xi_d}{a_d} \leq \overline{\lim}_{d\to\infty} \frac{\xi_d}{a_d} \leq z_1$ in probability, where the two random variables satisfy $\mathrm{P}(0 < z_2 \leq z_1 < \infty) = 1$.

The consistency concepts in Section 2.1.2 are modified correspondingly.

Theorem 2.2.2. *Under Assumptions 2.1.1 and 2.1.7, the results below hold.*

(a) If there exists a constant h, $1 \leq h \leq r$, such that $\frac{d}{\delta_h} \to 0$ and $\frac{d}{\delta_{h+1}} \to \infty$, then the sample eigenvalues satisfy

$$\begin{cases} \hat{\lambda}_j \overset{\mathrm{p}}{\sim} \lambda_j, & j \in H_l, 1 \leq l \leq h, \\ \frac{n\hat{\lambda}_j}{d} \overset{\mathrm{p}}{\to} K, & \sum_{l=1}^{h} q_h < j \leq n, \quad \textit{with } K = \lim_{d\to\infty} d^{-1} \sum_{j=m+1}^{d} \lambda_j; \end{cases} \tag{2.11}$$

in addition, the sample eigenvectors satisfy

$$\begin{cases} \textit{angle}\left(\hat{u}_j, \textit{span}\{u_k, k \in H_l\}\right) = \left[\mathrm{O_p}\left(a_l \vee \frac{d}{n\delta_l}\right)\right]^{\frac{1}{2}}, & j \in H_l, 1 \leq l \leq h, \\ |< \hat{u}_j, u_j >|^2 = \mathrm{O_p}\left(\frac{n\lambda_j}{d}\right), & \sum_{l=1}^{h} q_h < j \leq n, \end{cases} \tag{2.12}$$

which shows that the sample eigenvector whose index is in H_l, $1 \le l \le h$, is subspace consistent with space spanned by the population eigenvector whose labels are also in H_l, with the convergence rate $\left(a_l \vee \frac{d}{n\delta_l}\right)^{\frac{1}{2}}$, $l = 1, \cdots, h$. For $\sum_{l=1}^{h} q_h < j \le n$, the sample eigenvector $\hat{u}_j$ is strongly inconsistent with u_j, with the convergence rate $\left(\frac{n\lambda_j}{d}\right)^{\frac{1}{2}}$.

(b) If $\frac{d}{\delta_1} \to \infty$, then the sample eigenvalues satisfy

$$\frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{p}} K, \ j = 1, \cdots, n; \tag{2.13}$$

in addition, the sample eigenvectors satisfy

$$|< \hat{u}_j, u_j >|^2 = \text{O}_\text{p}\left(\frac{n\lambda_j}{d}\right), \quad j = 1, \cdots, n, \tag{2.14}$$

which shows that the sample eigenvector $\hat{u}_j$ is strongly inconsistent with u_j, with the convergence rate $(\frac{n\lambda_j}{d})^{\frac{1}{2}}$, $j = 1, \cdots, n$, respectively.

The results of Theorem 2.2.2 suggest that, for the population eigenvalues in each tier (which are of the same order), the corresponding sample eigenvalues can not be separated asymptotically; on the other hand, for the eigenvalues from different tiers, the corresponding sample eigenvalues can be separated asymptotically. Hence, if each tier only has one eigenvalue, i.e. $r = m$ and $q_1 = \cdots = q_r = 1$, the sample eigenvalues are asymptotically separable, in which case we can strengthen the result "$\hat{\lambda}_j \overset{\text{p}}{\sim} \lambda_j$" in (2.11) to "$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{p}} \frac{\chi^n_2}{n}$", as stated in (2.15) of Corollary 2.2.2 and (2.16) of Corollary 2.2.4.

2.2.2 Single Component Spike Models

We now consider special cases of Theorem 2.2.1 with $m = 1$, which are stated in the following Corollary 2.2.1 for single-component spike models.

Corollary 2.2.1. *Under Assumptions 2.1.1 and 2.1.2, the following holds.*

(a) If $\frac{d}{n\lambda_1} \to 0$, then the sample eigenvalues satisfy (2.5) *with $m = 1$. In addition,* (2.6)

suggests that $\hat{u}_1$ *is consistent with the convergence rate* $\left(\frac{d}{n\lambda_1}\right)^{\frac{1}{2}}$, *and for* $j = 2, \cdots, n \wedge d$, $\hat{u}_j$ *is subspace consistent with convergence rate* $\left(\frac{d}{n\lambda_1}\right)^{\frac{1}{2}}$.

(b) *If* $\frac{d}{n\lambda_1} \to 0$ *and* $\frac{d}{n} \to \infty$, *then it follows from* (2.7) *that*

$$\frac{\hat{\lambda}_1}{\lambda_1} \xrightarrow{\text{a.s}} 1; \qquad \hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}, \quad 2 \leq j \leq n \wedge d,$$

and (2.8) *suggests that* $\hat{u}_1$ *is consistent with convergence rate* $\left(\frac{d}{n\lambda_1}\right)^{\frac{1}{2}}$; *for* $j \in [2, n \wedge d]$, $\hat{u}_j$ *is strongly inconsistent with rate* $\left(\frac{n\lambda_j}{d}\right)^{\frac{1}{2}}$.

(c) *If* $\frac{d}{n\lambda_1} \to \infty$, *then* (2.9) *and* (2.10) *remain the same: for* $j = 1, \cdots, n \wedge d$, $\hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}$, *and* $\hat{u}_j$ *is strongly inconsistent with convergence rate* $(\frac{n\lambda_j}{d})^{\frac{1}{2}}$.

Having stated the main results for single-component spike models, we now offer several remarks regarding the conditions assumed in Corollary 2.2.1 and the connections with the existing results about PCA consistency.

- In Corollary 2.2.1, the dimension d can be fixed. In addition, assume $\lim_{n\to\infty}\lambda_2 = \cdots = \lim_{n\to\infty}\lambda_d = c < \underline{\lim}_{n\to\infty}\lambda_1 < \overline{\lim}_{n\to\infty}\lambda_1 < \infty$ (for a constant c), which corresponds to the classical asymptotic framework considered by Anderson (1963). Theorem 1 of Anderson (1963) implies that the maximal sample eigenvector is consistent, and the rest of the sample eigenvectors are subspace consistent with their corresponding eigenvectors, which is consistent with our Corollary 2.2.1(a).

- Assuming fixed λ_1 and $\frac{d}{n} \to c$ with c being a constant, Nadler (2008); Johnstone and Lu (2009); Benaych-Georges and Nadakuditi (2011) obtained the results in *Previous Results II - the random matrix domain* in Example 2.1.1, which indicate that, as $n \to \infty$, the maximal sample eigenvector $\hat{u}_1$ is consistent when $\frac{d}{n} \to 0$, and inconsistent when $\frac{d}{n} \to \infty$. Our Corollary 2.2.1 includes this as a special case. In addition, Corollary 2.2.1 offers more than just relaxing the fixed λ_1 assumption: it characterizes how an increasing λ_1 interacts with the ratio $\frac{d}{n}$, derives the corresponding convergence rate, and also studies

the asymptotic properties of the higher order sample eigenvalues and eigenvectors, all of which have not been investigated before.

- The asymptotic properties of the sample eigenvectors as in Corollary 2.2.1 remain valid under more general assumptions on the population eigenvalues. For example, if $\lim_{n\to\infty}\lambda_2 = \cdots = \lim_{n\to\infty}\lambda_d = c$ with c being a constant, the condition $\frac{\lambda_2}{\lambda_1} \to 0$ can be relaxed to assuming only that λ_1 is at least a constant away from λ_2 such that $\underline{\lim}_{n\to\infty}\lambda_1 > c$. Hence, for the models considered by Johnstone and Lu (2009), our results of the maximal sample eigenvector are the same as theirs. In addition, if $\frac{d}{n} \to 0$ or ∞, the result "$\hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}$" in Corollary 2.2.1 can be strengthened to "$\hat{\lambda}_j \xrightarrow{\text{a.s}} c\frac{d}{n}$".

Considering the fixed n cases (i.e. the HDLSS domain), Theorem 2.2.2 reduces to the following Corollary 2.2.2 for single spike models.

Corollary 2.2.2. *Under Assumptions 2.1.1 and 2.1.3, the following holds.*

(a) If $\frac{d}{\lambda_1} \to 0$, *then* (2.11) *is strengthened to*

$$\begin{cases} \frac{\hat{\lambda}_1}{\lambda_1} \xrightarrow{\text{p}} \frac{\chi^n_2}{n}, \\ \frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{p}} K,\ 2 \le j \le n, \quad \textit{with} \quad K = \lim_{d\to\infty} \frac{\sum_{j=2}^{d}\lambda_j}{d}, \end{cases} \tag{2.15}$$

and (2.12) *suggests that* $\hat{u}_1$ *is consistent with convergence rate* $\left(\frac{d}{n\lambda_1}\right)^{\frac{1}{2}}$, *and for* $j \in [2, n]$, $\hat{u}_j$ *is strongly inconsistent with convergence rate* $\left(\frac{n\lambda_j}{d}\right)^{\frac{1}{2}}$.

(b) If $\frac{d}{\lambda_1} \to \infty$, *then* (2.13) *becomes* $\frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{p}} K$, *for* $j = 1, \cdots, n$, *and* (2.14) *suggests that* $\hat{u}_j$ *is strongly inconsistent with convergence rate* $(\frac{n\lambda_j}{d})^{\frac{1}{2}}$.

Note that in (2.15): if $n \to \infty$, then we have $\frac{\chi^n_2}{n} \xrightarrow{\text{a.s}} 1$. This is consistent with the results in (a) and (b) of Corollary 2.2.1, where $\frac{\hat{\lambda}_1}{\lambda_1} \xrightarrow{\text{a.s}} 1$.

2.2.3 Multiple Component Spike Models

We now consider multiple component spike models with m dominating spikes that are individually identifiable. Corollary 2.2.3 follows from Theorem 2.2.1 by setting $r = m$ and

$q_1 = \cdots = q_r = 1$, i.e. each of the m tiers only contains one eigenvalue.

Corollary 2.2.3. *Under Assumptions 2.1.1 and 2.1.4, the following holds.*

(a) If $\frac{d}{n\lambda_m} \to 0$, *then* (2.5) *remains valid for the sample eigenvalues.* (2.6) *then suggests that* $\hat{u}_j$, $j \in [1, m]$, *is consistent with convergence rate* $\left(a_j \vee \frac{d}{n\lambda_j}\right)^{\frac{1}{2}}$, *and* $\hat{u}_j$, $j \in [m+1, n \wedge d]$, *is subspace consistent with convergence rate* $\left(a_m \vee \frac{d}{n\lambda_m}\right)^{\frac{1}{2}}$.

(b) If there exists a constant h, $1 \le h \le m$, *such that* $\frac{d}{n\lambda_h} \to 0$ *and* $\frac{d}{n\lambda_{h+1}} \to \infty$, *then* (2.7) *becomes*

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1, \quad 1 \le j \le h; \quad \hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}, \quad h+1 \le j \le n \wedge d,$$

and (2.8) *suggests that* $\hat{u}_j$, $j \in [1, h]$, *is consistent with convergence rate* $\left(a_j \vee \frac{d}{n\lambda_j}\right)^{\frac{1}{2}}$, *and* $\hat{u}_j$, $j \in [h+1, n \wedge d]$, *is strongly inconsistent with convergence rate* $\left(\frac{n\lambda_j}{d}\right)^{\frac{1}{2}}$.

(c) If $\frac{d}{n\lambda_1} \to \infty$, *then* (2.9) *and* (2.10) *remain the same: for* $j = 1, \cdots, n \wedge d$, $\hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}$, *and* $\hat{u}_j$ *is strongly inconsistent with convergence rate* $(\frac{n\lambda_j}{d})^{\frac{1}{2}}$.

We now discuss the conditions needed in the corollary and how the results connect with existing results in the literature.

- In Corollary 2.2.3, the dimension d can be fixed. In addition, consider $\lim_{n\to\infty}\lambda_{m+1} = \cdots = \lim_{n\to\infty}\lambda_d = c < \underline{\lim}_{n\to\infty}\lambda_m < \overline{\lim}_{n\to\infty}\lambda_1 < \infty$. Then, Corollary 2.2.3(a) is consistent with the classical results implied by Theorem 1 of Anderson (1963).

- Considering fixed $\lambda_1, \cdots, \lambda_m$ and $\frac{d}{n} \to c$, where $c \in (0, 1)$, Paul (2007) obtained the results in *Previous Results II - the random matrix domain* in Example 2.1.2. As one can see, our Corollary 2.2.3 relaxes the assumptions of $\frac{d}{n} \to c \in (0, 1)$ and that $\lambda_1, \cdots, \lambda_m$ are fixed. In addition, we characterize how increasing $\lambda_1, \cdots, \lambda_m$ interact with the ratio $\frac{d}{n}$ along with the corresponding convergence rates, and study the asymptotic properties of the higher order sample eigenvalues and eigenvectors, all of which have not been investigated before.

- The asymptotic properties of the sample eigenvectors in Corollary 2.2.3 will not change under more general assumptions on the population eigenvalues, as discussed after Theorem 2.2.1.

Corollary 2.2.4 below considers fixed n, and follows from Theorem 2.2.2.

Corollary 2.2.4. *Under Assumptions 2.1.1 and 2.1.5, the following holds.*

(a) If there exists a constant h, $1 \le h \le m$, such that $\frac{d}{\lambda_h} \to 0$ and $\frac{d}{\lambda_{h+1}} \to \infty$, then (2.11) *is strengthened to*

$$\begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{p}} \frac{\chi^2_n}{n}, & 1 \le j \le h, \\ \frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{p}} K, & h+1 \le j \le n. \end{cases} \tag{2.16}$$

(2.12) then suggests that $\hat{u}_j$, $j \in [1, h]$, is consistent with convergence rate $\left(a_j \vee \frac{d}{n\lambda_j}\right)^{\frac{1}{2}}$, and $\hat{u}_j$, $j \in [h+1, n]$, is strongly inconsistent with convergence rate $\left(\frac{n\lambda_j}{d}\right)^{\frac{1}{2}}$.

(b) If $\frac{d}{\lambda_1} \to \infty$, then (2.13) *and* (2.14) *remain valid: for $j = 1, \cdots, n$, $\frac{n\hat{\lambda}_j}{d} \xrightarrow{\text{p}} K$, and $\hat{u}_j$ is strongly inconsistent with convergence rate $(\frac{n\lambda_j}{d})^{\frac{1}{2}}$.*

We remark that in (2.16), if $n \to \infty$, then we have $\frac{\chi^n_2}{n} \xrightarrow{\text{a.s}} 1$. This is consistent with the result $\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1$, $j \in [1, h]$, in (a) and (b) of Corollary 2.2.3.

2.3 Balanced Positive and Negative Information

The theorems and corollaries in Section 2.2 characterize the asymptotic properties of PCA under our general framework when either positive information or negative information dominates the other one. We now consider the transient cases where positive information and negative information is balanced, i.e. of the same asymptotic order. Theorem 2.3.1 states the corresponding results for single component spike models. We also discuss the connection with existing results in Nadler (2008); Jung et al. (2012). Similar considerations for multiple component spike models are also very interesting, but beyond the scope of the present thesis.

Theorem 2.3.1. *In addition to Assumption 2.1.1, we also assume that as $n \to \infty$, the population eigenvalues have the following properties:*

- $\frac{\lambda_2}{\lambda_1} \to 0$,
- $\lambda_j \to c_\lambda$, $j = 2, \cdots, d$, *for a constant* c_λ.

If $\frac{d}{n\lambda_1} \to c \in (0, \infty)$, *then the sample eigenvalues satisfy*

$$\begin{cases} \frac{\hat{\lambda}_1}{\lambda_1} \xrightarrow{\text{a.s}} 1 + cc_\lambda, \\ \frac{n}{d}\hat{\lambda}_j \xrightarrow{\text{a.s}} c_\lambda, \ 2 \le j \le n \wedge d; \end{cases} \tag{2.17}$$

in addition, the sample eigenvectors satisfy

$$\begin{cases} |< \hat{u}_1, u_1 >|^2 = \frac{1}{1+cc_\lambda} + \mathrm{o}_{\text{a.s}}(1), \\ |< \hat{u}_j, u_j >|^2 = \mathrm{O}_{\text{a.s}}(\frac{n}{d}), \quad 2 \le j \le n \wedge d, \end{cases} \tag{2.18}$$

which shows that the limiting angle between the maximal sample eigenvector $\hat{u}_1$ and u_1 is between 0 and $\frac{\pi}{2}$, and each of the additional sample eigenvector $\hat{u}_j$ is strongly inconsistent with u_j, with the convergence rate $\left(\frac{n}{d}\right)^{\frac{1}{2}}$.

Below we comment on the results of Theorem 2.3.1.

- Theorems 2.2.1 and 2.3.1 together completely characterizes the phase transition behavior of the maximal sample eigenvector $\hat{u}_1$ as $\frac{d}{n\lambda_1}$ converges to a different limit: as $n \to \infty$, $\hat{u}_1$ starts from being consistent when $\frac{d}{n\lambda_1} \to 0$, to being in-between consistency and strong inconsistency (with the limiting angle between 0 and $\frac{\pi}{2}$) when $\frac{d}{n\lambda_1} \to c \in (0, \infty)$, and finally to being strongly inconsistent when $\frac{d}{n\lambda_1} \to \infty$.
- The results nicely complement existing results of Nadler (2008); Jung et al. (2012): Nadler (2008) considered cases with a constant λ_1 and $\frac{d}{n} \to c \in (0, \infty)$ as $n \to \infty$, and derived the absolute inner product between $\hat{u}_1$ and u_1; Jung et al. (2012) studied scenarios with fixed n and $\frac{d}{\lambda_1} \to c \in (0, \infty)$ as $d \to \infty$, and showed that the absolute inner product is random.

In the context of the illustrating Example 2.1.1, the results of Nadler (2008) correspond to the point on the horizontal axis with $\alpha = 0$ and $\gamma = 1$; the results of Jung et al. (2012) are for the point on the vertical axis with $\alpha = 1$ and $\gamma = 0$; finally, our results are for the solid line with $\alpha + \gamma = 1$, which separates the consistency and strong-inconsistency regions.

2.4 Future Work

There are several interesting problems that we will explore in the future. One is to extend our theorems to more general distributions. Another is to build the similar framework to study the asymptotic properties of functional PCA (Dauxois et al., 1982; Bosq, 2000; Hall and Hosseini-Nasab, 2006).

2.5 Proofs

In this section we provide technical proofs for the theorems and corollaries stated in Section 2.2. The theorems fall into two groups: (1) Theorems 2.2.1, 2.3.1, and Corollaries 2.2.1, 2.2.3 consider increasing sample size; (2) Theorem 2.2.2, and Corollaries 2.2.2, 2.2.4 consider the HDLSS settings where the sample size is fixed. Hence we only prove Theorems 2.2.1 and 2.3.1 in details below, and point out how the proof can be adjusted to prove Theorem 2.2.2.

In this thesis, 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. Hence, below in Section 2.5.1, we first describe an invariance property of this angle that will simplify our mathematical proof, and then state in Section 2.5.2 two lemmas that will be used in the proof. The proof process itself is organized as follows: some intuitive ideas are presented first in Section 2.5.3 to help understanding the proof, and then we prove the asymptotic properties of the sample eigenvalues and the sample eigenvectors in Sections 2.5.4 and 2.5.5, respectively.

2.5.1 Invariance Property of the Angle

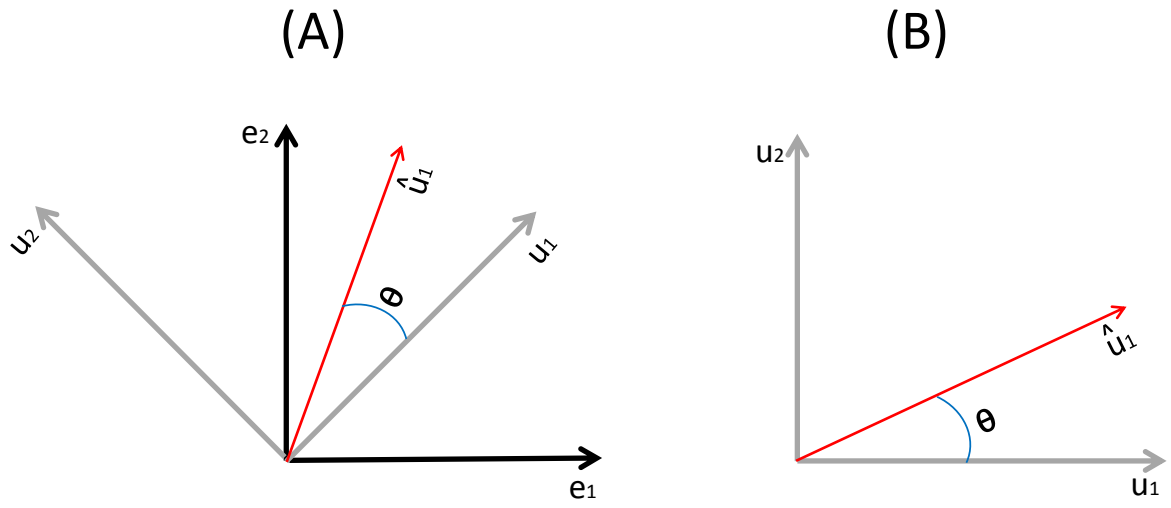

Figure 2.2: Independent between the angle and the basis choice

We note that the angle between the sample eigenvector and its population counterpart doesn't depend on the specific choice of the basis for the d-dimensional space. Because of this independence, we will choose the population eigenvectors u_i, $i = 1, \ldots, d$, as the basis of the d-dimensional space. Under this basis, the population covariance matrix of X_i, $i = 1, \ldots, n$, can be written as the following diagonal matrix:

$$\Sigma = \Lambda = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_d \end{pmatrix}. \tag{2.19}$$

This will simplify our mathematical analysis. Equivalently, without loss of generality, for the rest of the chapter, we assume that X_i, $i = 1, \ldots, n$, has the d-dimensional normal distribution with zero mean and the diagonal covariance matrix as in (2.19).

Now we use a toy example in Figure 2.2 to illustrate this invariance property of the angle. Assume that X_i has a 2-dimensional normal distribution with two eigenvectors u_1 and u_2, which are plotted in Panel (A) of Figure 2.2 under a specific basis set of e_1 and e_2. The θ denotes the angle between u_1 and its estimate $\hat{u}_1$. If we choose u_1 and u_2 as the new basis for

the 2-dimensional space, the angle between $\hat{u}_1$ and u_1 remains the same, as shown in Panel (B) of Figure 2.2.

2.5.2 Lemmas

The following two lemmas are needed to prove the theorems. The first one is about the asymptotic properties of the largest and smallest eigenvalues of the Wishart distribution, seen in Geman (1980); Silverstein (1985). The second one is the Wielandt's Inequality (Rao, 2002), which can be used to study asymptotic properties of eigenvalues of a random matrix.

Lemma 2.5.1. *Assume that $B = \frac{1}{s} V_s V_s^T$, where V_s is an $m \times s$ random matrix composed of iid standard normal random variables. As $s \to \infty$ and $\frac{m}{s} \to c \in [0, \infty)$, the largest and smallest non-zero eigenvalues of B, denoted as $\lambda_1(B)$ and $\lambda_{m \wedge s}(B)$, converge almost surely to $(1 + \sqrt{c})^2$ and $(1 - \sqrt{c})^2$, respectively.*

Remark: The results in the thesis cited are for $c \in (0, \infty)$, which can be easily extended to include the case of $c = 0$ by simple coupling arguments, as in (38) of Johnstone and Lu (2009).

Lemma 2.5.2. *(Wielandt's Inequality). If A, B are $m \times m$ real real symmetric matrices, then for all $k = 1, \ldots, m$,*

$$\left\{ \begin{array}{ccc} \lambda_k(A) & + & \lambda_m(B) \\ \lambda_{k+1}(A) & + & \lambda_{m-1}(B) \\ & \vdots & \\ \lambda_m(A) & + & \lambda_k(B) \end{array} \right\} \leq \lambda_k(A+B) \leq \left\{ \begin{array}{ccc} \lambda_k(A) & + & \lambda_1(B) \\ \lambda_{k-1}(A) & + & \lambda_2(B) \\ & \vdots & \\ \lambda_1(A) & + & \lambda_k(B) \end{array} \right\}.$$

2.5.3 Intuitive ideas

To help understand the formal proof in Sections 2.5.4 and 2.5.5, we first offer some intuitive discussion about the rationales behind our proof, which makes use of the connection between the sample covariance matrix $\hat{\Sigma}$ and its dual matrix $\hat{\Sigma}_D$.

To fix ideas, define $Z_i = \Lambda^{-\frac{1}{2}} X_i$ for $i = 1, \ldots, n$. Then, the Z_is are iid standard d-dimensional normal distribution. Denote the jth entry of Z_i as $z_{i,j}$ for $j = 1, \ldots, d$. For a fixed j, define

$$\widetilde{Z}_j = (z_{1,j}, \cdots, z_{n,j})^T, \tag{2.20}$$

which are iid standard n-dimensional normal distribution.

Note that the dual matrix of the sample covariance matrix can be expressed as

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

which has the same non-zero eigenvalues as the sample covariance matrix. Hence, we can study the asymptotic properties of the sample eigenvalues through the dual matrix, which we elaborate on below.

Single Component Spike Models

For simplicity, we start with single component spike models, and WLOG, assume that the eigenvalues satisfy $\lambda_1 > \lambda_2 = \cdots = \lambda_d = 1$. Then, the dual matrix can be rewritten as

$$\hat{\Sigma}_D = \frac{\lambda_1}{n} \widetilde{Z}_1 \widetilde{Z}_1^T + \frac{1}{n} \sum_{j=2}^{d} \widetilde{Z}_j \widetilde{Z}_j^T.$$

Denote the two summands in the above expression as A and B, respectively.

To understand the asymptotic properties of the sample eigenvectors, we first need to understand the asymptotic properties of the sample eigenvalues. For that purpose, we first note that the rank of the matrix A is 1, which suggests that A has only one non-zero eigenvalue, denoted as $\lambda_1(A) = n^{-1} \lambda_1 \widetilde{Z}_1^T \widetilde{Z}_1$. Furthermore, note that the matrix nB has the Wishart distribution, see e.g. (Muirhead (1982), p82); hence, Lemma 2.5.1 in Section 2.5.2 enables us to understand the asymptotic properties of the largest and smallest eigenvalues of the matrix nB, as well as the range of its non-zero eigenvalues. Finally, once we understand the asymptotic properties of the eigenvalues of A and B, *Wielandt's Inequality* (Lemma 2.5.2) allows us to study the asymptotic properties of the eigenvalues of the dual matrix $\hat{\Sigma}_D$.

Two ratios, $\frac{d}{n\lambda_1}$ and $\frac{d}{n}$, play crucial roles in the consistency of the sample eigenvectors. First, the ratio $\frac{d}{n\lambda_1}$ affects which one of the matrices A and B plays a dominating role in determining the maximal eigenvalue $\hat{\lambda}_1$ of $\hat{\Sigma}_D$, which then affects the consistency of the maximal sample eigenvector $\hat{u}_1$: if $\frac{d}{n\lambda_1} \to 0$, $\hat{\lambda}_1$ is determined by the matrix A and can be clearly separated from the other sample eigenvalues, which leads to the consistency of $\hat{u}_1$; if $\frac{d}{n\lambda_1} \to \infty$, $\hat{\lambda}_1$ is determined by the matrix B and can not be clearly separated from the other sample eigenvalues, which makes $\hat{u}_1$ strongly inconsistent; if $\frac{d}{n\lambda_1} \to c \in (0, \infty)$, then it is not clear which one matrix is dominating, and $\hat{u}_1$ is neither consistent nor strong inconsistent.

In addition, the ratio $\frac{d}{n}$ determines the consistency of the other sample eigenvectors: If $\frac{d}{n} \to 0$, they are subspace consistent with the subspace spanned by the corresponding population eigenvectors u_i, $i \geq 2$; If $\frac{d}{n} \to \infty$, then they are strongly inconsistent.

Multiple Component Spike Models

We now discuss the general ideas behind the proof for multiple component spike models. For simplicity, we assume that the first m eigenvalues can be grouped into two tiers, such that $\lambda_1 = \cdots = \lambda_q = \delta_1 \gg \lambda_{q+1} = \cdots = \lambda_m = \delta_2 \gg \lambda_{m+1} = \cdots = \lambda_d = 1$. Then, the dual matrix can be written as

$$\hat{\Sigma}_D = \frac{1}{n}\sum_{j=1}^{q} \delta_1 \widetilde{Z}_j \widetilde{Z}_j^T + \frac{1}{n}\sum_{j=q+1}^{m} \delta_2 \widetilde{Z}_j \widetilde{Z}_j^T + \frac{1}{n}\sum_{j=m+1}^{d} \widetilde{Z}_j \widetilde{Z}_j^T.$$

We denote the sum of the first two matrices in the above decomposition as A, and the third matrix as B.

Using similar arguments to those in Section 2.5.3, the consistency properties of the sample eigenvectors will depend on three ratios, $\frac{d}{n\delta_1}$, $\frac{d}{n\delta_2}$ and $\frac{d}{n}$, in the following manner:

- The ratio $\frac{d}{n\delta_1}$ determines the consistency of the sample eigenvectors in the first tier: If $\frac{d}{n\delta_1} \to 0$, the first q sample eigenvalues can be clearly separated from the other sample eigenvalues, which results in the subspace consistency of the corresponding sample vectors; If $\frac{d}{n\delta_1} \to \infty$, these sample eigenvalues can not be clearly separated from the others, and it follows that the corresponding sample eigenvectors are strongly inconsistent.

- The ratio $\frac{d}{n\delta_2}$ determines the consistency of the eigenvectors in the second tier: If $\frac{d}{n\delta_2} \to 0$, the sample eigenvalues in the second tier are separable from the others, and the corresponding sample eigenvectors are subspace consistent; If $\frac{d}{n\delta_2} \to \infty$, the sample eigenvalues in the second tier are separable from the others, which makes the corresponding sample eigenvectors strongly inconsistent.

- Finally, the ratio $\frac{d}{n}$ determines the consistency of the other sample eigenvectors: If $\frac{d}{n} \to 0$, the other sample eigenvectors are subspace consistent with the subspace spanned by the u_i, $i \geq m+1$; If $\frac{d}{n} \to \infty$, they are strongly inconsistent.

Different combinations of the limits of the three ratios will give us the various scenarios considered in Theorem 2.2.1.

2.5.4 Asymptotic Properties of the Sample Eigenvalues

We are now in a position to formally prove Theorem 2.2.1. In this section, we first derive the asymptotic properties of the sample eigenvalues, which will be used in studying the consistency of the sample eigenvectors in Section 2.5.5.

We consider general cases where the first m eigenvalues can be grouped into r tiers, and WLOG assume that $\lambda_1 = \cdots = \lambda_{q_1} = \delta_1, \cdots, \lambda_{\sum_{k=0}^{r-1} q_k+1} = \cdots = \lambda_m = \delta_r$ where $q_0 = 0$ and q_k are positive integers for $k \geq 1$. In addition, we assume that each ratio δ_j/δ_i, where $1 \leq i < j \leq r$, converges to a constant less than 1 as $n \to \infty$. (The following arguments can be extended to cases where only the upper limits of the ratios exist as stated in the theorems, through taking a converging subsequence of the diverging sequence of n.)

Now we will show the asymptotic properties of the sample eigenvalues as stated in Theorem 2.2.1. For that end, we first note that the dual matrix can be rewritten as the sum of two matrices as follows

$$\hat{\Sigma}_D = A + B, \quad \text{with} \quad A = \frac{1}{n}\sum_{j=1}^{m} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T, \quad B = \frac{1}{n}\sum_{j=m+1}^{d} \lambda_j \widetilde{Z}_j \widetilde{Z}_j^T, \tag{2.21}$$

where $\widetilde{Z}_j$ is defined in (2.20).

We then establish the asymptotic properties of the eigenvalues of A and B below in Lemmas 2.5.3 and 2.5.4, respectively. Finally, the asymptotic properties of the sample eigenvalues of Σ, which are the same as the eigenvalues of the dual matrix, naturally follow (Section 2.5.4).

Asymptotic Properties of the Eigenvalues of the Matrix A

Lemma 2.5.3. *As $n \to \infty$, the eigenvalues of the matrix A in (2.21) satistfy*

$$\frac{\lambda_k(A)}{\lambda_k} \xrightarrow{\text{a.s}} 1, \quad \textit{for} \quad k = 1, \cdots, m. \tag{2.22}$$

Proof. We first construct the dual matrix A^* of the matrix A. For $i = 1, \cdots, n$, let X_i^* be the m-dimensional random vector formed by the first m entries of X_i, i.e. $X_i^* = (I_m, 0_{m\times(d-m)})X_i$. Then, X_i^* is normal with mean zero and the following covariance matrix Σ^*:

$$\Sigma^* = \Lambda^* = \begin{pmatrix} \lambda_1^{-\frac{1}{2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m^{-\frac{1}{2}} \end{pmatrix}.$$

Defined the sample covariance matrix of X_i^* as

$$\begin{aligned} A^* &= \frac{1}{n}\sum_{j=1}^{m} X_i^*(X_i^*)^T \\ &= \lambda_1 \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n z_{i,1}^2 & \cdots & (\frac{\lambda_m}{\lambda_1})^{\frac{1}{2}}\frac{1}{n}\sum_{i=1}^n z_{i,1}z_{i,m} \\ \vdots & \ddots & \vdots \\ (\frac{\lambda_m}{\lambda_1})^{\frac{1}{2}}\frac{1}{n}\sum_{i=1}^n z_{i,1}z_{i,m} & \cdots & \frac{\lambda_m}{\lambda_1}\frac{1}{n}\sum_{i=1}^n z_{i,m}^2 \end{pmatrix}, \end{aligned} \tag{2.23}$$

where the $z_{i,j}$'s are defined in (2.20) and are iid standard normal random variables.

Note that the above matrix A^* is the dual matrix of A, then A and A^* share the same non-zero eigenvalues. Below we study the eigenvalues of the dual matrix A^*.

The iid properties of the $z_{i,j}$'s suggest that

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

which can be combined with (2.23) to suggest that

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

where $b_k = \lim_{n\to\infty}\frac{\lambda_k}{\lambda_1} \le 1$, $k = 1, \cdots, m$. It then follows that

$$\frac{\lambda_1(A)}{\lambda_1} = \frac{\lambda_1(A^*)}{\lambda_1} \xrightarrow{\text{a.s}} 1, \quad \text{as} \quad n \to \infty.$$

Similarly, for $k = 2, \cdots, m$, we have that

$$\frac{\lambda_1(\frac{1}{n}\sum_{j=k}^{m}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T)}{\lambda_k} \xrightarrow{\text{a.s}} 1, \quad \text{as} \quad n \to \infty. \tag{2.24}$$

Next we try to derive upper and lower bounds for $\lambda_k(A)$. First, Lemma 2.5.2 suggests that, for $k \ge 2$,

$$\lambda_k(A) = \lambda_k(\frac{1}{n}\sum_{j=1}^{m}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T) \le \lambda_1(\frac{1}{n}\sum_{j=k}^{m}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T) + \lambda_k(\frac{1}{n}\sum_{j=1}^{k-1}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T). \tag{2.25}$$

Since the rank of $\frac{1}{n}\sum_{j=1}^{k-1}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T$ is at most $k-1$, then it follows that

$$\lambda_k(\frac{1}{n}\sum_{j=1}^{k-1}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T) = 0,$$

which can be combined with (2.24) to show that

$$\frac{\lambda_k(A)}{\lambda_k} \le 1, \quad \text{as} \quad n \to \infty. \tag{2.26}$$

Now for the lower bound, the expression (5.9) in Jung and Marron (2009) suggests that

$$\lambda_1(\frac{\lambda_k}{n}\widetilde{Z}_k\widetilde{Z}_k^T) + \lambda_n(\frac{1}{n}\sum_{j=k+1}^{m}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T) \leq \lambda_k(A). \tag{2.27}$$

Given that the rank of $\frac{1}{n}\sum_{j=k+1}^{m}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T$ is at most m with $m < n$, then we know that

$$\lambda_n(\frac{1}{n}\sum_{j=k+1}^{m}\lambda_j\widetilde{Z}_j\widetilde{Z}_j^T) = 0, \tag{2.28}$$

which, together with (2.27), suggests that as $n \to \infty$,

$$\frac{\lambda_k(A)}{\lambda_k} \geq \frac{1}{\lambda_k}\lambda_1(\frac{\lambda_k}{n}\widetilde{Z}_k\widetilde{Z}_k^T). \tag{2.29}$$

In addition, note that as $n \to \infty$,

$$\frac{1}{\lambda_k}\lambda_1(\frac{\lambda_k}{n}\widetilde{Z}_k\widetilde{Z}_k^T) = \frac{1}{n}\widetilde{Z}_k^T\widetilde{Z}_k \xrightarrow{\text{a.s}} 1, \tag{2.30}$$

which, together with (2.29), gives that

$$\frac{\lambda_k(A)}{\lambda_k} \geq 1, \quad \text{as} \quad n \to \infty. \tag{2.31}$$

The combination of (2.26) and (2.31) suggests (2.22). □

Asymptotic Properties of the Eigenvalues of B

Lemma 2.5.4. *As $n \to \infty$, the eigenvalues of the matrix B as defined in* (2.21) *satisfy*

$$\lambda_k(B) \overset{\text{a.s}}{\sim} \frac{d}{n},\ k = 1, \cdots, n \wedge (d-m), \tag{2.32}$$

Proof. We start with assuming that $\lambda_{m+1} = \cdots = \lambda_d = 1$. WLOG, we assume that $\frac{d}{n}$ has a limit as $n \to \infty$. (Otherwise, we can always choose a subsequence of $\{n\}$ such that $\frac{d}{n}$ has a

limit.) Furthermore, we assume that $\frac{d}{n} \to c \leq 1$. (If $c > 1$, we can consider the dual matrix of B, whose dimension is $(d-m)$-by-$(d-m)$, and study its eigenvalues.)

Define $V = [\widetilde{Z}_{m+1}, \widetilde{Z}_{m+2}, \cdots, \widetilde{Z}_d]$, whose dimension is n-by-$(d-m)$. It then follows that $B = \frac{1}{n}VV^T = \frac{d-m}{n}(\frac{1}{d-m}VV^T)$.

Lemma 2.5.1 suggests that $\lambda_1(\frac{1}{d-m}VV^T) \overset{\text{a.s}}{\sim} \lambda_{n\wedge(d-m)}(\frac{1}{d-m}VV^T) \overset{\text{a.s}}{\sim} 1$. It follows that

$$\lambda_k(B) = \frac{d-m}{n}\lambda_k(\frac{1}{d-m}VV^T) \overset{\text{a.s}}{\sim} \frac{d}{n},\ k = 1, \cdots, n \wedge (d-m), \quad \text{as } n \to \infty,$$

which then yields (2.32).

The above arguments remain valid for cases where $\lambda_{m+1} \sim \lambda_d \sim 1$. □

Asymptotic Properties of the Sample Eigenvalues

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

According to Lemma 2.5.2, we have that

$$\lambda_j(A) + \lambda_n(B) \leq \hat{\lambda}_j \leq \lambda_j(A) + \lambda_1(B),$$

which suggests

$$\frac{1}{\lambda_j}\lambda_j(A) + \frac{1}{\lambda_j}\lambda_n(B) \leq \frac{\hat{\lambda}_j}{\lambda_j} \leq \frac{1}{\lambda_j}\lambda_j(A) + \frac{1}{\lambda_j}\lambda_1(B). \tag{2.33}$$

In addition, note that Lemma 2.5.4 shows that $\frac{1}{\lambda_j}\lambda_n(B) \leq \frac{1}{\lambda_j}\lambda_{n\wedge(d-m)}(B) \overset{\text{a.s}}{\sim} \frac{d}{\lambda_j n}$ and $\frac{1}{\lambda_j}\lambda_1(B) \overset{\text{a.s}}{\sim} \frac{d}{\lambda_j n}$. Below we consider three scenarios separately.

First, if there exists $h \in [1, r]$ such that $\frac{d}{n\delta_h} \to 0$, then $\frac{d}{n\lambda_j} \to 0$, for $j \in H_l$, $l = 1, \cdots, h$, where H_l is the index set of the eigenvalues in the lth tier. Thus, we have

$$\frac{1}{\lambda_j}\lambda_n(B) \xrightarrow{\text{a.s}} 0, \quad \text{and } \frac{1}{\lambda_j}\lambda_1(B) \xrightarrow{\text{a.s}} 0, \quad j \in H_l, l = 1, \cdots, h. \tag{2.34}$$

The above (2.34), together with (2.33) and Lemma 2.5.3, leads to

$$\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1, \quad j \in H_l, l = 1, \cdots, h. \tag{2.35}$$

Secondly, if $\frac{d}{n\delta_h} \to \infty$, then we have $d \gg n$ and $n \wedge (d-m) = n$. For $j \in H_l$, $l \geq h$, Lemma 2.5.2 suggests that

$$\frac{n}{d}\lambda_j(A) + \frac{n}{d}\lambda_n(B) \leq \frac{n}{d}\hat{\lambda}_j \leq \frac{n}{d}\lambda_j(A) + \frac{n}{d}\lambda_1(B). \tag{2.36}$$

Lemma 2.5.3, together with the condition $\frac{d}{n\delta_h} \to \infty$, suggests that $\frac{n}{d}\lambda_j(A) \xrightarrow{\text{a.s}} 0$ for $j \in H_l$ and $l \geq h$. Using Lemma 2.5.4, we have that $\frac{n}{d}\lambda_{n-m}(B) \overset{\text{a.s}}{\sim} 1$ and $\frac{n}{d}\lambda_1(B) \overset{\text{a.s}}{\sim} 1$. The above, combined with (2.36), suggests that

$$\hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}, \quad \sum_{l=1}^{h-1} q_l < j \leq n \wedge d, \quad \text{a.s.} \tag{2.37}$$

Finally, if $\frac{d}{n\delta_r} \to 0$, (2.34) suggests that $\frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1$, $1 \leq j \leq m$. In addition, Lemma 2.5.2 suggests that

$$\frac{n}{d}\lambda_{j+n-n\wedge(d-m)}(A) + \frac{n}{d}\lambda_{n\wedge(d-m)}(B) \leq \frac{n}{d}\hat{\lambda}_j \leq \frac{n}{d}\lambda_j(A) + \frac{n}{d}\lambda_1(B). \tag{2.38}$$

Note that rank of A is less than or equal to m and it means that for $j > m$, $\lambda_{j+n-n\wedge(d-m)}(A) = \lambda_j(A) = 0$. Furthermore, from Lemma 2.5.4, we have that $\frac{n}{d}\lambda_1(B) \overset{\text{a.s}}{\sim} \frac{n}{d}\lambda_{n\wedge(d-m)}(B) \overset{\text{a.s}}{\sim} 1$. Combining above with (2.38), we have $\hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}$, $m+1 \leq j \leq n \wedge (d-m)$. For $j \in [n \wedge (d-m) + 1 \leq j \leq n \wedge d]$, given that $\hat{\lambda}_j \leq \hat{\lambda}_{n\wedge(d-m)}$, it follows that $\hat{\lambda}_j = \mathrm{O}_{\text{a.s}}(\frac{d}{n})$.

The above arguments can be summarized as follow: if $\frac{d}{n\delta_r} \to 0$,

$$\begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s}} 1, & 1 \leq j \leq m \\ \hat{\lambda}_j \overset{\text{a.s}}{\sim} \frac{d}{n}, & m+1 \leq j \leq n \wedge d - m \\ \hat{\lambda}_j = \mathrm{O}_{\text{a.s}}(\frac{d}{n}), & n \wedge (d-m) + 1 \leq j \leq n \wedge d \end{cases}. \tag{2.39}$$

Note that if $n \wedge (d-m) + 1 > n \wedge d$, the last term disappears.

Combining (2.35), (2.37), and (2.39), we can get the corresponding results (2.5), (2.7), and (2.9) in Theorem 2.2.1.

2.5.5 Asymptotic Properties of the Sample Eigenvectors

We are now ready to derive the asymptotic properties of the sample eigenvectors

$$\hat{u}_j = (\hat{u}_{1,j}, \cdots, \hat{u}_{d,j})^T, \quad j = 1, \cdots, n \wedge d.$$

First, we state two results that simplify the proof. As discussed in Section 2.5.1, we choose the population eigevectors u_j $(j = 1, \cdots, d)$ 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 yields that

$$|< \hat{u}_j, u_j >|^2 = |< \hat{u}_j, e_j >|^2 = \hat{u}_{j,j}^2, \tag{2.40}$$

and for any index set H,

$$\cos\left[angle\left(\hat{u}_j, \text{span}\{u_k, k \in H\}\right)\right] = \sum_{k \in H} \hat{u}_{k,j}^2. \tag{2.41}$$

Define

$$\hat{U}_{i,j} = (\hat{u}_{k,l})_{k \in H_i, l \in H_j}, \quad 1 \le i, j \le r+1, \tag{2.42}$$

where H_i is defined in (2.2), $i = 1, \cdots, r+1$. Then, the sample eigenvectors matrix can be rewrote as following.

$$\hat{U} = [\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_d] = \begin{pmatrix} \hat{U}_{1,1} & \hat{U}_{1,2} & \cdots & \hat{U}_{1,r+1} \\ \hat{U}_{2,1} & \hat{U}_{2,2} & \cdots & \hat{U}_{2,r+1} \\ \vdots & \vdots & & \vdots \\ \hat{U}_{r+1,1} & \hat{U}_{r+1,2} & \cdots & \hat{U}_{r+1,r+1} \end{pmatrix}. \tag{2.43}$$

Scenario (b) in Theorem 2.2.1

Now consider the scenario (b) in Theorem 2.2.1 in that there exists a constant $h \in [1, r]$, such that $\frac{d}{n\delta_h} \to 0$ and $\frac{d}{n\delta_{h+1}} \to \infty$. Define $a_l = \max_{1\le k\le l} \frac{\delta_{k+1}}{\delta_k}$, $l = 1, \cdots, r$. From (2.41), in order to show subspace consistent with the convergence rate $\left(a_l \vee \frac{d}{n\delta_l}\right)^{\frac{1}{2}}$ in (2.7), we just need to show

$$\sum_{k\in H_l} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{a.s}(a_l) \wedge \mathrm{O}_{a.s}(\frac{d}{n\delta_l}), \quad j \in H_l, l = 1, \cdots, h. \tag{2.44}$$

The following proof is just to show (2.44) for $l = 1$. For $l = 2, \cdots, h$, the process is similar and we skip it to avoid the repetition.

Note that for $l = 1$, the left part of equation (2.44) is just the sum of square column elements of matrix $\hat{U}_{1,1}$, where $\hat{U}_{1,1}$ is defined (2.42). Thus, in order to show (2.44) for $l = 1$, we just need to show that the sum of square column elements of matrix $\hat{U}_{1,1}$ converges to 1 with the convergence rate $a_1 \vee \frac{d}{n\delta_1}$. The following proof contains two steps: the first one is to show sum of square column elements of matrix $\hat{U}_{1,1}$ converges to 1; the second step is to show the convergence rate $a_1 \vee \frac{d}{n\delta_1}$.

Now, we will show that the sum of square column entries of the matrix $\hat{U}_{1,1}$ converges to 1. Let $Z = (Z_1, \cdots, Z_n)$, where $Z_i = \Lambda^{-\frac{1}{2}} X_i$ as defined in Section 2.5.3. Denote $S = (s_{k,l})_{d\times d} = \Lambda^{-\frac{1}{2}} \hat{U} \hat{\Lambda}^{\frac{1}{2}}$ where $\hat{U}$ is the sample eigevector matrix and $\hat{\Lambda}$ is the sample eigenvalue matrix defined in (2.1), the eigendecomposition of the sample variance matrix $\hat{\Sigma}$. We can show that

$$SS^T = \frac{1}{n} ZZ^T.$$

Considering the k-th diagonal entry of the matrices on the two sides and noting that $s_{k,j} = \lambda_k^{-\frac{1}{2}} \hat{\lambda}_j^{\frac{1}{2}} \hat{u}_{k,j}$, we have the following

$$\frac{1}{n} \sum_{i=1}^{n} z_{i,k}^2 = \sum_{j=1}^{d} s_{k,j}^2 = \lambda_k^{-1} \sum_{j=1}^{d} \hat{\lambda}_j \hat{u}_{k,j}^2, \quad k = 1, \cdots, d. \tag{2.45}$$

As shown earlier, $\frac{1}{n}\sum_{i=1}^{n} z_{i,k}^2 \xrightarrow{\text{a.s}} 1$, which suggests that $\hat{u}_{k,j}^2 \le \frac{\lambda_k}{\hat{\lambda}_j}$ as $n \to \infty$. Then, as shown

in Section 2.5.4, the sample eigenvalues $\hat{\lambda}_j$ satisfy (2.7), which suggests that

$$\hat{u}_{k,j}^2 = \begin{cases} \mathrm{O}_{\mathrm{a.s}}(\frac{\lambda_k}{\lambda_j}) & j \in H_l, l = 1, \cdots, h \\ \mathrm{O}_{\mathrm{a.s}}(\frac{n\lambda_k}{d}) & j = \sum_{l=1}^{h} q_l + 1, \cdots, n \wedge d \end{cases}. \tag{2.46}$$

In addition, the k-th diagonal entry of $S^T S$ is less or equal than its largest eigenvalues. Since $S^T S$ share the same non-zero eigenvalues as $\frac{1}{n} Z^T Z$, we have

$$\hat{\lambda}_j \sum_{k=1}^{d} \lambda_k^{-1} \hat{u}_{k,j}^2 = \sum_{k=1}^{d} s_{k,j}^2 \leq \lambda_{\max}(\frac{1}{n} Z^T Z), \quad j \in H_l, l = 1, \cdots, h. \tag{2.47}$$

The cross products matrix $Z^T Z$ has a standard n-dimensional Wishart $W_n(d, I)$ distribution with d degrees of freedom and identity covariance matrix, see e.g. (Muirhead (1982), p82). Using lemma 2.5.2, we have

$$\lambda_{\max}(\frac{1}{n} Z Z^T) \overset{\mathrm{a.s}}{\sim} (\frac{d}{n}). \tag{2.48}$$

Using (2.35), (2.47), and (2.48), we have

$$\sum_{k=m+1}^{d} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}}(\frac{d}{n\lambda_j}), \quad j \in H_l, l = 1, \cdots, h. \tag{2.49}$$

Since $\lambda_j \ll \frac{d}{n}$, $j = \sum_l^h q_l + 1, \cdots, m$, (2.46) and (2.49), then we have

$$\sum_{k=\sum_l^h q_l + 1}^{d} \hat{u}_{k,j}^2 = \mathrm{O}_{\mathrm{a.s}}\left(\frac{d}{n\lambda_j}\right), \quad j \in H_l, l = 1, \cdots, h. \tag{2.50}$$

Note that $\lambda_k = \delta_1$, $k \in H_1$. and then from (2.45), for $k \in H_1$,

$$\begin{aligned} 1 + \mathrm{o}_{a.s}(1) = \frac{1}{n} \sum_{i=1}^{n} z_{i,k}^2 &= \lambda_k^{-1} \sum_{j=1}^{d} \hat{\lambda}_j \hat{u}_{k,j}^2 = \delta_1^{-1} \sum_{j=1}^{d} \hat{\lambda}_j \hat{u}_{k,j}^2 \\ &\leq \delta_1^{-1} \sum_{j \in H_1} \hat{\lambda}_1 \hat{u}_{k,j}^2 + \delta_1^{-1} \sum_{j \notin H_1} \hat{\lambda}_{q_1+1} \hat{u}_{k,j}^2 \\ &= \delta_1^{-1} (\hat{\lambda}_1 - \hat{\lambda}_{q_1+1}) \sum_{j \in H_1} \hat{u}_{k,j}^2 + \delta_1^{-1} \hat{\lambda}_{q_1+1}. \end{aligned} \tag{2.51}$$

In addition, from (2.35), we have $\delta_1^{-1}(\hat{\lambda}_1 - \hat{\lambda}_{q_1+1}) \xrightarrow{\text{a.s}} \delta_1^{-1}(\delta_1 - \delta_2) = (1-c)$ and $\delta_1^{-1}\hat{\lambda}_{q_1+1} = c + \mathrm{o}_{a.s}(1)$, where $c = \lim_{n\to\infty} \frac{\delta_2}{\delta_1} < 1$.

Combining above with (2.51), we have

$$\begin{aligned} 1 + \mathrm{o}_{a.s}(1) &\leq (1-c)\underline{\lim}_{n\to\infty} \sum_{j\in H_1} \hat{u}_{k,j}^2 + c \\ &\leq (1-c)\overline{\lim}_{n\to\infty} \sum_{j\in H_1} \hat{u}_{k,j}^2 + c \leq 1, \end{aligned}$$

which yields $\sum_{j\in H_1} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{a.s}(1), \quad k \in H_1$. It means that the sum of square row elements of $\hat{U}_{1,1}$ converges to 1. Since the sum of square row or column elements of $\hat{U}_{1,1}$ is less than or equal to 1, it follows that the sum of square column elements of $\hat{U}_{1,1}$ converges to 1.

The second step is to show that the sum of square column elements of $\hat{U}_{1,1}$ converge to 1 with the convergence rate $a_1 \vee \frac{d}{n\delta_1}$. Since we have showed that the sum of square row elements of $\hat{U}_{1,1}$ converges to 1, it follows that the sum of square row elements of $\hat{U}_{1,2}$ converges to 0. Furthermore, the sum of the square column elements of $\hat{U}_{1,2}$ converges to 0, as following:

$$\sum_{k\in H_1} \hat{u}_{k,j}^2 = \mathrm{o}_{a.s}(1), \quad j \in H_2. \tag{2.52}$$

WLOG, we assume that $\frac{\delta_3}{\delta_2} \to 0$. If the limit is greater than 0, we can combine index H_2 and H_3 together to check whether $\frac{\delta_4}{\delta_2} \to 0$ converges to 0. If not, we still combine index together until the big jump appears. Since $\frac{\delta_3}{\delta_2} \to 0$, (2.46) and (2.50), it follows that

$$\sum_{k\in H_3\cup\cdots H_{r+1}} \hat{u}_{k,j}^2 = \mathrm{o}_{a.s}(1), \quad j \in H_2. \tag{2.53}$$

From (2.52) and (2.53), we have

$$\sum_{k\in H_2} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{a.s}(1), \quad j \in H_2,$$

It means that sum of square column elements of $\hat{U}_{2,2}$ converges to 1. Since the sum of square row or column elements of $\hat{U}_{2,2}$ is less than or equal to 1, it follows that the sum of square

row elements of $\hat{U}_{2,2}$ converges to 1, as following

$$\sum_{j \in H_2} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{a.s}(1), \quad k \in H_2. \tag{2.54}$$

Since $\hat{\lambda}_j \xrightarrow{a.s} \lambda_j = \delta_2$, $j \in H_2$ and (2.54), it follows that for $k \in H_2$

$$\begin{aligned} 1 + \mathrm{o}_{a.s}(1) &= \frac{1}{n} \sum_{i=1}^{n} z_{i,k}^2 = \lambda_k^{-1} \sum_{j=1}^{d} \hat{\lambda}_j \hat{u}_{k,j}^2 \\ &\geq \delta_2^{-1} \sum_{j \in H_1} \hat{\lambda}_j \hat{u}_{k,j}^2 + \delta_2^{-1} \sum_{j \in H_2} \hat{\lambda}_j \hat{u}_{k,j}^2 \\ &= \delta_2^{-1} \sum_{j \in H_1} \hat{\lambda}_j \hat{u}_{k,j}^2 + 1 + \mathrm{o}_{a.s}(1), \end{aligned}$$

which yields $\delta_2^{-1} \sum_{j \in H_1} \hat{\lambda}_j \hat{u}_{k,j}^2 = \mathrm{o}_{a.s}(1)$, $k \in H_2$. Since $\hat{\lambda}_j \xrightarrow{a.s} \lambda_j = \delta_1$, $j \in H_1$, it follows that

$$\sum_{j \in H_1} \hat{u}_{k,j}^2 = \mathrm{o}_{a.s}(\frac{\delta_2}{\delta_1}), \quad k \in H_2,$$

which yields

$$\sum_{k \in H_2} \hat{u}_{k,j}^2 = \mathrm{o}_{a.s}(\frac{\delta_2}{\delta_1}), \quad j \in H_1. \tag{2.55}$$

In addition, from (2.46) and (2.50), we have

$$\sum_{k \in H_3 \cup \cdots H_{r+1}} \hat{u}_{k,j}^2 = \mathrm{o}_{a.s}(\frac{\delta_2}{\delta_1}), \quad j \in H_1. \tag{2.56}$$

From (2.55), (2.56) and $\frac{\delta_2}{\delta_1} \gg \frac{d}{n\delta_1}$, we have

$$\sum_{k \in H_1} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{a.s}(\frac{\delta_2}{\delta_1}) = 1 + \mathrm{o}_{a.s}(a_1) \wedge \mathrm{O}_{a.s}(\frac{d}{n\delta_1}), \quad j \in H_1.$$

Until now, we showed that the sum of square column elements of $\hat{U}_{1,1}$ converge to 1 with the convergence rate $a_1 \vee \frac{d}{n\delta_1}$ and it means that (2.44) is established for $l = 1$. The proof of (2.44) is similar for $l = 2, \cdots, h$. Thus, we have showed the subspace consistent in (2.8) in Theorem 2.2.1.

Now, we will show the strong inconsistency in (2.8) in Theorem 2.2.1. Using (2.46), we have

$$|< \hat{u}_j, u_j >|^2 = \hat{u}_{j,j}^2 = \mathrm{O_{a.s}}\left(\frac{n\lambda_j}{d}\right), \quad j = \sum_{l=1}^{h} q_l + 1, \cdots, n \wedge d. \tag{2.57}$$

Until now, we have finished the proof of scenario (b) in Theorem 2.2.1.

Scenario (a) in Theorem 2.2.1

Having proved the scenario (b) above, now consider the scenario (a) where $\frac{d}{n\delta_r} \to 0$. Then (2.44) in Section 2.5.5 becomes

$$\sum_{k \in H_l} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{a.s}(a_l) \wedge \mathrm{O}_{a.s}(\frac{d}{n\delta_l}), \quad j \in H_l, l = 1, \cdots, r, \tag{2.58}$$

which yields

$$\sum_{k \in H_l} \hat{u}_{k,j}^2 = 1 + \mathrm{o}_{a.s}(a_r) \wedge \mathrm{O}_{a.s}(\frac{d}{n\delta_r}), \quad m+1 \leq j \leq n \wedge d. \tag{2.59}$$

From (2.41), (2.58) and (2.59), we have (2.6) in Theorem 2.2.1. Until now, we have finished the proof of scenario (a) in Theorem 2.2.1

Scenario (c) in Theorem 2.2.1

Finally, for the scenario (c) where $\frac{d}{n\delta_1} \to 0$, using (2.46), we have

$$|< \hat{u}_j, u_j >|^2 = \hat{u}_{j,j}^2 = \mathrm{O_{a.s}}\left(\frac{n\lambda_j}{d}\right), \quad j = 1, \cdots, n \wedge d.$$

Thus, (2.10) in Theorem 2.2.1 is established.

Having finished the proof of Theorem 2.2.1, the proofs of Corollaries 2.2.1 and 2.2.3 naturally follow as they are special cases of Theorem 2.2.1.

2.5.6 Proof of Theorem 2.2.2 and Corollaries 2.2.2, 2.2.4

Considering the HDLSS setup, the proofs of Theorem 2.2.2 and Corollaries 2.2.2, 2.2.4 are similar to the above one in Section 2.5.5. One major difference is to replace the following

result with diverging sample size n

$$\frac{1}{n}\sum_{i=1}^{n} z_{i,k}^2 \xrightarrow{a.s} 1, \quad as \quad n \to \infty,$$

with this one that holds for fixed sample size n and diverging dimension d:

$$\frac{1}{n}\sum_{i=1}^{n} z_{i,k}^2 = \mathrm{O_p}(1), \quad as \quad d \to \infty.$$

Such a replacement should be performed in the appropriate places of the proof for Theorem 2.2.1.

2.5.7 Proof of Theorem 2.3.1

Under the assumption of Theorem 2.3.1, Lemma 2.5.3 is still established, where $m = 1$. Lemma 2.5.4 becomes the following lemma:

Lemma 2.5.5. *Under the assumption of Theorem 2.3.1, as $n \to \infty$, the eigenvalues of the matrix B as defined in* (2.21) *satisfy*

$$\frac{n}{d}\lambda_k(B) \xrightarrow{a.s} c_\lambda, \ k = 1, \cdots, n \wedge (d-1), \tag{2.60}$$

Proof. Note that under the assumption of Theorem 2.3.1, we have $\frac{d}{n} \to \infty$. If $\lambda_j \to c_\lambda$, $j = 2, \cdots, d$ and $\frac{d}{n} \to 0$ or ∞, " $\overset{\text{a.s}}{\sim}$ " in Lemma 2.5.4 can be replace by " $\xrightarrow{a.s}$ " and (2.60) is established. □

Asymptotic Properties of Sample Eigenvalues

Now, we will show the asymptotic properties of sample eigenvalues. According to Lemma 2.5.2, we have that

$$\lambda_j(A) + \lambda_n(B) \le \hat{\lambda}_j \le \lambda_j(A) + \lambda_j(B), \tag{2.61}$$

which suggests

$$\frac{1}{\lambda_1}\lambda_1(A) + \frac{1}{\lambda_1}\lambda_n(B) \leq \frac{\hat{\lambda}_1}{\lambda_1} \leq \frac{1}{\lambda_1}\lambda_1(A) + \frac{1}{\lambda_1}\lambda_1(B). \tag{2.62}$$

Note that under the assumption of Theorem 2.3.1, we have $\frac{d}{n} \to \infty$ and then $n \wedge (d-1) = n$. Then Lemma 2.5.5 showed that $\frac{1}{\lambda_1}\lambda_n(B) = \frac{1}{\lambda_1}\lambda_{n\wedge(d-1)}(B)$ or $\frac{1}{\lambda_1}\lambda_1(B) \xrightarrow{a.s} c_\gamma \frac{d}{n\lambda_1} \to cc_\gamma$. In addition, Lemma 2.5.3 showed that $\frac{1}{\lambda_1}\lambda_1(A) \xrightarrow{a.s} 1$. Thus, from (2.62), we have

$$\frac{\hat{\lambda}_1}{\lambda_1} \xrightarrow{a.s} 1 + cc_\lambda, \quad n \to \infty. \tag{2.63}$$

Since rank of A equal to 1, it follows that $\lambda_j(A) = 0$ for $j \geq 2$. Note that Lemma 2.5.5 showed that $\lambda_n(B) = \lambda_{n\wedge(d-1)}(B)$ or $\lambda_1(B) \xrightarrow{a.s} c_\gamma \frac{d}{n}$. Using (2.61), we have

$$\frac{n}{d}\hat{\lambda}_1 \xrightarrow{a.s} c_\lambda, \quad n \to \infty. \tag{2.64}$$

From 2.63 and 2.64, we have (2.17) in Theorem 2.3.1. Until now, we have finished the proof of asymptotic properties of sample eigenvalues in Theorem 2.3.1.

Asymptotic Properties of Sample Eigenvectors

Now, we will show the asymptotic properties of sample eigenvectors. Let $m = 1$ and then $r = 1$ and $q_1 = 1$. Using (2.46), we have

$$|< \hat{u}_j, u_j >|^2 = \hat{u}_{j,j}^2 = \mathrm{O}_{\mathrm{a.s}}\left(\frac{n\lambda_j}{d}\right) = \mathrm{O}_{\mathrm{a.s}}\left(\frac{n}{d}\right), \quad j = 2, \cdots, n \wedge d. \tag{2.65}$$

The first diagonal entry of $S^T S$ ($\hat{\lambda}_1 \sum_{k=1}^{d} \lambda_k^{-1} \hat{u}_{k,1}^2 = \sum_{k=1}^{d} s_{k,1}^2$) is between its smallest than largest eigenvalue. Then (2.47) becomes

$$\lambda_{\min}(\frac{1}{n}Z^T Z) \leq \hat{\lambda}_1 \sum_{k=1}^{d} \lambda_k^{-1} \hat{u}_{1,k}^2 \leq \lambda_{\max}(\frac{1}{n}Z^T Z),$$

which yields

$$\lambda_1^{-1}\lambda_{\min}(\frac{1}{n}Z^TZ) \leq \lambda_1^{-1}\hat{\lambda}_1\sum_{k=1}^{d}\lambda_k^{-1}\hat{u}_{1,k}^2 \leq \lambda_1^{-1}\lambda_{\max}(\frac{1}{n}Z^TZ), \tag{2.66}$$

Note that $\frac{n}{d} \to 0$ and using Lemma 2.5.1, we have $\lambda_1^{-1}\lambda_{\min}(\frac{1}{n}Z^TZ)$ or $\lambda_1^{-1}\lambda_{\max}(\frac{1}{n}Z^TZ) \xrightarrow{a.s} \frac{d}{n\lambda_1} = c + o(1)$. Using (2.17) and assumption of population eigenvalues in Theorem 2.3.1, we have $\lambda_1^{-1}\hat{\lambda}_1\sum_{k=1}^{d}\lambda_k^{-1}\hat{u}_{k,1}^2 \xrightarrow{a.s} \frac{1+cc_\lambda}{c_\lambda}(1-\hat{u}_{1,1}^2)$. Thus, from (2.66), we have

$$|<\hat{u}_1, u_1>|^2 = \hat{u}_{1,1}^2 = \frac{1}{1+cc_\lambda} + o_{\text{a.s}}(1) \tag{2.67}$$

From (2.65)and (2.67), we get (2.18). Until now, we finished the proof of Theorem 2.3.1.

Chapter 3

Consistency of Sparse PCA in HDLSS

3.1 Introduction

Principal Component Analysis (PCA) is an important visualization and dimension reduction tool for High Dimension, Low Sample Size (HDLSS) data. However, the linear combinations found by PCA typically involve all the variables, with non-zero loadings, which can be challenging to interpret. To overcome this weakness, we will study sparse PCA methods that generate sparse principal components (PCs), i.e. PCs with only a few non-zero loadings. Several sparse PCA methods have been proposed to facilitate the interpretation of HDLSS data, see for example Zou et al. (2006); d'Aspremont et al. (2007); Shen and Huang (2008); Leng and Wang (2009); Witten et al. (2009); Johnstone and Lu (2009); Ma (2012); Lee et al. (2012, forthcoming).

Sparse PCA is primarily motivated by modern data sets of very high dimension; hence we prefer the statistical viewpoint of the HDLSS asymptotics. Such asymptotics are based on the limit as the dimension $d \to \infty$ with the sample size n being fixed, as originally studied by Casella and Hwang (1982) in the context of James-Stein estimation, and more recently by Hall et al. (2005); Ahn et al. (2007); Jung and Marron (2009); Yata and Aoshima (2012); Ahn et al. (2012, forthcoming); Jung et al. (2012) in various multivariate analysis contexts. Conventional PCA was first studied using HDLSS asymptotics by Ahn et al. (2007), and much more thoroughly analyzed by Jung and Marron (2009); Jung et al. (2012), with more recent development in Yata and Aoshima (2012). Chapter 3 is the first that studies the HDLSS

asymptotic properties of *Sparse* PCA, and takes the statistical lessons learned from this type of asymptotics in a timely and orthogonal direction from that of Jung and Marron (2009); Jung et al. (2012).

One main contribution of this chapter is a *clear and complete characterization* of HDLSS asymptotic conditions about sparse PCA consistency, inconsistency, as well as strong inconsistency. First of all, we identify in Sections 3.2 and 3.3 sparsity conditions where conventional PCA is strongly inconsistent (for scenarios with relatively small population eigenvalues as carefully studied in Jung and Marron (2009); Jung et al. (2012), and also noted by Lee et al. (2010b)), yet sparse PCA methods are consistent. Furthermore, the mathematical boundaries of the sparse PCA consistency are clearly established in Section 3.4, through showing that an oracle version of sparse PCA is marginally inconsistent on the boundaries, and strongly inconsistent beyond the consistent region. The formulation of *strong inconsistency* has not yet been studied in the random matrix literature (Johnstone and Lu, 2009; Amini and Wainwright, 2009). Following the random matrix work, we also focus on the single component spiked covariance model. Our results depend on a *spike index*, α, which measures the dominance of the first eigenvalue, and on a *sparsity index*, β, which measures the number of non-zero entries of the maximal eigenvector. The two indices α and β are formally defined later in Example 3.1.1 of Section 3.1.1.

Our results offer major new insights relative to those of Jung and Marron (2009); Jung et al. (2012) (when restricting to the maximal eigenvector), who studied HDLSS asymptotic properties of conventional PCA without considering sparsity. This thesis clearly characterizes the benefit of imposing sparsity constraints when the maximal eigenvector is sparse, by revealing a new domain of consistency within the inconsistent region of conventional PCA. In addition, actual rates of convergence are obtained, which are not studied by Jung and Marron (2009); Jung et al. (2012). Our key theoretical findings in connection with the earlier results are intuitively illustrated below for the exemplary model of Example 3.1.1. (Our results remain valid for more general single component spike models as shown in Sections 3.2 to 3.4.) The consistency and inconsistency results for Example 3.1.1 are summarized below as functions of α and β, and illustrated graphically in Figure 3.1:

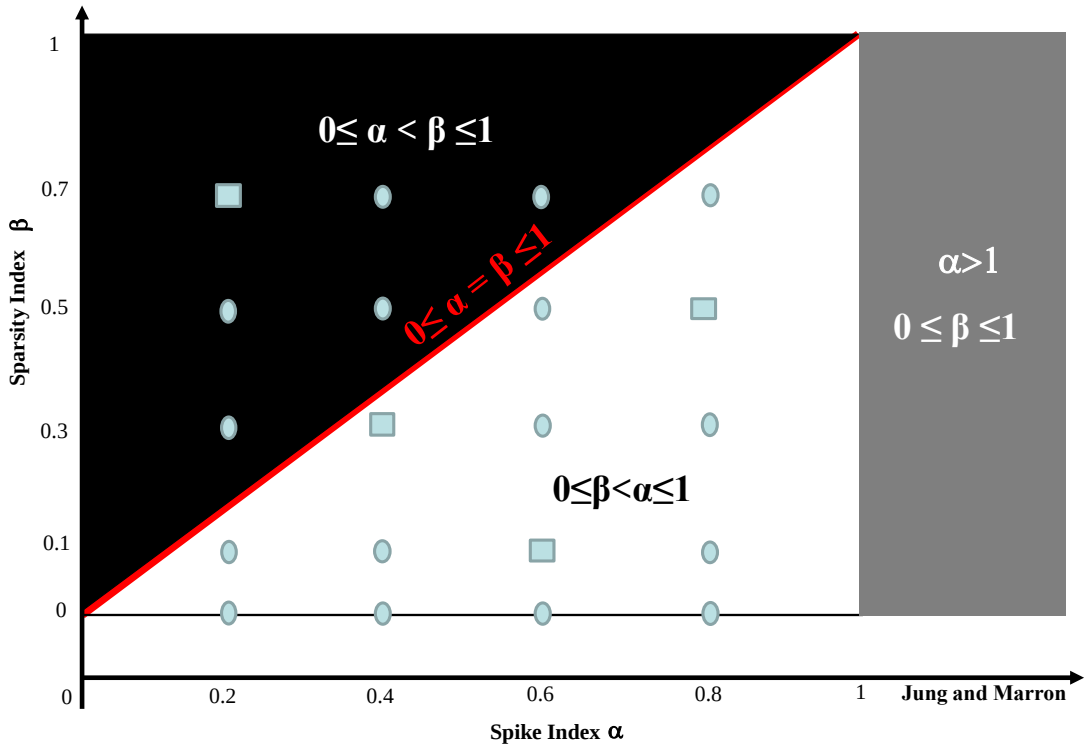

Figure 3.1: Consistent areas for PCA and sparse PCA

- **Previous Results (grey rectangle)**: Jung and Marron (2009) showed that the first empirical PCA eigenvector is consistent with the maximal eigenvector when the spike index α is greater than 1, only involving the horizontal axis.

- **Consistency (white triangle)**: We will show that sparse PCA is consistent even when α is less than or equal to 1 (hence PCA is either strongly inconsistent (Jung and Marron, 2009) or marginally inconsistent (Jung et al., 2012)), as long as α is greater than the sparsity index β, involving both horizontal and vertical axes. This is done in Section 3.2 for a simple thresholding method and in Section 3.3 for the RSPCA method of Shen and Huang (2008).

- **Strong Inconsistency (black triangle) and Marginal Inconsistency (red diagonal line)**: In Section 3.4 we show that even an oracle sparse PCA procedure is strongly inconsistent, when α is smaller than β, and marginally inconsistent when $\alpha = \beta$, the boundary between the consistent and strongly inconsistent regions.

Besides the HDLSS asymptotics, other asymptotic frameworks have been used to study sparse PCA. Under the random matrix framework where both n and d tend to infinity, Johnstone and Lu (2009); Amini and Wainwright (2009) considered the single spike covariance model (originally proposed by Johnstone (2001)). In particular, Johnstone and Lu (2009) showed that conventional PCA is consistent if and only if $d(n)/n \to 0$; furthermore, when $\log(d \vee n)/n \to 0$, they proved that PCA could regain consistency when being performed on a subset of variables with the largest sample variances. Note that our asymptotic framework of $d \to \infty$ with n fixed is not considered by Johnstone and Lu (2009), who also modeled the sparsity differently from us. Amini and Wainwright (2009) further restricted the maximal eigenvector to have k non-zero entries, and studied support set recovery properties of the thresholding subset PCA procedure of Johnstone and Lu (2009) and the DSPCA procedure of d'Aspremont et al. (2007). Different from Amini and Wainwright (2009), the thesis studies asymptotic properties of estimating the actual maximal eigenvector, instead of its support set. Hence, our results are complementary to those of Amini and Wainwright (2009), as elaborated below in the context of Example 3.1.2. Paul and Johnstone (2007b) developed the augmented sparse PCA procedure along with its optimal rate of convergence property. Ma (2012) proposed an iterative thresholding procedure for estimating principal subspaces, and established its nice theoretical properties. Considering the sample size $n \to \infty$ with dimension d fixed, Leng and Wang (2009) proposed an adaptive lasso sparse PCA, and established its consistency for selecting non-zero loadings.

Asymptotics on sparsity have been investigated in other high dimensional settings, such as regression and variable selection by Meinshausen and Bühlmann (2006); Candes and Tao (2007); van de Geer (2008); Bickel et al. (2009); Meier et al. (2009); Omidiran and Wainwright (2010); Obozinski et al. (2011), sparse covariance matrix estimation by Bickel and Levina (2008a,b); El Karoui (2008), classification by Bühlmann (2006), and density estimation by Bunea et al. (2010), for example.

3.1.1 Notation and Assumptions

All quantities are indexed by the dimension d in the current chapter. However, when it will not lead to confusion, the subscript d will be omitted for convenience. Recall above, let the population covariance matrix be Σ. The eigen-decomposition of Σ is

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

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

Recall Assumption 2.1.1 that $X_1, \ldots, X_n$ are random samples from a d-dimensional normal distribution $N(0, \Sigma_d)$. Denote the data matrix by $X = [X_1, \ldots, X_n]_{d\times n}$ and the sample covariance matrix by $\hat{\Sigma} = n^{-1}XX^T$. Then, the sample covariance matrix $\hat{\Sigma}$ can be similarly decomposed as

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

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

Let $\bar{u}_j$ be any sample based estimator of u_j, e.g. $\bar{u}_j = \hat{u}_j$ for $j = 1, \ldots, d$. Recall three consistency concepts from Section 2.2.1:

- **Consistency**: The direction $\bar{u}_j$ is *consistent* with its population counterpart u_j if

$$\text{Angle}(\bar{u}_j, u_j) \equiv \arccos(|< \bar{u}_j, u_j >|) \xrightarrow{p} 0, \text{as } d \to \infty, \tag{3.1}$$

  where $< \cdot, \cdot >$ denotes the inner product between two vectors.

- **Strong Inconsistency**: The direction $\bar{u}_j$ is *strongly inconsistent* with its population counterpart u_j if $\text{Angle}(\bar{u}_j, u_j) \xrightarrow{p} \frac{\pi}{2}$, as $d \to \infty$.

- **Consistency with convergence rate d^ι**: The direction $\bar{u}_j$ is consistent with its population counterpart u_j with the convergence rate d^ι if $|< \bar{u}_j, u_j >| = 1 + o_p(d^{-\iota})$, where the notation $G_d \equiv o_p(d^{-\iota})$ means that $d^\iota G_d \xrightarrow{p} 0$, as $d \to \infty$.

In addition, we consider another important concept in the current chapter:

- **Marginal Inconsistency**: The direction $\bar{u}_j$ is *marginally inconsistent* with u_j if $\text{Angle}(\bar{u}_j, u_j)$ converges to a (possibly random) quantity in $(0,\ \frac{\pi}{2})$, as $d \to \infty$.

We now use two illustrative examples to highlight our key theoretical results. The examples are chosen mainly for intuitive illustration. Our theorems cover more general single component spike models (Sections 3.2 to 3.4).

Example 3.1.1. *Assume that $X_1, \ldots, X_n$ are random sample vectors from a d-dimensional normal distribution $N(0, \Sigma_d)$, where the covariance matrix Σ_d has the eigenvalues as*

$$\lambda_1 = d^\alpha, \lambda_2 = \ldots = \lambda_d = 1, \alpha \geq 0. \tag{3.2}$$

This is a special case of the single component spike covariance Gaussian model considered before by, for example, Johnstone (2001); Paul (2007); Johnstone and Lu (2009); Amini and Wainwright (2009).

Without loss of generality (WLOG), we further assume that the first eigenvector u_1 is proportional to the d-dimensional vector

$$\dot{u}_1 = (\overbrace{1, \ldots, 1}^{\lfloor d^\beta \rfloor}, 0, \ldots, 0)^T,$$

where $0 \leq \beta \leq 1$ and $\lfloor d^\beta \rfloor$ denotes the integer part of d^β. For example, if $\beta = 0$, the first population eigenvector becomes $u_1 = (1, 0, \ldots, 0)^T$. (Note that in general the non-zero entries do not have to be the first $\lfloor d^\beta \rfloor$ elements, nor do they need to have equal values.)

We formally define α as the spike index *that measures the strength of the spike, and β as the* sparsity index *that quantifies the sparsity of the maximal eigenvector u_1, where $\lfloor d^\beta \rfloor$ is the number of its non-zero elements. Under Model* (3.2)*, Jung and Marron (2009) showed that the first empirical eigenvector (the PC direction) $\hat{u}_1$ is consistent with u_1 when $\alpha > 1$; however for $\alpha < 1$, it is strongly inconsistent. Jung et al. (2012) then showed that $\hat{u}_1$ is in between consistent and strongly inconsistent on the boundary when $\alpha = 1$. Again, one main*

point of the thesis is an exploration of conditions under which sparse methods can lead to consistency even when the spike index $\alpha \leq 1$, by exploiting sparsity.

Example 3.1.2. *Our theorems are also applicable to the sparse single-component spike model considered by Amini and Wainwright (2009), although we have a different focus from them as illustrated below. The covariance matrix Σ_d can be expressed using our notation as*

$$\Sigma_d = (\lambda_1 - 1) z^* z^{*T} + \begin{pmatrix} I_{\lfloor d^\beta \rfloor} & 0 \\ 0 & \Gamma_{d-\lfloor d^\beta \rfloor} \end{pmatrix},$$

where the first eigenvalue $\lambda_1 > 1$, the first $\lfloor d^\beta \rfloor$ entries of the maximal eigenvector are non-zero with values of $\pm 1/\sqrt{\lfloor d^\beta \rfloor}$, and $\Gamma_{d-\lfloor d^\beta \rfloor}$ is a symmetric positive semi-definite matrix with the maximal eigenvalue $\lambda_{max}(\Gamma_{d-\lfloor d^\beta \rfloor}) \leq 1$. For this example, consider cases where all eigenvalues of $\Gamma_{d-\lfloor d^\beta \rfloor}$ are one. Hence, the eigenvalues of Σ_d are $\lambda_1 > 1 = \lambda_2 = \cdots = \lambda_d$.

Amini and Wainwright (2009) focused on the consistent recovery of the support set of the maximal eigenvector z^*. We, however, are interested in the consistency of the actual direction vector. We note that the two types of consistency are not equivalent. For the above model, the two results are summarized below.

- *Assuming fixed λ_1 and $n, d \to \infty$, Amini and Wainwright (2009) showed that the support set can be recovered if $n > c_u(\lfloor d^\beta \rfloor)^2 \log(d - \lfloor d^\beta \rfloor)$, while it can not be recovered if $n < c_l(\lfloor d^\beta \rfloor)^2 \log(d - \lfloor d^\beta \rfloor)$, where c_u and c_l are two constants.*

- *For fixed n, $\lambda_1 = d^\alpha$ and $d \to \infty$, our Theorems 3.2.2 and 3.3.4 show that sparse PCA is consistent when $\alpha > \beta$, although the support set may not be consistently recovered (Theorem 3 of AW); and our Theorem 3.4.1 indicates that sparse PCA is strongly inconsistent when $\alpha < \beta$, even when one knows the exact support set.*

3.1.2 Roadmap

The organization of the rest of this chapter is as follows. For easy access to the main ideas, Section 3.2 proves the consistency of a simple thresholding (ST) method that generates sparse

PC directions, under the sparsity and small spike conditions where the conventional PCA is strongly inconsistent. Section 3.3 then generalizes these ideas to establish the consistency of the RSPCA method of Shen and Huang (2008). Section 4.5 identifies the region and its boundaries where the strong inconsistency and marginal inconsistency of an appropriate oracle sparse PCA procedure are proved. Section 3.5 reports simulation results to illustrate both consistency and strong inconsistency of PCA and sparse PCA. Section 3.6 concludes Chapter 3 with some discussion of future work. Section 3.8 contains the proofs of the main theorems.

3.2 Consistency of Simple Thresholding Sparse PCA

In Example 3.1.1, the first eigenvector of the sample covariance matrix $\hat{u}_1$ is strongly inconsistent with u_1 when $\alpha < 1$, because it attempts to estimate too many parameters. Sparse data analytic methods assume that many of these parameters are zero, which can allow greatly improved estimation of the first PC direction u_1. Here, this issue is explored in the context of sparse PCA. The sample covariance matrix based estimator, $\hat{u}_1$, can be improved by exploiting the fact that u_1 has many zero elements.

We first study a natural simple thresholding (ST) method where entries with small absolute values are replaced by zero. (Starting with the ST approach makes it easier to demonstrate the key ideas that are also useful for establishing the consistency of a more sophisticated sparse PCA method in Section 3.3.) In HDLSS contexts, it is challenging to apply thresholding directly to $\hat{u}_1$, because the number of its entries grows rapidly as $d \to \infty$, which naturally shrinks their magnitudes given that $\hat{u}_1$ has norm one. Thresholding is more conveniently formulated in terms of the *dual covariance matrix* (Jung et al., 2012).

Denote the dual sample covariance matrix by $S = \frac{1}{n} X^T X$ and the first dual eigenvector by $\tilde{v}_1$. The sample eigenvector $\hat{u}_1$ is connected with the dual eigenvector $\tilde{v}_1$ through the following transformation,

$$\tilde{u}_1 = (\tilde{u}_{1,1}, \ldots, \tilde{u}_{d,1})^T = X\tilde{v}_1, \tag{3.3}$$

and the sample estimate is then given by $\hat{u}_1 = \tilde{u}_1 / \|\tilde{u}_1\|$ (Jung et al., 2012).

Given a sequence of threshold values ζ, define the thresholded entries as

$$\breve{u}_{k,1} = \begin{cases} \tilde{u}_{k,1} & \text{if } |\tilde{u}_{k,1}| > \zeta, \\ 0 & \text{if } |\tilde{u}_{k,1}| \leq \zeta, \end{cases} \quad \text{for} \quad k = 1, \ldots, d. \tag{3.4}$$

Denote $\breve{u}_1 = (\breve{u}_{1,1}, \ldots, \breve{u}_{d,1})^T$ and normalize it to get the simple thresholding (ST) estimator $\hat{u}_1^{\mathrm{ST}} = \breve{u}_1 / \|\breve{u}_1\|$.

For the model in Example 3.1.1, given an eigenvalue of strength $\alpha \in (0, 1)$, (recall that $\lambda_1 = d^\alpha$ and $\hat{u}_1$ is strongly inconsistent), below we explore conditions on the threshold sequence ζ under which the ST estimator $\hat{u}_1^{\mathrm{ST}}$ is in fact consistent with u_1. First of all, the threshold ζ can not be too large; otherwise all the entries will be zeroed out. It will be seen in Theorem 3.2.1 that a sufficient condition for this is $\zeta \leq d^{\frac{\gamma}{2}}$, where $\gamma \in (0, \alpha)$. Secondly, the threshold ζ can not be too small, or pure noise terms will be included. A parallel sufficient condition is shown to be $\zeta \geq \log^\delta(d)\lambda_2^{\frac{1}{2}}$, where $\delta \in (\frac{1}{2}, \infty)$.

Below we formally establish conditions on the eigenvalues of the population covariance matrix Σ_d and the thresholding parameter ζ, which give consistency of $\hat{u}_1^{\mathrm{ST}}$ to u_1. The proofs are provided in Section 3.8.

To fix ideas, we first consider the extreme sparsity case $u_1 = (1, 0, \ldots, 0)^T$. Suppose that $\lambda_1 \sim d^\alpha$, in the sense that $0 < c_1 \leq \underline{\lim}_{d\to\infty} \frac{\lambda_1}{d^\alpha} \leq \overline{\lim}_{d\to\infty} \frac{\lambda_1}{d^\alpha} \leq c_2$, for two constants c_1 and c_2. WLOG, assume $\sum_{j=2}^d \lambda_j \sim d$. As in Jung and Marron (2009), denote the measure of sphericity for $\{\lambda_2, \cdots, \lambda_d\}$ as

$$\varepsilon_2 \equiv \frac{(\sum_{j=2}^d \lambda_j)^2}{d \sum_{j=2}^d \lambda_j^2},$$

which can be used as the basis of a hypothesis test for equality of eigenvalues, and assume the ε_2-condition: $\varepsilon_2 \gg \frac{1}{d}$, i.e

$$(d\varepsilon_2)^{-1} = \frac{\sum_{j=2}^d \lambda_j^2}{(\sum_{j=2}^d \lambda_j)^2} \to 0, \text{as } d \to \infty. \tag{3.5}$$

Now we need to impose the following conditions on the eigenvalues:

- Assume that the ε_2-condition (3.5) is satisfied, which guarantees that the dual matrix S_d has a limit. Hence the first dual eigenvector $\hat{v}_1$ will have a limit and it will then help build up the consistency of $\hat{u}_1^{\mathrm{ST}}$.

- In addition, we need the second eigenvalue λ_2 to be an obvious distance away from the first eigenvalue λ_1. If not, it will be hard to distinguish the first and second empirical eigenvectors as observed by Jung and Marron (2009), among others. In that case the appropriate amount of thresholding on the first empirical eigenvector becomes unclear. Therefore, we assume that $\lambda_2 \sim d^{\theta}$, where $\theta < \alpha$.

Theorem 3.2.1. *Suppose that $X_1, \ldots, X_n$ are random samples from a d-dimensional normal distribution $N(0, \Sigma_d)$ and the first population eigenvector $u_1 = (1, 0, \ldots, 0)^T$. If the following conditions are satisfied:*

(a) $\lambda_1 \sim d^{\alpha}$, $\lambda_2 \sim d^{\theta}$, *and* $\sum_{j=2}^{d} \lambda_j \sim d$, *where* $\theta \in [0, \alpha)$ *and* $\alpha \in (0, 1]$,

(b) the ε_2-condition (3.5) *is satisfied,*

(c) $\log^{\delta}(d)\, d^{\frac{\theta}{2}} \leq \zeta \leq d^{\frac{\gamma}{2}}$, *where* $\delta \in (\frac{1}{2}, \infty)$ *and* $\gamma \in (\theta, \alpha)$,

then the simple thresholding estimator $\hat{u}_1^{\mathrm{ST}}$ is consistent with u_1.

In fact, $u_1 = (1, 0, \ldots, 0)^T$ in Theorem 3.2.1 is a very extreme case. The following theorem considers the general case $u_1 = (u_{1,1}, \ldots, u_{d,1})^T$, where only $\lfloor d^{\beta} \rfloor$ elements of u_1 are non-zero. WLOG, we assume that the first $\lfloor d^{\beta} \rfloor$ entries are non-zero just for notational convenience.

Define

$$Z_i \equiv (z_{1,i}, \ldots, z_{d,i})^T = (X_i^T u_1, \ldots, X_i^T u_d)^T, \quad i = 1, \ldots, n. \tag{3.6}$$

We can show that Z_i are iid $N\left(0, \mathrm{diag}\{\lambda_1, \ldots, \lambda_d\}\right)$ random vectors. Let

$$W_i \equiv (w_{1,i}, \ldots, w_{d,i})^T = (\lambda_1^{-\frac{1}{2}} z_{1,i}, \ldots, \lambda_d^{-\frac{1}{2}} z_{d,i})^T, \quad i = 1, \ldots, n, \tag{3.7}$$

and the W_i are iid $N(0, I_d)$ random vectors, where I_d is the d-dimensional identity matrix.

The following conditions are also needed to ensure the consistency of $\hat{u}_1^{\mathrm{ST}}$:

- The non-zero entries of the population eigenvector u_1 need to be a certain distance away from zero. In fact, if the non-zero entries of the first population eigenvector are close to zero, the corresponding entries of the first empirical eigenvector would also be small and look like pure noise entries. Thus, we assume

$$\max_{1\le k\le \lfloor d^{\beta}\rfloor}|u_{k,1}|^{-1} \sim d^{\frac{\eta}{2}}, \quad \text{where} \quad \eta \in [0,\alpha).$$

- From (3.6), we have

$$X_i = \sum_{j=1}^{d} z_{j,i}u_j, \quad i = 1,\dots,n.$$

Since $z_{1,i}$ has the largest variance λ_1, then $z_{1,i}u_1$ contributes the most to the variance of X_i, $i = 1,\dots,n$. Note that $z_{1,i}u_1$ is consistent with u_1, and so $z_{1,i}u_1$ is the key to making the simple thresholding method work. So we need to show that the remaining parts

$$H_i \equiv (h_{1,i},\dots,h_{d,i})^T = \sum_{j=2}^{d} z_{j,i}u_j, \quad i = 1,\dots,n \tag{3.8}$$

have a negligible effect on the direction vector $\hat{u}_1^{\mathrm{ST}}$.

- Suppose that the H_i are iid $N(0,\Delta_d)$, where $\Delta_d = (m_{kl})_{d\times d}$. A sufficient condition to make their effect negligible is the following mixing condition of Leadbetter et al. (1983):

$$|m_{kl}| \le {m_{kk}}^{\frac{1}{2}}{m_{ll}}^{\frac{1}{2}}\rho_{|k-l|}, \quad 1 \le k \ne l \le \lfloor d^{\beta}\rfloor, \tag{3.9}$$

where $\rho_t < 1$ for all $t > 1$ and $\rho_t \log(t) \longrightarrow 0$, as $t \to \infty$. This mixing condition can guarantee that $\max_{1\le i\le n}|h_{1,i}|$ has a quick convergence rate, as $d \to \infty$. It enables us to neglect the influence of H_i for sufficiently large d and make $z_{j,i}u_1$ the dominant component, which then gives consistency to the first population eigenvector u_1. Thus the thresholding estimator $\hat{u}_1^{\mathrm{ST}}$ becomes consistent.

We now state one of the main theorems:

Theorem 3.2.2. *Assume that $X_1, \ldots, X_n$ are random samples from a d-dimensional normal distribution $N(0, \Sigma_d)$. Define Z_i, W_i and H_i as in (3.6), (3.7), and (3.8) for $i = 1, \ldots, n$. The first population eigenvector is $u_1 = (u_{1,1}, \ldots, u_{d,1})^T$ with $u_{k,1} \neq 0, k = 1, \ldots, \lfloor d^\beta \rfloor$, and otherwise $u_{k,1} = 0$.*

If the following conditions are satisfied:

(a) $\lambda_1 \sim d^\alpha$, $\lambda_2 \sim d^\theta$, and $\sum_{j=2}^d \lambda_j \sim d$, where $\theta \in [0, \alpha)$ and $\alpha \in (0, 1]$,

(b) the ε_2-condition (3.5) is satisfied,

(c) $max_{1 \leq k \leq \lfloor d^\beta \rfloor} |u_{k,1}|^{-1} \sim d^{\frac{\eta}{2}}$, where $\eta \in [0, \alpha)$,

(d) H_i satisfies the mixing condition (3.9), $i = 1, \ldots, n$,

(e) $\log^\delta(d)\, d^{\frac{\theta}{2}} \leq \zeta \leq d^{\frac{\gamma}{2}}$, where $\delta \in (\frac{1}{2}, \infty)$ and $\gamma \in (\theta, \alpha - \eta)$,

then the thresholding estimator $\hat{u}_1^{\mathrm{ST}}$ is consistent with u_1.

We offer a couple of remarks regarding Theorem 3.2.2. First of all, the theorem naturally reduces to Theorem 3.2.1 if we let the sparsity index $\beta = 0$. More importantly, this theorem, and the following ones in Sections 3.2 to 3.4, show that the concepts depicted in Figure 3.1 hold much more generally than just the models in Examples 3.1.1 and 3.1.2. In particular, in the above theorem, setting $\theta = 0$ and $\eta = \beta$ would give the results plotted in Figure 3.1.

In addition, for different thresholding parameter ζ, the ST estimator $\hat{u}_1^{\mathrm{ST}}$ is consistent with different convergence rate, as stated in the following theorem. The notation $\zeta = o(d^\rho)$ below means that $\zeta d^{-\rho} \to 0$ as $d \to \infty$.

Theorem 3.2.3. *For the thresholding parameter $\zeta = o(d^{\frac{\alpha - \eta - \kappa}{2}})$, where $\kappa \in [0, \alpha - \eta - \theta)$, the corresponding thresholding estimator $\hat{u}_1^{\mathrm{ST}}$ is consistent with u_1, with a convergence rate of $d^{\frac{\kappa}{2}}$.*

3.3 Consistency of RSPCA

As noted in Section 3.1, several sparse PCA methods have been proposed in the literature. Here we perform a detailed HDLSS asymptotic analysis of the sparse PCA procedure developed by Shen and Huang (2008). For simplicity, we refer to it as the *regularized sparse PCA*, or RSPCA for short. All the detailed proofs are again provided in Section 3.8.

We start with briefly reviewing the methodological details of RSPCA. (For more details, see Shen and Huang (2008).) Given a d-by-n data matrix X, consider the following penalized sum-of-squares criterion:

$$\|X - uv^T\|_F^2 + P_\zeta(u), \quad \text{subject to} \quad \|v\| = 1, \tag{3.10}$$

where u is a d-vector, v is a unit n-vector, $\|\cdot\|_F$ denotes the Frobenius norm, and $P_\zeta(u) = \sum_{j=1}^{d} p_\zeta(|u_{j,1}|)$ is a penalty function with $\zeta \geq 0$ being the penalty parameter. The penalty function can be any sparsity-inducing penalty. In particular, Shen and Huang (2008) considered the soft thresholding (or L_1 or LASSO) penalty of Tibshirani (1996), the hard thresholding penalty of Donoho and Johnstone (1994), and the smoothly clipped absolute deviation (SCAD) penalty of Fan and Li (2001).

Without the penalty term or when $\zeta = 0$, minimization of (3.10) can be obtained via singular value decomposition (SVD) (Eckart and Young, 1936), which results in the best rank-one approximation of X as $\tilde{u}_1\tilde{v}_1^T$, where $\tilde{u}_1$ and $\tilde{v}_1^T$ minimize the criterion (3.10). The normalized $\tilde{u}_1$ turns out to be the first empirical PC loading vector. With the penalty term, Shen and Huang (2008) define the sparse PC loading vector as $\hat{u}_1 = \tilde{u}_1/\|\tilde{u}_1\|$ where $\tilde{u}_1$ is now the minimizer of (3.10) with the penalty term included. The minimization now needs to be performed iteratively. For a given $\tilde{v}_1$ in the criterion (3.10), we can get the minimizing vector as $\tilde{u}_1 = h_\zeta(X\tilde{v}_1)$, where h_ζ is a thresholding function that depends on the particular penalty function used and the penalty (or thresholding) parameter ζ. See Shen and Huang (2008) for more details. The thresholding is applied to the vector $X\tilde{v}_1$ componentwise.

Shen and Huang (2008) proposed the following iterative procedure for minimizing the criterion (3.10):

The RSPCA Algorithm

---

1. Initialize:

   (a) Use SVD to obtain the best rank-one approximation $\tilde{u}_1\tilde{v}_1^T$ of the data matrix X, where $\tilde{v}_1$ is a unit vector.

   (b) Set $\tilde{u}_1^{\text{old}} = \tilde{u}_1$ and $\tilde{v}_1^{\text{old}} = \tilde{v}_1$.

2. Update:

   (a) $\tilde{u}_1^{\text{new}} = h_\zeta\left(X\tilde{v}_1^{\text{old}}\right)$.

   (b) $\tilde{v}_1^{\text{new}} = \frac{X^T\tilde{u}_1^{\text{new}}}{\|X^T\tilde{u}_1^{\text{new}}\|}$.

3. Repeat Step 2 setting $\tilde{u}_1^{\text{old}} = \tilde{u}_1^{\text{new}}$ and $\tilde{v}_1^{\text{old}} = \tilde{v}_1^{\text{new}}$ until convergence.

4. Normalize the final $\tilde{u}_1^{\text{new}}$ to get $\hat{u}_1$, the desired sparse loading vector.

There exists a nice connection between the ST method of Section 3.2 and RSPCA with hard thresholding. The ST estimator $\hat{u}_1^{\text{ST}}$ is exactly the sparse loading vector $\hat{u}_1$ obtained from the first iteration of the RSPCA iterative algorithm, when the hard thresholding penalty is used. Such connection suggests that we can extend the proofs for the theorems regarding the property of the ST estimator to establish consistency theorems for RSPCA below. Although nicely connected, RSPCA actually performs better numerically over ST as we will illustrate in Section 3.5.

Below we develop conditions under which the RSPCA estimator $\hat{u}_1$ is consistent with the population eigenvector u_1 when a proper thresholding parameter ζ is used. All three of the soft thresholding, hard thresholding or SCAD penalties are considered. To prove the consistency of RSPCA, we first establish in Theorem 3.3.1 the consistency of the first-step RSPCA estimator obtained during the initial iteration, which can then be used to show in Theorem 3.3.4 that the follow-up updated RSPCA estimator remains consistent. Convergence rates are given in Theorems 3.3.2 and 3.3.5.

The following Theorem 3.3.1 states conditions when the first-step RSPCA estimator $\hat{u}_1$ is consistent with u_1 under a proper thresholding parameter ζ. Given the aforementioned connection between ST and RSPCA, the consistency conditions are the same as those needed for the consistency of ST. See the discussion of Theorem 3.2.2 for the implications of these conditions.

Theorem 3.3.1. *Under the assumptions and conditions of Theorem 3.2.2, the first-step sparse loading vector $\hat{u}_1$ is consistent with u_1.*

Similar to the ST estimator (Theorem 3.2.3), for different parameters ζ, the RSPCA estimator $\hat{u}_1$ is consistent with u_1 with different convergence rates. The result is given in the following Theorem 3.3.2.

Theorem 3.3.2. *For the thresholding parameter $\zeta = o(d^{\frac{\alpha-\eta-\kappa}{2}})$, where $\kappa \in [0, \alpha-\eta-\theta)$, the sparse loading vector $\hat{u}_1$ in Theorem 3.3.1 is consistent with u_1, with a convergence rate of $d^{\frac{\kappa}{2}}$.*

To obtain an updated estimate for v_1 we set $\hat{u}_1^{\text{old}}$ to be the consistent first-step RSPCA estimate, and obtain $\tilde{v}_1^{\text{new}} = X^T\hat{u}_1^{\text{old}}/\|X^T\hat{u}_1^{\text{old}}\|$. Theorem 3.3.3 below studies the asymptotic properties of $\tilde{v}_1^{\text{new}}$.

Theorem 3.3.3. *Assume that $\hat{u}_1^{\text{old}}$ is consistent with u_1 with the convergence rate $d^{\frac{\kappa}{2}}$, where $\kappa \in [1-\alpha, \infty)$. If the ε_2-condition is satisfied, then*

$$\tilde{v}_1^{\text{new}} \xrightarrow{p} \frac{\tilde{W}_1}{\|\tilde{W}_1\|}, \quad as \quad d \to \infty,$$

where $\tilde{W}_1 = (w_{1,1}, \cdots, w_{1,n})$ follows a standard n-dimensional normal distribution $N(0, I_n)$ and the $w_{j,i}$ are defined in (3.7).

Given the above established asymptotic properties of $\tilde{v}_1^{\text{new}}$, we can now study the asymptotic properties of the updated RSPCA estimator

$$\hat{u}_1^{\text{new}} = \frac{\tilde{u}_1^{\text{new}}}{\|\tilde{u}_1^{\text{new}}\|}, \quad \text{with} \quad \tilde{u}_1^{\text{new}} = h_\zeta(X\tilde{v}_1^{\text{new}}), \tag{3.11}$$

as defined in the iterative RSPCA algorithm. The following Theorem 3.3.4 shows that, with a proper choice of ζ, the updated RSPCA estimator $\hat{u}_1^{\text{new}}$ remains to be consistent with the population eigenvector u_1.

Theorem 3.3.4. *Under the assumptions and conditions of Theorems 3.2.2 and 3.3.3, the updated loading vector $\hat{u}_1^{\text{new}}$ (3.11) is consistent with u_1.*

The following Theorem 3.3.5 studies the convergence rate property of $\hat{u}_1^{\text{new}}$. For different ζ, $\hat{u}_1^{\text{new}}$ is consistent with u_1 with different convergence rates.

Theorem 3.3.5. *For the thresholding parameter $\zeta = o(d^{\frac{\alpha-\eta-\kappa}{2}})$, where $\kappa \in [0, \alpha-\eta-\theta)$, the updated sparse loading vector $\hat{u}_1^{\text{new}}$ in Theorem 3.3.4 is consistent with u_1, with convergence rate $d^{\frac{\kappa}{2}}$.*

Theorems 3.3.2 and 3.3.5 imply that, if $\alpha - \eta - \theta > 1 - \alpha$, then by choosing the thresholding parameter to be $\zeta = o(d^{\frac{\alpha-\eta-\kappa}{2}})$, we can make the updated RSPCA loading vector $\hat{u}_1^{\text{new}}$consistent with u_1 at every updating step.

Interestingly, Theorems 3.2.3 and 3.3.5 suggest that the ST estimator and the RSPCA estimator share the same rate of convergence. However, we note that the RSPCA estimator has better finite sample performance than ST: RSPCA is more stable due to the multiple iterations involved in the estimation, which reduce the estimation variability. The improvement of RSPCA over ST is illustrated in Figure 3.4.

3.4 Strong Inconsistency and Marginal Inconsistency

We have shown that we can attain consistency using sparse PCA, when the spike index α is greater than the sparsity index β. This motivates the question of consistency using sparse PCA when α is smaller than or equal to β. To answer this question, we consider an oracle estimator which "knows" the exact positions of the zero entries of the maximal eigenvector u_1. We will show that even this oracle estimator is strongly inconsistent when α is smaller than β, and marginally inconsistent when $\alpha = \beta$. Compared with this oracle sparse PCA,

threshold methods can perform no better because they also need to estimate location of the zero entries; hence threshold methods will also be inconsistent.

To make this precise, we study the procedure to generate the oracle estimator for general single component models. Similar to Sections 3.2 and 3.3, assume that the first $\lfloor d^\beta \rfloor$ entries of the maximal eigenvector u_1 are non-zero and the rest are all zero: $u_1 = (u_{1,1}, \ldots, u_{d,1})^T$, where $u_{k,1} \neq 0, k = 1, \ldots, \lfloor d^\beta \rfloor$; otherwise $u_{k,1} = 0$.

Let $X_i^* = (x_{1,i}, \ldots, x_{\lfloor d^\beta \rfloor, i})^T \sim N(0, \Sigma^*_{\lfloor d^\beta \rfloor})$, where $\Sigma^*_{\lfloor d^\beta \rfloor}$ is the covariance matrix of X_i^*, $i = 1, \ldots, n$. Then, the eigen-decomposition of $\Sigma^*_{\lfloor d^\beta \rfloor}$ is

$$\Sigma^*_{\lfloor d^\beta \rfloor} = U^*_{\lfloor d^\beta \rfloor} \Lambda^*_{\lfloor d^\beta \rfloor} (U^*_{\lfloor d^\beta \rfloor})^T,$$

where Λ_d^* is the diagonal matrix of eigenvalues $\lambda_1^* \geq \lambda_2^* \geq \ldots \geq \lambda^*_{\lfloor d^\beta \rfloor}$, and $U^*_{\lfloor d^\beta \rfloor}$ is the matrix of the corresponding eigenvectors so that $U^*_{\lfloor d^\beta \rfloor} = [u_1^*, \ldots, u^*_{\lfloor d^\beta \rfloor}]$.

Since the last $d - \lfloor d^\beta \rfloor$ entries of the maximal eigenvector u_1 equal zero, it follows that the first eigenvector u_1^* of $\Sigma^*_{\lfloor d^\beta \rfloor}$ is formed by the non-zero entries of u_1, i.e. $u_1^* = (u_{1,1}, \ldots, u_{\lfloor d^\beta \rfloor, 1})^T$. So we have

$$u_1 = ((u_1^*)^T, \overbrace{0, \ldots, 0}^{d - \lfloor d^\beta \rfloor})^T. \tag{3.12}$$

Consider the following data matrix $X^*_{\lfloor d^\beta \rfloor} = [X_1^*, \ldots, X_n^*]$, and denote the sample covariance matrix by $\hat{\Sigma}^*_{\lfloor d^\beta \rfloor} = n^{-1} X^*_{\lfloor d^\beta \rfloor} X^T_{\lfloor d^\beta \rfloor}$. Then, the sample covariance matrix $\hat{\Sigma}^*_{\lfloor d^\beta \rfloor}$ can be similarly decomposed as

$$\hat{\Sigma}^*_{\lfloor d^\beta \rfloor} = \hat{U}^*_{\lfloor d^\beta \rfloor} \hat{\Lambda}^*_{\lfloor d^\beta \rfloor} (\hat{U}^*_{\lfloor d^\beta \rfloor})^T,$$

where $\hat{\Lambda}^*_{\lfloor d^\beta \rfloor}$ is the diagonal matrix of the sample eigenvalues $\hat{\lambda}_1^* \geq \hat{\lambda}_2^* \geq \ldots \geq \hat{\lambda}^*_{\lfloor d^\beta \rfloor}$, and $\hat{U}^*_{\lfloor d^\beta \rfloor}$ is the matrix of the corresponding sample eigenvectors so that $\hat{U}^*_{\lfloor d^\beta \rfloor} = [\hat{u}_1^*, \ldots, \hat{u}_d^*]$.

Then, we define the oracle (OR) estimator as

$$\hat{u}_1^{\mathrm{OR}} = ((\hat{u}_1^*)^T, \overbrace{0, \ldots, 0}^{d-\lfloor d^\beta \rfloor})^T. \tag{3.13}$$

The following theorem formally states its inconsistency properties.

Theorem 3.4.1. *Assume that $X_1, \ldots, X_n$ are random samples from a d-dimensional normal distribution $N(0, \Sigma_d)$.*

(a) If $\lambda_1 \sim d^\alpha$, $\lambda_2 \sim d^\theta$, $\lambda_d \sim 1$ and $\sum_{j=2}^d \lambda_j \sim d$, where $\alpha < \beta$ and $\theta \in [0, \frac{\beta}{2})$, then the oracle estimator $\hat{u}_1^{\mathrm{OR}}$ in (3.13) is strongly inconsistent with u_1.

(b) If $\lambda_1/d^\alpha \to c \in (0, \infty)$ and $\lambda_j \to c_\lambda$, $j = 2, \cdots, d$, where $\alpha = \beta$, then the oracle estimator $\hat{u}_1^{\mathrm{OR}}$ in (3.13) is marginally inconsistent with u_1 in that

$$|< \hat{u}_1^{\mathrm{OR}}, u_1 >| \Rightarrow \left(1 + \frac{c_\lambda}{c\chi_n^2}\right)^{-\frac{1}{2}}, \tag{3.14}$$

where $\Rightarrow$ denotes convergence in distribution, and χ_n^2 denotes the chi-squared distribution with n degrees of freedom.

3.5 Simulations for Sparse PCA

We perform simulation studies to illustrate the performance of the ST method and the RSPCA with the hard thresholding penalty. We fix the sample size at $n = 25$ and consider a range of the dimension $d = 500,\ 1000,\ 2500,\ 5000,\ 10000$. The diverging d allows us to study the convergence behavior of the methods. To generate the data matrix X, we first need to construct the population covariance matrix for X that approximates the conditions of Theorems 3.2.2 and 3.3.1 when the spike index α is greater than the sparsity index β.

For the population covariance matrix, we consider the motivating model in Example 3.1.1 for the first population eigenvector and the eigenvalues, where the first eigenvalue $\lambda_1 = d^\alpha$ and the rest equal one, i.e. $\lambda_j = 1$, $j \geq 2$. For the additional population eigenvectors u_j,

$2 \le j \le \lfloor d^\beta \rfloor$, let the last $d - \lfloor d^\beta \rfloor$ entries of these eigenvectors be zero. In particular, let the eigenvectors u_j, $2 \le j \le \lfloor d^\beta \rfloor$, be proportional to

$$\dot{u}_j = (\overbrace{1, \ldots, 1}^{j-1}, -j+1, 0, \ldots, 0)^T.$$

After normalizing $\dot{u}_j$, we get the j-th eigenvector $u_j = \dot{u}_j / \|\dot{u}_j\|$. For $j > \lfloor d^\beta \rfloor$, let the j-th eigenvector have just one non-zero entry in the j-th position such that $u_j = (\overbrace{0, \ldots, 0}^{j-1}, 1, 0, \ldots, 0)^T$.

Then the data matrix is generated as

$$X = d^{\frac{\alpha}{2}} u_1 z_1^T + \sum_{j=2}^{d} u_j z_j^T,$$

where the z_j follows the n-dimensional standard normal distribution.

We select twenty spike and sparsity pairs (α, β) with $\alpha = \{0.2, 0.4, 0.6, 0.8\}$ and $\beta = \{0, 0.1, 0.3, 0.5, 0.7\}$, which are shown in Figure 3.1. We perform simulation for all twenty pairs under each specific dimension d, a total of 100 simulation setups. For each triplet (α, β, d), we generate 100 realizations of X. Results for four representative (α, β) pairs (indicated by the squares in Figure 3.1) are reported below for the case of $d = 10000$, unless indicated otherwise. Additional simulation results can be found in online supplement (Shen et al., 2012b).

First of all, the plots in Figure 3.2 summarize the results for the pair $(\alpha, \beta) = (0.6, 0.1)$, corresponding to one of the square dots in the white (consistent) triangular area of Figure 3.1. For each replication of X and a range of the thresholding parameter ζ, we obtained the ST estimator $\hat{u}_1^{\mathrm{ST}}$ and the RSPCA estimator $\hat{u}_1$. Then we calculate the angle between the estimates $\hat{u}_1^{\mathrm{ST}}$ (or $\hat{u}_1$) and the true eigenvector u_1 through (3.1). Plotting this angle as a function of the thresholding parameter ζ gives the curve in Panel (A) of Figure 3.2. Since ST and RSPCA perform very similarly in this case, only the RSPCA plots are shown in Figure 3.2. The 100 simulation realizations generate the one-hundred curves in the panel. We rescale the thresholding parameter as $\log_{10}(\zeta + 10^{-5})$, to help reveal clearly the tendency of

the angle curves as the thresholding parameter increases.

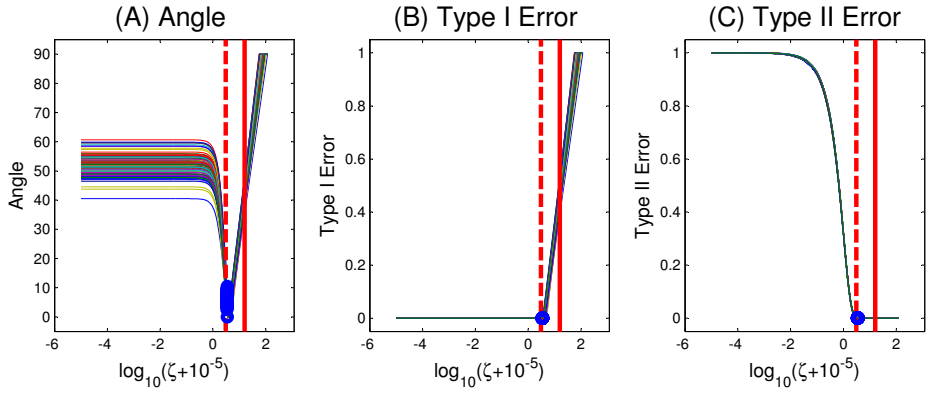

Figure 3.2: Performance summary of RSPCA

In these angle curves, the angles with $\zeta = 0$ (essentially the left edge of each curve) correspond to those obtained by the conventional PCA. Note that these angles are all over 40 degrees which confirms the results of Jung and Marron (2009) that when the spike index $\alpha < 1$, the conventional PCA is not consistent for u_1. As ζ increases, the angle remains stable for a while, then decreases to almost 0 degree, before eventually starting to increase to 90 degrees. The plot suggests that RSPCA does improve over PCA for a range of ζ. When ζ is large enough, all the entries of the RSPCA estimator will be zeroed out, so the estimator eventually becomes a d-dimensional zero vector, and the angle to u_1 goes to 90 degree.

Theorem 3.3.4 suggests a *consistent* range of the thresholding parameter ζ, i.e. when ζ is inside this range, RSPCA would give a consistent estimator for u_1. The boundaries of this consistent range are indicated by the dashed and solid vertical lines in the angle plot. The range is very reasonable in the current case as the angles inside it are all small, and this offers an empirical validation for our asymptotic results.

Zou et al. (2007) suggest the use of the Bayesian Information Criterion (BIC) (Schwarz, 1978) to select the number of the non-zero coefficients for a lasso regression. Lee et al. (2010a) apply this idea to the sparse PCA context. Below we want to investigate how BIC performs numerically.

According to Lee et al. (2010a), for a fixed $\tilde{v}_1$, minimization of (3.10) with respect to $\tilde{u}_1$ is equivalent to minimizing the following penalized regression criterion:

$$\|X - \tilde{u}_1\tilde{v}_1^T\|_F^2 + P_\zeta(\tilde{u}_1) = \|Y - (I_d \bigotimes \tilde{v}_1)\tilde{u}_1\|^2 + P_\zeta(\tilde{u}_1), \tag{3.15}$$

where $Y = (X^1, \ldots, X^d)^T$, with X^i being the i-th row of X, and $\bigotimes$ is the Kronecker product. Following Lee et al. (2010a), we define the BIC for (3.15) as

$$\text{BIC}(\zeta) = \frac{\|Y - \hat{Y}\|^2}{nd\hat{\sigma}^2} + \frac{\log(nd)}{nd}\widehat{df}(\zeta), \tag{3.16}$$

where $\hat{\sigma}^2$ is the ordinary-least squares estimate of the error variance, and $\widehat{df}(\zeta)$ is the degree of sparsity for the thresholding parameter ζ, i.e. the number of non-zero entries in $\tilde{u}_1$. For every step of the iterative procedure of RSPCA, we can use BIC (3.16) to select the thresholding parameter and then obtain the corresponding sparse PC direction, until the algorithm converges.

For every angle curve in the angle plots of Figure 3.2, we use a blue circle to indicate the thresholding parameter ζ that is selected by BIC during the last iterative step of RSPCA, and the corresponding angle. In the current $\alpha = 0.6$, $\beta = 0.1$ context, BIC works well, and all the BIC-selected ζ values are very close, so the 100 circles are essentially over plotted on each other. BIC also works well for the other spike and sparsity pairs (α, β) we considered where $\alpha > \beta$, as shown in Shen et al. (2012b).

Another measure of the success of a sparse estimator is in terms of which entries are zeroed. Type I Error is the proportion of non-zero entries in u_1 that are mistakenly estimated as zero. Type II Error is the proportion of zero entries in u_1 that are mistakenly estimated as non-zero. Similar to the angle curves in Panel (A), there are one hundred Type I Error and Type II Error curves in Panels (B) and (C) of Figure 3.2, respectively. The dashed and solid vertical lines are the same as those in Panel (A). Note that for all the thresholding parameters in the range indicated by the lines, the errors are very small, which is again consistent with the asymptotic results of Theorem 3.3.4. The circles again are selected by BIC and they

correspond to the same thresholding parameter, as in the angle plots. Thus, BIC works well here. BIC also generates similarly very small errors for the other spike and sparsity pairs (α, β) in Figure 3.1 that satisfy $\alpha > \beta$.

In addition to Theorem 3.3.4's consistency region, our Theorem 3.3.5 shows that $\hat{u}_1$ converges at a different rate for different thresholding parameter ζ. We empirically demonstrate this below in Figure 3.3 for the case of $\alpha = 0.8$ and $\beta = 0.5$. We choose two threshold values ζ_1 and ζ_2 just inside the RSPCA consistency region. Theorem 3.3.5 suggests that $\hat{u}_1$ corresponding to ζ_1 would converge faster as the dimension d increases, which is exactly the case in Figure 3.3. Similar phenomena can be observed for the other (α, β) scenarios as well as for the ST estimator, as shown in our online supplement (Shen et al., 2012b).

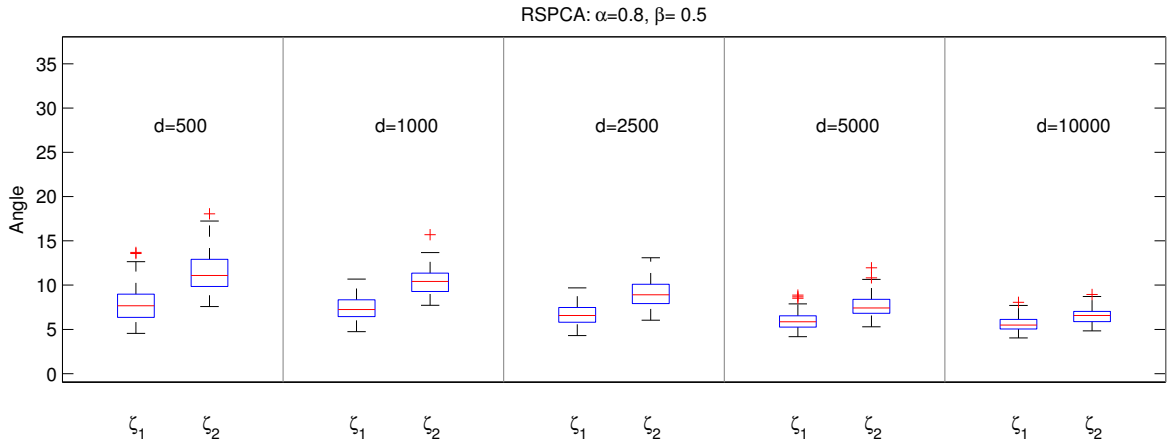

Figure 3.3: Demonstration of convergence for growing d

Next we compare the performance among PCA, ST and RSPCA. In almost all cases, ST and RSPCA give better results than PCA and in some extreme cases, the three methods have similar poor performance. Although in most cases both ST and RSPCA have similar performance, however, there are some cases (for example when $\alpha = 0.4$ and $\beta = 0.3$), where RSPCA performs better than ST. For every simulation replication, we use BIC to select the thresholding parameter and obtain the ST and RSPCA estimators. We then calculate the angle, Type I Error and Type II Error for the three estimators, as well as the difference

between ST and RSPCA (ST minus RSPCA). The 100 values of each measure are summarized using box plots in Figure 3.4.

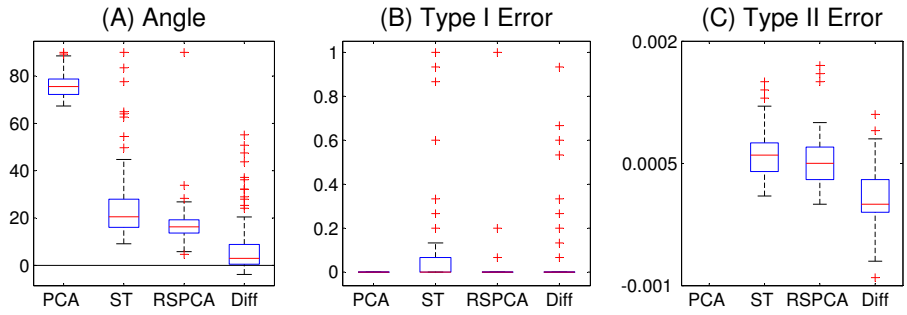

Figure 3.4: Comparison of PCA, ST and RSPCA

Panel (A) of Figure 3.4 shows the box plots of the angles between the maximal eigenvector and the estimates obtained by PCA, ST and RSPCA, as well as the differences between the angles corresponding to ST and RSPCA. Note that the PCA angles are large, compared with ST and RSPCA, indicating the worse performance of PCA. The angle of ST seems larger than RSPCA. For a deeper view of this comparison, the pairwise differences are studied in the fourth box plot of the panel. The angle differences are almost always positive, with some differences bigger than 50 degrees, which suggests that RSPCA has a better performance than ST. Similar conclusions can be made from the box plots of the errors in Panels (B) and (C). The box plot for PCA is not shown in Panel (C) because the corresponding Type II Error almost always equals one, which is far outside the shown range of interest.

The observed improvement of RSPCA over ST deserves some discussion, in connection with Theorems 3.2.3 and 3.3.5 that suggest the same rate of convergence for the two. This simulation compares finite sample performance between the two methods, and we believe the observed (and anticipated) improvement is due to the multiple iterations involved in RSPCA that reduce the variability and make the estimation more stable. Hence such improvement appears asymptotically in the constant coefficient, rather the rate.

Finally, Theorems 3.2.2 and 3.3.4 consider the condition that the spike index α is greater than the sparsity index β. When α is smaller than β, neither ST nor RSPCA is expected to give consistent estimation for the first population eigenvector u_1, as discussed in Section 3.4. For the spike and sparsity pairs (α, β) such that $\alpha < \beta$, the simulation results also confirm this point. Here, we display the simulation plots for the spike and sparsity pair $(\alpha, \beta) = (0.2, 0.7)$ in Figure 3.5 as a representative of such simulations. Since ST and RSPCA have very similar performance here, we just show the simulation results for RSPCA. Similar to Figure 3.2, the circles in Figure 3.5 correspond to the thresholding parameter selected by BIC. From the angle plots, we can see that the angles, selected by BIC, are close to 90 degrees, which suggests the failure of BIC in this case. In fact, all the angle curves are above 80 degrees. Thus, neither ST nor RSPCA generates a reasonable sparse estimator. This is a common phenomenon when the spike index α is smaller than the sparsity index β. It is consistent with the theoretical investigation in Section 3.4.

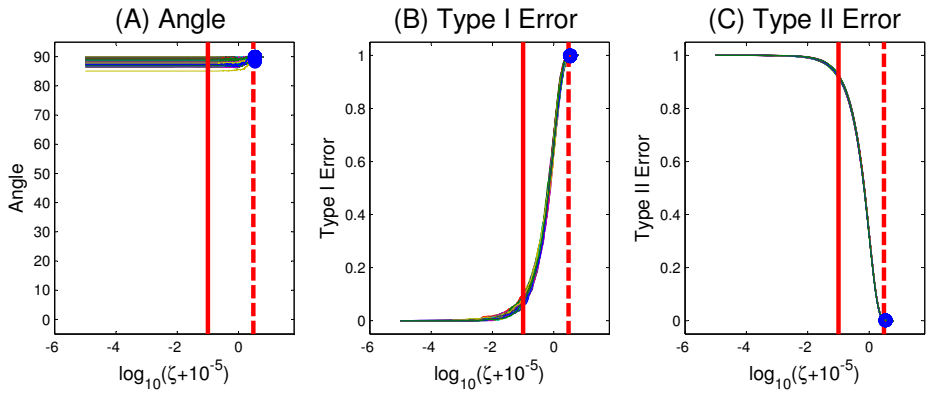

Figure 3.5: Performance summary of RSPCA

Furthermore, the corresponding Type I Error, generated by ST or RSPCA with BIC, is close to one. This further confirms that BIC doesn't work when the spike index α is smaller than the sparsity index β. ST and RSPCA with $\zeta = 0$ is just the conventional PCA, and typically will not generate a sparse estimator. This entails that the Type I Error and Type II Error, corresponding to $\zeta = 0$, respectively equals zero and one. As the thresholding

parameter increases, more and more entries are thresholded out; hence Type I Error increases to one and Type II Error decreases to zero.

3.6 Discussion

In this chapter, we consider single-component Gaussian spike models under the HDLSS asymptotic framework where the sample size n is fixed, the dimension d and the maximal eigenvalue λ_1 both go to infinity. As a comparison, Johnstone and Lu (2009) consider cases that n and d go to infinity together with n/d converging to a constant, while fixing λ_1. The sizes of n and λ_1 contribute positively to (i.e. encourage) the consistency of PCA and sparse PCA, while the dimension d contributes negatively to (i.e. discourages) the consistency. It is interesting in future work to explore the transition from the random matrix asymptotic domain of Johnstone and Lu (2009) to our HDLSS domain of asymptotics. It is also interesting to consider a multiple component spike model as in Paul and Johnstone (2007b); Jung and Marron (2009); Ma (2012). One technical challenge is to figure out what kind of sparsity assumptions (in terms of strength and location) can be simultaneously imposed on the multiple components to make the theoretical setup meaningful in practice. Simultaneous selection of the multiple thresholding parameters involved would be challenging as well. Shen and Huang (2008) suggested to apply RSPCA to extract the leading sparse PCs in a sequential manner, to allow different amounts of sparsity for each PC, which has been shown to have practical advantages.

We also hope to extend our theorems to more general distributions. However, this will be challenging as sparse PCA methods may not work in some extreme non-Gaussian cases, as illustrated in the following example.

Example 3.6.1. *Let $\alpha \in (0,1)$ and $X = (x_1, \ldots, x_d)^T$, where $\{x_j, j = 1, \ldots, d\}$ are independent discrete random variables distributed as follows:*

$$x_1 = \begin{cases} d^{\frac{\alpha}{2}}, & \text{with probability } \frac{1}{2}, \\ -d^{\frac{\alpha}{2}}, & \text{with probability } \frac{1}{2}; \end{cases}$$

$$x_j = \begin{cases} d^{\frac{\alpha+1}{4}}, & \text{with probability } \ d^{-\frac{\alpha+1}{2}}, \\ -d^{\frac{\alpha+1}{4}}, & \text{with probability } \ d^{-\frac{\alpha+1}{2}}, \\ 0, & \text{with probability } \ 1 - 2d^{-\frac{\alpha+1}{2}}. \end{cases} \qquad \text{for} \quad j = 2 \ldots, d$$

Then X has mean 0 and variance-covariance matrix as

$$\Sigma_d = d^{\alpha} u_1 u_1^T + \sum_{j=2}^{d} u_j u_j^T, \quad \text{with} \quad u_k = (0, \ldots, 0, \overbrace{1}^{k}, 0, \ldots, 0)^T.$$

Suppose that we only have sample size $n = 1$, i.e. $X_1 = (x_{1,1}, \ldots, x_{d,1})^T$, then the first empirical eigenvector

$$\hat{u}_1 = (\hat{u}_{1,1}, \ldots, \hat{u}_{d,1})^T = \frac{1}{\sqrt{\sum_{j=1}^{d} x_{j,1}^2}} (x_{1,1}, \ldots, x_{d,1})^T.$$

Under this condition, we can prove that $\mathrm{P}\left(\mathrm{argmax}_k |\hat{u}_{k,d}| = 1\right)$ goes to zero as $d \to \infty$. This suggests that the absolute value of the first entry of $\hat{u}_1$ can not be greater than the others with probability 1, so we can not always threshold out the right entries which would result in the failure of sparse PCA. We perform simulation studies to numerically demonstrate this point, which can be found in our online supplement (Shen et al., 2012b).

3.7 Future Work

There are several interesting problems that we will explore in the future. One is to extend sparse PCA method to other PCs and to study their asymptotic properties. Another is to study the performance of the sparse PCA methods in Johnstone and Lu (2009) and Amini and Wainwright (2009).

3.8 Proofs

3.8.1 Proofs of Theorem 3.2.2 and Theorem 3.2.3

In order to prove Theorem 3.2.2 and Theorem 3.2.3, we use the dependent extreme value result from Lemma 6.1.1 and Theorem 6.1.3 of Leadbetter et al. (1983). An immediate consequence of these is:

Proposition 3.8.1. *Suppose that the standard normal sequence* $\{\xi_i, i = 1, \ldots, \lfloor d^\beta \rfloor\}$ *satisfies the mixing condition (3.9). Let the positive constants* $\{c_i\}$ *be such that* $\sum_{i=1}^{\lfloor d^\beta \rfloor}(1 - \Phi(c_i))$ *is bounded and such that* $C_{\lfloor d^\beta \rfloor} = min_{1 \le i \le \lfloor d^\beta \rfloor}\ c_i \ge c(log(\lfloor d^\beta \rfloor))^{\frac{1}{2}}$ *for some* $c > 0$.

Then

$$P\left[\bigcap_{i=1}^{\lfloor d^\beta \rfloor}\{\xi_i \le c_i\}\right] - \prod_{i=1}^{\lfloor d^\beta \rfloor}\Phi(c_i) \longrightarrow 0, as\ d \to \infty \tag{3.17}$$

holds, where Φ *is the standard normal distribution function. If further*

$$\sum_{i=1}^{\lfloor d^\beta \rfloor}(1 - \Phi(c_i)) \longrightarrow \jmath, as\ d \to \infty, \tag{3.18}$$

holds for some $\jmath \ge 0$, *then*

$$P\left[\bigcap_{i=1}^{\lfloor d^\beta \rfloor}\{\xi_i \le c_i\}\right] \longrightarrow e^{-\jmath}, as\ d \to \infty. \tag{3.19}$$

Proposition 3.8.1 is used to control the right side of (3.8) through the following Lemma.

Lemma 3.8.1. *Suppose that* $\xi_i \sim N(0, \delta_{i,i})$ *satisfies the mixing condition (3.9), where* δ_{ij} *is the covariance of the normal sequence* $\{\xi_i\}$, $i, j = 1, \cdots, \lfloor d^\beta \rfloor$. *If* $C_{\lfloor d^\beta \rfloor} \ge (log(\lfloor d^\beta \rfloor))^\delta max_{1 \le i \le \lfloor d^\beta \rfloor}\delta_{ii}^{\frac{1}{2}}$, *where* $\delta \in (\frac{1}{2}, \infty)$, *then*

$$C_{\lfloor d^\beta \rfloor}^{-1} max_{1 \le i \le \lfloor d^\beta \rfloor}|\xi_i| \xrightarrow{p} 0, as\ d \to \infty. \tag{3.20}$$

Proof. Note that for every $\tau > 0$

$$
\begin{aligned}
&P\left[C_{\lfloor d^\beta \rfloor}^{-1}\max_{1\le i\le \lfloor d^\beta\rfloor}|\xi_i| > \tau\right] = P\left[\max_{1\le i\le \lfloor d^\beta\rfloor}|\xi_i| > C_{\lfloor d^\beta\rfloor}\tau\right] \qquad (3.21)\\
&\le P\left[\left\{\max_{1\le i\le \lfloor d^\beta\rfloor}\xi_i > C_{\lfloor d^\beta\rfloor}\tau\right\}\bigcup\left\{\max_{1\le i\le \lfloor d^\beta\rfloor}(-\xi_i) > C_{\lfloor d^\beta\rfloor}\tau\right\}\right]\\
&\le P\left[\max_{1\le i\le \lfloor d^\beta\rfloor}\xi_i > C_{\lfloor d^\beta\rfloor}\tau\right] + P\left[\max_{1\le i\le \lfloor d^\beta\rfloor}(-\xi_i) > C_{\lfloor d^\beta\rfloor}\tau\right]\\
&= 2P\left[\max_{1\le i\le \lfloor d^\beta\rfloor}\xi_i > C_{\lfloor d^\beta\rfloor}\tau\right]\\
&\le 2\left(1 - P\left[\bigcap_{i=1}^{\lfloor d^\beta\rfloor}\left\{\xi_i\delta_{ii}^{-\frac{1}{2}} \le c(log(\lfloor d^\beta\rfloor))^\delta\right\}\right]\right),
\end{aligned}
$$

where c is a positive constant. Since

$$
\sum_{i=1}^{\lfloor d^\beta\rfloor}\left(1 - \Phi\left(c(log(\lfloor d^\beta\rfloor))^\delta\right)\right) \longrightarrow 0, as\ d\to\infty,
$$

it follows from Proposition 3.8.1 that

$$
P\left[\bigcap_{i=1}^{\lfloor d^\beta\rfloor}\left\{\xi_i\delta_{ii}^{-\frac{1}{2}} \le c(log(\lfloor d^\beta\rfloor))^\delta\right\}\right] \longrightarrow 1, as\ d\to\infty. \qquad (3.22)
$$

From (3.21) and (3.22), we can get

$$
C_{\lfloor d^\beta\rfloor}^{-1}\max_{1\le i\le \lfloor d^\beta\rfloor}|\xi_i| \xrightarrow{p} 0, as\ d\to\infty.
$$

□

Now we will begin the proof of Theorem 3.2.2 and Theorem 3.2.3. Denote $\tilde{X}_j = (x_{j,1},\cdots,x_{j,n})^T$, $\tilde{Z}_j = (z_{j,1},\cdots,z_{j,n})^T$, $\tilde{W}_j = (w_{j,1},\cdots,w_{j,n})^T$and $\tilde{H}_j = (h_{j,1},\cdots,h_{j,n})$, $j = 1,\cdots,d$. Note that

$$
|<\hat{u}_1^{Th}, u_1>| = \frac{|\sum_{j=1}^{[d^\beta]}\breve{\mu}_{j,1}\mu_{j,1}|}{\sqrt{\sum_{j=1}^{d}(\breve{\mu}_{j,1})^2}} = \frac{\lambda_1^{-\frac{1}{2}}|\sum_{j=1}^{[d^\beta]}\breve{\mu}_{j,1}\mu_{j,1}|}{\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{d}(\breve{\mu}_{j,1})^2}}. \qquad (3.23)
$$

Since $\tilde{X}_j = \mu_{j,1}\tilde{Z}_1 + \tilde{H}_j, j = 1, \cdots, d$, it follows that $\tilde{v}_1^T\tilde{X}_j = \mu_{j,1}\tilde{v}_1^T\tilde{Z}_1 + \tilde{v}_1^T\tilde{H}_j$, which yields

$$\begin{aligned}
\breve{\mu}_{j,1} &= \mu_{j,1}\tilde{v}_1^T\tilde{Z}_1 1_{\{|\tilde{v}_1^T\tilde{X}_j|>\lambda\}} + \tilde{v}_1^T\tilde{H}_j 1_{\{|\tilde{v}_1^T\tilde{X}_j|>\lambda\}} \\
&= \mu_{j,1}\tilde{v}_1^T\tilde{Z}_1 + \mu_{j,1}\tilde{v}_1^T\tilde{Z}_1 1_{\{|\tilde{v}_1^T\tilde{X}_j|\leq\lambda\}} + \tilde{v}_1^T\tilde{H}_j 1_{\{|\tilde{v}_1^T\tilde{X}_j|>\lambda\}},
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j=1}^{[d^\beta]}\breve{\mu}_{j,1}\mu_{j,1} &= \sum_{j=1}^{[d^\beta]}\mu_{j,1}^2\tilde{v}_1^T\tilde{Z}_1 1_{\{|\tilde{v}_1^T\tilde{X}_j|>\lambda\}} + \sum_{j=1}^{[d^\beta]}\mu_{j,1}\tilde{v}_1^T\tilde{H}_j 1_{\{|\tilde{v}_1^T\tilde{X}_j|>\lambda\}} \\
&= \tilde{v}_1^T\tilde{Z}_1 + \sum_{j=1}^{[d^\beta]}\mu_{j,1}^2\tilde{v}_1^T\tilde{Z}_1 1_{\{|\tilde{v}_1^T\tilde{X}_j|\leq\lambda\}} + \sum_{j=1}^{[d^\beta]}\mu_{j,1}\tilde{v}_1^T\tilde{H}_j 1_{\{|\tilde{v}_1^T\tilde{X}_j|>\lambda\}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\lambda_1^{-\frac{1}{2}}|\sum_{j=1}^{[d^\beta]}\breve{\mu}_{j,1}\mu_{j,1}| &\leq \lambda_1^{-\frac{1}{2}}\sum_{j=1}^{[d^\beta]}\mu_{j,1}^2|\tilde{v}_1^T\tilde{Z}_1| + \lambda_1^{-\frac{1}{2}}\sum_{j=1}^{[d^\beta]}|\mu_{j,1}\tilde{v}_1^T\tilde{H}_j| \qquad (3.24)\\
&= |\tilde{v}_1^T\tilde{W}_1| + \sum_{j=1}^{[d^\beta]}\sum_{i=1}^{n}\lambda_1^{-\frac{1}{2}}|\mu_{j,1}h_{j,i}|,
\end{aligned}$$

and

$$\begin{aligned}
&\lambda_1^{-\frac{1}{2}}|\sum_{j=1}^{[d^\beta]}\breve{\mu}_{j,1}\mu_{j,1}| \qquad (3.25)\\
&\geq \lambda_1^{-\frac{1}{2}}|\tilde{v}_1^T\tilde{Z}_1| - \lambda_1^{-\frac{1}{2}}\sum_{j=1}^{[d^\beta]}\mu_{j,1}^2|\tilde{v}_1^T\tilde{Z}_1|1_{\{|\tilde{v}_1^T\tilde{X}_j|\leq\lambda\}} - \lambda_1^{-\frac{1}{2}}\sum_{j=1}^{[d^\beta]}|\mu_{j,1}\tilde{v}_1^T\tilde{H}_j| \\
&\geq |\tilde{v}_1^T\tilde{W}_1| - |\tilde{v}_1^T\tilde{W}_1|\sum_{j=1}^{[d^\beta]}\mu_{j,1}^2 1_{\{|\tilde{v}_1^T\tilde{X}_j|\leq\lambda\}} - \sum_{j=1}^{[d^\beta]}\sum_{i=1}^{n}\lambda_1^{-\frac{1}{2}}|\mu_{j,1}h_{j,i}|,
\end{aligned}$$

and

$$\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{d}\breve{\mu}_{j,1}^2} \le \lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{[d^\beta]}\breve{\mu}_{j,1}^2} + \lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=[d^\beta]+1}^{d}\breve{\mu}_{j,1}^2} \tag{3.26}$$

$$\le \lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{[d^\beta]}\left(\mu_{j,1}\tilde{v}_1^T\tilde{Z}_1\right)^2} + \lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{[d^\beta]}(\tilde{v}_1^T\tilde{H}_j)^2} + \lambda_1^{-\frac{1}{2}}\sum_{j=[d^\beta]+1}^{d}|\breve{\mu}_{j,1}|$$

$$= |\tilde{v}_1^T\tilde{W}_1| + \sqrt{\sum_{j=1}^{[d^\beta]}\lambda_1^{-1}(\sum_{i=1}^{n}|h_{j,i}|)^2} +$$

$$+ \sum_{j=[d^\beta]+1}^{d}\sum_{i=1}^{n}\lambda_1^{-\frac{1}{2}}|h_{j,i}|1_{\{\sum_{i=1}^{n}|h_{j,i}|>\lambda\}},$$

and

$$\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{d}\breve{\mu}_{j,1}^2} \ge \lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{[d^\beta]}\breve{\mu}_{j,1}^2} \tag{3.27}$$

$$\ge |\tilde{v}_1^T\tilde{W}_1| - |\tilde{v}_1^T\tilde{W}_1|\sqrt{\sum_{j=1}^{[d^\beta]}\mu_{j,1}^2 1_{\{|\tilde{v}_1^T\tilde{X}_j|\le\lambda\}}} - \sqrt{\sum_{j=1}^{[d^\beta]}\lambda_1^{-1}(\sum_{i=1}^{n}|h_{j,i}|)^2}.$$

Next we will show that

$$\sum_{j=1}^{[d^\beta]}\sum_{i=1}^{n}\lambda_1^{-\frac{1}{2}}|\mu_{j,1}h_{j,i}| = o_p(d^{-\frac{\varsigma}{2}}), \text{where } \varsigma \in [0, \alpha-\eta-\theta). \tag{3.28}$$

Since $H_i = (h_{1,i}, \cdots, h_{d,i})^T = \sum_{k=2}^{d} z_{k,i}\mu_k, i = 1, \cdots, n$, it follows that $h_{j,i} = \sum_{k=2}^{d}\mu_{j,k}z_{k,i} = \sum_{k=2}^{d}\mu_{j,k}\lambda_k^{\frac{1}{2}}w_{k,i} \sim \mathrm{N}(0, \sigma_{j,i}^2)$, where $\sigma_{j,i}^2 \le \lambda_2$, $j = 1, \cdots, [d^\beta]$, $i = 1, \cdots, n$. Thus, for fix τ

$$P\left[\sum_{j=1}^{[d^\beta]}\sum_{i=1}^{n}d^{\frac{\varsigma}{2}}\lambda_1^{-\frac{1}{2}}|\mu_{j,1}h_{j,i}| \ge \tau\right] \le P\left[\bigcup_{j=1}^{[d^\beta]}\left\{\sum_{i=1}^{n}|\mu_{j,1}h_{j,i}| \ge d^{-\frac{\varsigma}{2}}\lambda_1^{\frac{1}{2}}\tau\mu_{j,1}^2\right\}\right]$$

$$\le \sum_{j=1}^{[d^\beta]}\sum_{i=1}^{n}P\left[|h_{j,i}| \ge n^{-1}d^{-\frac{\varsigma}{2}}\lambda_1^{\frac{1}{2}}\tau|\mu_{j,1}|\right] \le \sum_{j=1}^{[d^\beta]}\sum_{i=1}^{n}P\left[|h_{j,i}\sigma_{j,i}^{-1}| \ge c^* d^{\frac{\alpha-\eta-\theta-\varsigma}{2}}\right]$$

$$= 2n[d^\beta]\int_{cd^{(\alpha-\eta-\theta-\varsigma)/2}}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}dx \longrightarrow 0, \text{as } d \to \infty,$$

where c is constant. Similar, we can show that

$$\sum_{j=1}^{[d^{\beta}]} \lambda_1^{-1} (\sum_{i=1}^{n} |h_{j,i}|)^2 = o_p(d^{\frac{\varsigma}{2}}), \tag{3.29}$$

and

$$\sum_{j=[d^{\beta}]+1}^{d} \sum_{i=1}^{n} \lambda_1^{-\frac{1}{2}} |h_{j,i}| 1_{\{\sum_{i=1}^{n} |h_{j,i}| > \lambda\}} = o_p(d^{-\frac{\varsigma}{2}}), \tag{3.30}$$

where $\varsigma \in [0, \alpha - \eta - \theta)$. Finally, we want to show that

$$\sum_{j=1}^{[d^{\beta}]} \mu_{j,1}^2 1_{\{|\tilde{v}_1^T \tilde{X}_j| \leq \lambda\}} = o_p(d^{-\frac{\varsigma^{'}}{2}}), \tag{3.31}$$

where $\varsigma^{'}$ satisfies that $d^{\frac{\varsigma^{'}+\eta-\alpha}{2}} \lambda = o(1)$. Since we can always find a subsequence of $\{\frac{\lambda_1}{\sum_{j=2}^{d} \lambda_j}\}$ and make it convergent to a nonnegative constant, for simplicity, we just assume that $\lim_{d\to\infty} \frac{\lambda_1}{\sum_{j=2}^{d} \lambda_j} = C$. If $C = 0$, then the spike index $\alpha < 1$ and Jung and Marron (2009) shows that

$$c_d^{-1} S_d \quad \xrightarrow{p} \quad I_n, \text{ as } d \to \infty,$$

where $c_d = n^{-1} \sum_{j=1}^{d} \lambda_j$. Since the eigenvector $\tilde{v}_1^T$ of $c_d^{-1} S_d$ can be chosen so that they are continuous from Acker (1974), it follows that $\tilde{v}_1^T \Rightarrow v_1$, as $d \to \infty$, where $\Rightarrow$ denotes the convergence in distribution and v_1 is the first eigenvector of n-dimensional identity matrix. If $C = 0$, then the spike index $\alpha = 1$ and Jung et al. (2012) shows that $\tilde{v}_1^T \Rightarrow \frac{\tilde{W}_1}{\|\tilde{W}_1\|}$, as $d \to \infty$. Therefore, we have

$$| \tilde{v}_1^T \tilde{W}_1 | \Rightarrow | v_1^T \tilde{W}_1 | \text{ or } \| \tilde{W}_1 \|, \text{as } d \to \infty. \tag{3.32}$$

Since $d^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\lambda = o(1)$, $d^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\sum_{i=1}^{n}\max_{1\leq j\leq[d^{\beta}]}|h_{j,i}| = o_p(1)$, and

$$\begin{aligned}
&\sum_{j=1}^{[d^{\beta}]} d^{\frac{\varsigma}{2}^{'}}\mu_{j,1}^2 1_{\{|\tilde{v}_1^T\tilde{X}_j|\leq\lambda\}} \leq \sum_{j=1}^{[d^{\beta}]} d^{\frac{\varsigma}{2}^{'}}\mu_{j,1}^2 1_{\{|\mu_{j,1}\tilde{v}_1^T\tilde{Z}_1|\leq|\tilde{v}_1^T\tilde{H}_j|+\lambda\}} \\
&\leq \sum_{j=1}^{[d^{\beta}]} d^{\frac{\varsigma}{2}^{'}}\mu_{j,1}^2 1_{\left\{|\tilde{v}_1^T\tilde{W}_1|\leq\lambda_1^{-\frac{1}{2}}\max_{1\leq j\leq[d^{\beta}]}|\mu_{j,1}|^{-1}\left(\sum_{i=1}^{n}\max_{1\leq j\leq[d^{\beta}]}|h_{j,i}|+\lambda\right)\right\}} \\
&\leq \frac{cd^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\sum_{i=1}^{n}\max_{1\leq j\leq[d^{\beta}]}|h_{j,i}| + cd^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\lambda}{|\tilde{v}_1^T\tilde{W}_1|},
\end{aligned}$$

where c is a constant, it follows that (3.31) is established. Then we have

$$\lambda_1^{-\frac{1}{2}}|\sum_{j=1}^{[d^{\beta}]}\breve{\mu}_{j,1}\mu_{j,1}| = |\tilde{v}_1^T\tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}), \tag{3.33}$$

from (3.24), (3.25), (3.28), and (3.31) , and

$$\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{d}\breve{\mu}_{j,1}^2} = |\tilde{v}_1^T\tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}), \tag{3.34}$$

from (3.26), (3.27), (3.29), (3.30) and (3.31). Furthermore, we will have

$$|<\hat{u}_1^{Th}, u_1>| = \frac{|\tilde{v}_1^T\tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}})}{|\tilde{v}_1^T\tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}})} = 1 + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}),$$

from (3.23), (3.32), (3.33), and (3.34). It means that $\hat{u}_1^{Th}$ is consistent with u_1 with convergence rate $d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}$. Note that $d^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\lambda = o(1)$. If $\lambda = o(d^{\frac{\alpha-\eta-\varsigma}{2}})$, then we can take $\varsigma^{'} = \varsigma$. Then $\hat{u}_1^{Th}$ is consistent with u_1 with convergence rate $d^{\frac{\varsigma}{2}}$. This finishes the proof of Theorem 3.2.2 and Theorem 3.2.3.

3.8.2 Proofs of Theorem 3.3.1, 3.3.2, 3.3.3, 3.3.4 and 3.3.5

The proof of Theorem 3.3.1 and 3.3.2 are similar as they appears in Section 3.8.2. The proof of Theorem 3.3.3 is shown in Section 3.8.2. Theorem 3.3.4 and 3.3.5 are again similar with

the proof shown in Section 3.8.2.

Proofs of Theorem 3.3.1 and Theorem 3.3.2

Assume that $\hat{u}_1 = \frac{\breve{u}_1^p}{\|\breve{u}_1^p\|}$ and the entries of $\breve{u}_1^p$ have the formula $\breve{\mu}_{j,1}^p = \tilde{\mu}_{j,1} 1_{\{|\tilde{\mu}_{j,1}|>\lambda'\}} + \varsigma_j 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}$, where $\tilde{\mu}_{j,1}$ is defined in (3.3), $\lambda^{'} = \lambda$ or $a\lambda$ and $\mid \varsigma_j \mid < c\lambda$ (c is a constant). Denote $\breve{\mu}_{j,1} = \tilde{\mu}_{j,1} 1_{\{|\tilde{\mu}_{j,1}|>\lambda'\}}$, then $\breve{\mu}_{j,1}^p = \breve{\mu}_{j,1} + \varsigma_j 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}$. Note that

$$| < \hat{\mu}_1, \mu_1 > | = \frac{|\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}^p \mu_{j,1}|}{\sqrt{\sum_{j=1}^d (\breve{\mu}_{j,1}^p)^2}} = \frac{\lambda_1^{-\frac{1}{2}} |\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}^p \mu_{j,1}|}{\lambda_1^{-\frac{1}{2}} \sqrt{\sum_{j=1}^d (\breve{\mu}_{j,1}^p)^2}}. \tag{3.35}$$

In addition, we have

$$\begin{aligned} &\lambda_1^{-\frac{1}{2}} |\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}\mu_{j,1}| - c\lambda\lambda_1^{-\frac{1}{2}} \sum_{j=1}^{[d^\beta]} \mid \mu_{j,1} \mid \leq \lambda_1^{-\frac{1}{2}} |\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}^p \mu_{j,1}| \leq \\ &\leq \lambda_1^{-\frac{1}{2}} |\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}\mu_{j,1}| + c\lambda\lambda_1^{-\frac{1}{2}} \sum_{j=1}^{[d^\beta]} \mid \mu_{j,1} \mid, \end{aligned} \tag{3.36}$$

and

$$\begin{aligned} &\lambda_1^{-\frac{1}{2}} \sqrt{\sum_{j=1}^d (\breve{\mu}_{j,1})^2} - c\lambda\lambda_1^{-\frac{1}{2}} \sqrt{\sum_{j=1}^d 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}} \leq \lambda_1^{-\frac{1}{2}} \sqrt{\sum_{j=1}^d (\breve{\mu}_{j,1}^p)^2} \leq \\ &\leq \lambda_1^{-\frac{1}{2}} \sqrt{\sum_{j=1}^d (\breve{\mu}_{j,1})^2} + c\lambda\lambda_1^{-\frac{1}{2}} \sqrt{\sum_{j=1}^d 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}}. \end{aligned} \tag{3.37}$$

Since $\lambda^{'}$ satisfies the condition (e) in Theorem 3.2.2, it follows that $\breve{\mu}_{j,1}$ has the same property as $\breve{\mu}_{j,1}$ in the proof of Theorem 3.2.2. Recall (3.33) and (3.34) from the proof of Theorem 3.2.2

$$\lambda_1^{-\frac{1}{2}} |\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}\mu_{j,1}| = |\tilde{v}_1^T \tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}), \tag{3.38}$$

and

$$\lambda_1^{-\frac{1}{2}} |\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}\mu_{j,1}| = |\tilde{v}_1^T \tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}), \tag{3.39}$$

Where $\varsigma \in [0, \alpha-\eta-\theta)$ and $\varsigma^{'}$ satisfies $d^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\lambda = o(1)$. Since (recall $\beta \leq \eta$) $d^{\frac{\varsigma^{'}}{2}}\lambda\lambda_1^{-\frac{1}{2}} \sum_{j=1}^{[d^\beta]} \mid \mu_{j,1} \mid \leq d^{\frac{\varsigma^{'}}{2}}\lambda\lambda_1^{-\frac{1}{2}} d^{\frac{\beta}{2}} \leq c d^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\lambda = o(1)$, it follows that

$$\lambda\lambda_1^{-\frac{1}{2}} \sum_{j=1}^{[d^\beta]} \mid \mu_{j,1} \mid = o(d^{-\frac{\varsigma^{'}}{2}}). \tag{3.40}$$

Thus, (3.36), (3.38) and (3.40) together illustrate that

$$\lambda_1^{-\frac{1}{2}} |\sum_{j=1}^{[d^\beta]} \breve{\mu}_{j,1}^p\mu_{j,1}| = |\tilde{v}_1^T \tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}) \tag{3.41}$$

Now, we will show that

$$\lambda\lambda_1^{-\frac{1}{2}} \sqrt{\sum_{j=1}^{d} 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}} = o_p(d^{-\frac{\varsigma^{'}}{2}}). \tag{3.42}$$

For fixed τ,

$$\begin{aligned}
&P\left[d^{\varsigma^{'}}\lambda^2\lambda_1^{-1} \sum_{j=[d^\beta]+1}^{d} 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}} \geq \tau\right] \\
&\leq \sum_{j=[d^\beta]+1}^{d} P\left[1_{\{|\tilde{\mu}_{j,1}|>\lambda\}} \geq d^{-\varsigma^{'}}\lambda^{-2}\lambda_1 d^{-1}\tau\right] \\
&\leq \sum_{j=[d^\beta]+1}^{d} \sum_{i=1}^{n} P\left[\mid h_{j,i}\sigma_{j,i}^{-1} \mid > c d^\theta\lambda\right] \\
&\leq n(d-[d^\beta]) \int_{cd^\theta\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx \longrightarrow 0, \text{as } d \to \infty,
\end{aligned}$$

which yields

$$\sqrt{\lambda^2\lambda_1^{-1}\sum_{j=[d^\beta]+1}^{d}1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}} = o_p(d^{-\frac{\varsigma^{'}}{2}}). \tag{3.43}$$

In addition, since $\lambda\lambda_1^{-\frac{1}{2}}d^{\frac{\beta}{2}} = o(d^{-\frac{\varsigma^{'}}{2}})$ and (3.43), it follows that

$$\begin{aligned}
&\lambda\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{d}1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}} \\
&\leq \lambda\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{[d^\beta]}1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}} + \lambda\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=[d^\beta]+1}^{d}1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}} \\
&\leq \lambda\lambda_1^{-\frac{1}{2}}d^{\frac{\beta}{2}} + \sqrt{\lambda^2\lambda_1^{-1}\sum_{j=[d^\beta]+1}^{d}1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}},
\end{aligned}$$

we have (3.42). Then, we can get

$$\lambda_1^{-\frac{1}{2}}\sqrt{\sum_{j=1}^{d}(\breve{\mu}_{j,1}^{p})^2} = |\tilde{v}_1^T\tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}). \tag{3.44}$$

from (3.37),(3.39) and (3.42). Furthermore, we have

$$|<\hat{u}_1, u_1>| = \frac{|\tilde{v}_1^T\tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}})}{|\tilde{v}_1^T\tilde{W}_1| + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}})} = 1 + o_p(d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}),$$

from (3.35),(3.41) and (3.44). This means that $\breve{u}_1^p$ is consistent with u_1 with convergence rate $d^{-\frac{\min\{\varsigma,\varsigma^{'}\}}{2}}$. Note that $d^{\frac{\varsigma^{'}+\eta-\alpha}{2}}\lambda = o(1)$. If $\lambda = o(d^{\frac{\alpha-\eta-\varsigma}{2}})$, then we can take $\varsigma^{'} = \varsigma$. Therefore, $\breve{u}_1^p$ is consistent with u_1 with convergence rate $d^{\frac{\varsigma}{2}}$.

- Let $\breve{u}_1^{soft} = h_\lambda^{soft}(X_{(d)}\tilde{v}_1)$, then the entries $\breve{u}_{j,1}^{soft}$ of $\breve{u}_1^{soft}$ has the formula $\breve{u}_{j,1}^{soft} = \tilde{\mu}_{j,1}1_{\{|\tilde{\mu}_{j,1}|>\lambda\}} - \text{sign}(\tilde{\mu}_{j,1})\lambda 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}$.
- Let $\breve{u}_1^{hard} = h_\lambda^{hard}(X_{(d)}\tilde{v}_1)$, then the entries $\breve{u}_{j,1}^{hard}$ of $\breve{u}_1^{hard}$ has the formula $\breve{u}_{j,1}^{hard} = \tilde{\mu}_{j,1}1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}$.

- Let $\breve{u}_1^{SCAD} = h_\lambda^{SCAD}(X_{(d)}\tilde{v}_1)$, then the entries $\breve{u}_{j,1}^{SCAD}$ of $\breve{u}^{SCAD}$ has the formula $\breve{u}_{j,1}^{SCAD} = \tilde{\mu}_{j,1}1_{\{|\tilde{\mu}_{j,1}|>a\lambda\}} + \varsigma_j 1_{\{|\tilde{\mu}_{j,1}|>\lambda\}}$, where $\mid \varsigma_j \mid< a\lambda$.

Since $\breve{u}_1^{soft}$,$\breve{u}_1^{hard}$ and $\breve{u}_1^{SCAD}$ are just special cases of $\breve{u}_1^{p}$, it follows that $\hat{u}_1$ (the normalized vector of $\breve{u}_1^{soft}$, $\breve{u}_1^{hard}$ or $\breve{u}_1^{SCAD}$) is consistent with u_1 with convergence rate $d^{\frac{\varsigma}{2}}$.

Proof of Theorem 3.3.3

note that

$$\tilde{v}_1^{new} = \frac{\lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \lambda_1^{-\frac{1}{2}} X_{(d)}^T u_1}{\parallel \lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \lambda_1^{-\frac{1}{2}} X_{(d)}^T u_1 \parallel} = \frac{\lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \tilde{W}_1}{\parallel \lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \tilde{W}_1 \parallel}.$$

Jung and Marron (2009) shows that $\frac{\hat{\lambda}_1}{d} \xrightarrow{p} c$, as $d \to \infty$, where c is a constant. Since $\parallel \hat{u}_1^{old} - u_1 \parallel = o_p(d^{\frac{\varsigma}{2}})$, where $\varsigma \geq 1-\alpha$, it follows that $\lambda_1^{-\frac{1}{2}}\hat{\lambda}_1^{\frac{1}{2}} \parallel \hat{u}_1^{old} - u_1 \parallel = o_p(1)$. Therefore, we have

$$\parallel \lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) \parallel \leq n\lambda_1^{-\frac{1}{2}}\hat{\lambda}_1^{\frac{1}{2}} \parallel \hat{u}_1^{old} - u_1 \parallel = o_p(1),$$

which yields

$$\lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \tilde{W}_1 = \tilde{W}_1 + o_p(1),$$

and

$$\begin{aligned} \parallel \lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \tilde{W}_1 \parallel &\leq \parallel \lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) \parallel + \parallel \tilde{W}_1 \parallel \\ &= \parallel \tilde{W}_1 \parallel + o_p(1). \end{aligned}$$

Thus,

$$\begin{aligned}\tilde{v}_1^{new} &= \frac{\lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \tilde{W}_1}{\| \lambda_1^{-\frac{1}{2}} X_{(d)}^T(\hat{u}_1^{old} - u_1) + \tilde{W}_1 \|} \\ &= \frac{\tilde{W}_1 + o_p(1)}{\| \tilde{W}_1 \| + o_p(1)} \xrightarrow{p} \frac{\tilde{W}_1}{\| \tilde{W}_1 \|}, \text{as } d \to \infty.\end{aligned}$$

Proof of Theorem 3.3.4 and Theorem 3.3.5

Note that $\tilde{v}_1^{new} \xrightarrow{p} \frac{\tilde{W}_1}{\|\tilde{W}_1\|}$, as $d \to \infty$ from Theorem 3.3.3, then we just need to replace $\hat{u}_1$ with $\tilde{u}_1^{new}$ and $\tilde{v}_1$ with $\tilde{v}_1^{new}$ in the proofs of Theorem 3.3.4 and Theorem 3.3.5 to get Theorem 3.3.4 and Theorem 3.3.5.

3.8.3 Proof of Theorem 3.4.1

Since $X_i^* = (I_{\lfloor d^\beta \rfloor}, (0)_{\lfloor d^\beta \rfloor \times (d - \lfloor d^\beta \rfloor)}) X_i$, where $I_{\lfloor d^\beta \rfloor}$ denotes the $\lfloor d^\beta \rfloor$-dimensional identity matrix and $(0)_{\lfloor d^\beta \rfloor \times (d - \lfloor d^\beta \rfloor)}$ is the $\lfloor d^\beta \rfloor$-by-$(d - \lfloor d^\beta \rfloor)$ zero matrix, $j = 1, \ldots, n$, it follows that

$$\Sigma^*_{\lfloor d^\beta \rfloor} = (I_{\lfloor d^\beta \rfloor}, (0)_{\lfloor d^\beta \rfloor \times (d - \lfloor d^\beta \rfloor)}) \Sigma_d (I_{\lfloor d^\beta \rfloor}, (0)_{\lfloor d^\beta \rfloor \times (d - \lfloor d^\beta \rfloor)})^T,$$

which yields

$$\lambda_1^* = \lambda_1, \lambda_2 \geq \lambda_j^* \geq \lambda_d, j = 2, \ldots, \lfloor d^\beta \rfloor. \tag{3.45}$$

Therefore,

$$\frac{\sum_{j=2}^{\lfloor d^\beta \rfloor} \lambda_j^{*2}}{(\sum_{j=2}^{\lfloor d^\beta \rfloor} \lambda_j^*)^2} \leq \frac{\lfloor d^\beta \rfloor \lambda_2^2}{(\lfloor d^\beta \rfloor)^2 \lambda_d^2} = \frac{O(d^\beta) O(d^{2\theta})}{O(d^{2\beta})} = o(1), \tag{3.46}$$

and

$$\frac{\lambda_1^*}{\sum_{j=2}^{\lfloor d^\beta \rfloor} \lambda_j^*} \leq \frac{\lambda_1}{\lfloor d^\beta \rfloor \lambda_d} = \frac{O(d^\alpha)}{O(d^\beta)} = o(1). \tag{3.47}$$

If we rescale λ_j^*, $j = 1, \ldots, \lfloor d^\beta \rfloor$, (3.46) satisfies the ε_2 assumption of Jung and Marron (2009) and (3.47) satisfies the assumption $\lambda_1 = O(d^\alpha)$ and $\sum_{j=2}^{d} \lambda_j = O(d)$, where $\alpha < 1$. For this case, Jung and Marron (2009) have shown that $\hat{u}_1^*$ is strongly inconsistent with u_1^*. This means that the oracle estimator $\hat{u}_1^{\mathrm{OR}}$ is strongly inconsistent with u_1.

For the boundary case of $\alpha = \beta$, WLOG, we assume that $\lambda_1 = cd^\alpha$ and $\lambda_j = c_\lambda$, $j = 2, \cdots, d$. Then (3.45) becomes

$$\lambda_1^* = \lambda_1 = cd^\alpha, \lambda_j^* = c_\lambda, j = 2, \ldots, \lfloor d^\beta \rfloor.$$

Note that the dimension of X_i^* is $\lfloor d^\beta \rfloor = \lfloor d^\alpha \rfloor$. If we treat the effective dimension $\lfloor d^\alpha \rfloor$ as d in Jung et al. (2012), it then follows that

$$|< \hat{u}_1^*, u_1^* >| \Rightarrow \left(1 + \frac{c_\lambda}{c\chi_n^2}\right)^{-\frac{1}{2}}.$$

The above, together with (3.12) and (3.13), leads to the inconsistency of $\hat{u}_1^{\mathrm{OR}}$ (3.14).

Chapter 4

Analysis of Tree Data

4.1 Introduction

The tree structured data objects, i.e. the brain artery trees (Aylward and Bullitt, 2002) in Section 4.2, are important in medical image analysis. The application of the statistical analysis, i.e. PCA, to the tree data can be challenging because the tree space is a non-Euclidean space, as seen in Wang and Marron (2007). We present an approach, the *Dyck path representation*, to build a bridge between tree space (a non-Euclidean space) and curve space (standard Euclidean space). Then, we can exploit the power of functional data analysis to study statistical properties of tree data sets. Besides the Dyck path representation, we also develop a novel *branch length representation* to connect tree space and curve space. Furthermore, we present a *pruned tree* idea to statistically analyze properties of tree structured data at a range of scales. Projection is a fundamental tool in classical Functional Data Analysis (FDA), see Ramsay and Silverman (2002) and Ramsay and Silverman (2005). We project on PC directions (Jolliffe, 2002) to explore population variation, and further transform the variation in terms of trees, as in Section 4.5.1. In addition, partial least square (PLS) (Wold et al., 1984), and canonical correlation analysis (CCA) directions (Härdle and Simar, 2007), are used here to study population relationship with age. we also use distance weighted discrimination (DWD) direction (Marron et al., 2007; Qiao et al., 2010) , and *Direction-Projection-Permutation* (*DiProPerm*) test (Wichers et al., 2007), to explore population relationship with gender.

4.1.1 Roadmap

The organization of the rest of Chapter 4 is as follows. Section 4.2 described the tree data that we analyzed in this chapter. Section 4.2.1 presented several correspondence methods, i.e the *descendant correspondence*, to embed the brain artery trees into 2 dimension, and Section 4.2.2 discussed the *equivalence relation* and *equivalence class* idea to summarize the tree data to help the statistical analysis. Section 4.3 showed four tree representation methods, such as the combinational approach in Section 4.3.1, the phylogenetic trees in Section 4.3.2, the Dyck path representation in Section 4.3.3, and the branch length representation in Section 4.3.4. Section 4.4 discussed some existed analytic methods, such as the combinational approach in Section 4.4.1 and the phylogenetic methods in Section 4.4.2, and their analysis results. Section 4.5 showed the Dyck path analysis and details were presented in Section 4.5.1. Section 4.5 presented the pruned tree idea and showed the Dyck path and branch length analysis under the pruned trees structure in Section 4.6.1 and 4.6.2 respectively. The Section 4.7 gave out the conclusion and discussed some analysis results. The section 4.8 presented the some future work that we plan to finish in the future.

4.2 Data Description

A driving real data example is a set of blood artery trees. This data set is from a study of Magnetic Resonance Angiography (MRA) brain images (Dumoulin and Hart, 1986), of a set of 98 human subjects. Covariates of interest include both gender and age (from 19 to 79). The raw data can be found at Handle (2008). One slice of a 3 dimensional MRA image is shown in Panel (A) of Figure 4.1 and it highlights regions of blood flow as white. These white regions are aggregated in 3 dimensions by a tube tracking algorithm and then combined into brain artery trees, as in Aylward and Bullitt (2002). The result is shown in Panel (B) of Figure 4.1. These trees are colored according to the region of the brain (using color indicated in the caption in Figure 4.1) and are studied separately. The information stored in these color trees is very rich. Every color tree contains a set of branch segments which further contain a sequence of spheres, as calculated in Aylward and Bullitt (2002). Each sphere has

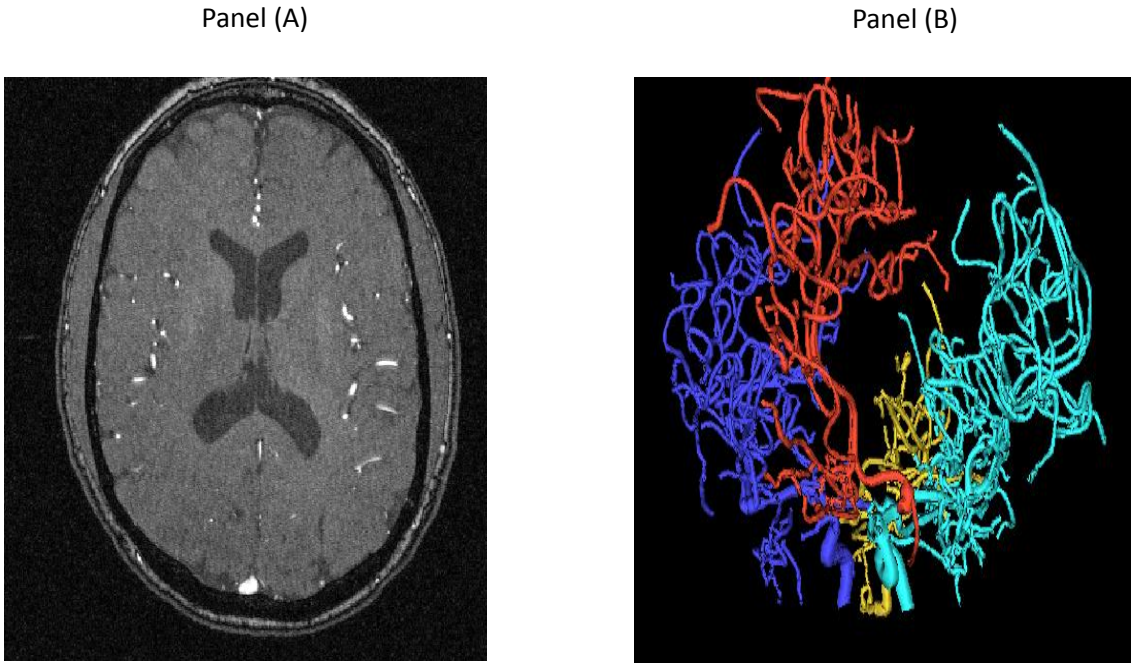

Figure 4.1: A single slice MRA image for one person

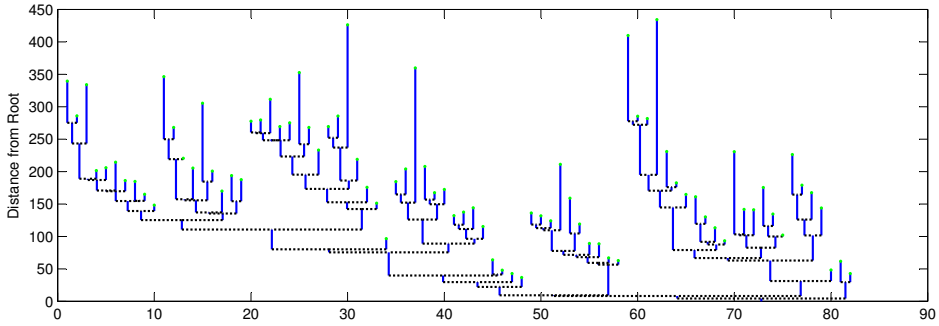

Figure 4.2: The binary tree from the back (gold) tree

a center with x, y, z coordinates, indicating the location of the center of the brain artery and its local radius , indicating the arterial thickness at that point. These spheres are shown in the visual rendering in Panel (B) of Figure 4.1.

4.2.1 Correspondence

In order to represent the topology of each tree, e.g. the back (gold) tree in Figure 4.1, several methods have been proposed in Aydin et al. (2009) to represent the 3 dimensional trees, in Figure 4.1, as simple binary trees embedded in 2 dimensions. This embedding is called *correspondence*, following standard terminology from image and shape analysis, see e.g. Dryden and Mardia (1998). The descendant correspondence method in Aydin et al. (2009) is used here to represent 3 dimensional trees, e.g. the back tree in Figure 4.1, as simple binary trees, e.g. as shown in Figure 4.2. Most tree branches are vessel segments between two neighboring splitting points of blood vessels. A leaf branch is the vessel segment between the end point of a blood vessel and its nearest splitting point. The arc length of each vessel segment, following the vessel curve, is defined to be the *branch length*. Every non-leaf branch of the trees in Figure 4.1 connects its children branches to its parent. The descendant correspondence method of 2 dimensional embedding puts the branch that has more descendant branches to the left, generating the binary tree in Figure 4.2. The blue line segments in Figure 4.2 represent tree branches and their length is the corresponding branch length. The root branch at the bottom of Figure 4.2 corresponds to the initial fat gold tree trunk shown near the bottom of Figure 4.1. We will focus on binary trees of this type and try to explore population structure of 98 such objects. An alternative representation is the *thickness correspondence* method of Aydin et al. (2009) that can also be used to represent a 3 dimensional tree as a simple binary tree. It might be interesting to study thickness correspondence in the future.

4.2.2 An Equivalence Relation

A more formal approach to correspondence in populations of tree structured objects is based on the mathematical notion of equivalence relation, see e.g. Wikipedia (2011b). Equivalence relations partition populations into equivalence classes, see e.g. Wikipedia (2011a). A very useful equivalence relation is defined through flipping. A flip can occur at each tree vertex, and is a switch of its children branches, as shown in Figure 4.3. Tree 1 is equivalent to Tree

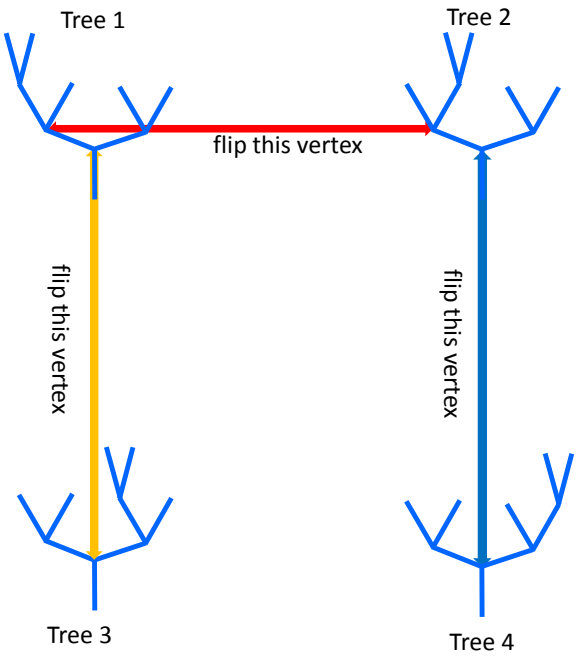

Figure 4.3: Equivalence from the flipping indicated by the colored arrows

2 and Tree 3 by flipping at the vertex indicated by the red and yellow arrows respectively, as shown in Figure 4.3. Tree 2 is equivalent to Tree 4 by flipping at the vertex indicated by blue arrow. These four binary trees are members of the same equivalence class. Flipping defines an equivalence relation and thus defines equivalence classes. After identifying the equivalence classes, we select the most representative tree for each class. For example, the descendent correspondence selects the tree, with more left branches than right, as the representative tree (e.g. Tree 1 in Figure 4.3). This makes every branch of the binary tree (e.g. Tree 1) have more left descendant branches than right. We define such a property as the *descendant correspondence property*. A possible approach to get the most representative model is through the concept of *invariant* as discussed in chapter 4 of Dryden and Mardia (1998).

4.3 Tree Representation

Statistical analysis, including PCA, has been very successful in curve data analysis, as seen in Ramsay and Silverman (2002) and Ramsay and Silverman (2005). An important reason behind this success is that curve space is a standard Euclidean space. However, tree space is strongly non-Euclidean space, as seen in Wang (2005) and the fundamental tools of functional data analysis, such as PCA, linear subspace, projection, are no longer available. Since PCA has been a broadly useful tool to explore variation in data, including curve data, it is very useful to propose analogues of PCA in tree space to explore the variation of tree populations. Since these fundamental tools for functional data analysis require data mathematical representation of the data, we first consider how to represent tree data. Several tree representation methods will be introduced in the following.

4.3.1 Combinatorics Approach

Wang and Marron (2007) developed a novel mathematical framework for the statistical analysis of tree objects and generalized PCA techniques (based on optimization problems) to tree objects to analyze the variation of the population tree structure. However, Wang and Marron (2007) solved the resulting optimization problems only for small toy examples and did not develop a general solution. Aydin et al. (2009) further analyzed these optimization problems and gave a general solution, based of the PCA like notion of *tree-lines*. Then, Aydin et al. (2009) applied these PCs to analyze the variation of a set of tree objects.

4.3.2 Phylogenetic Trees

Another way of representing trees is the phylogenetic tree approach, see Holmes (1999) for a good introduction. Phylogenetic trees aim to describe family relationships. Closeness in the phylogenetic tree corresponds to fairly recent common ancestors. Sisters in the phylogenetic tree defined by the same ancestor are called clades or monophyletic groups and are relatively similar. Finding such groups is one goal of a phylogenetic tree study. Holmes (1999); Li et al. (2000); Holmes (2003a) present several methods to analyze phylogenetic trees that may

be able to give some hints for the statistical analysis of population of blood vessel trees in Figure 4.1.

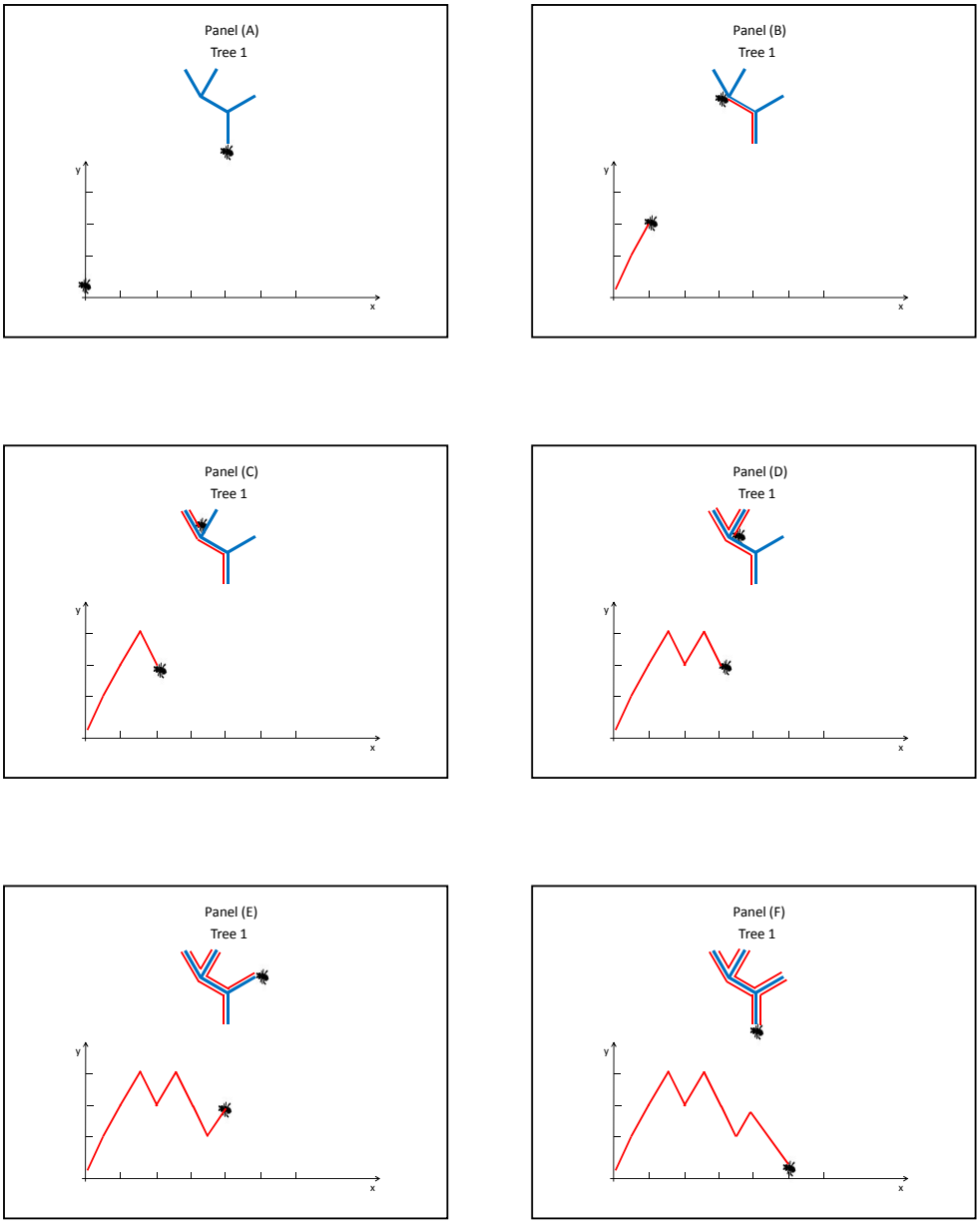

Figure 4.4: Dyck path of a tree

4.3.3 Dyck Path Representation

The main challenge for application of functional data analysis to tree populations in Figure 4.1 is the non-Euclidean property of tree space. This leaves fundamental concepts that underly functional data analysis, such as linear subspace, projection, and linear combination, unavailable. In contrast to the ideas of Wang and Marron (2007); Aydin et al. (2009), who developed a novel mathematical framework for statistical analysis of tree objects, we instead build a bridge between tree space and curve space and then exploit the power of functional data analysis to analyze populations of tree objects. We establish a one-to-one correspondence between binary trees and curves. The variation of these curves is studied using PCA. This gives an analog of PCA to explore the variation of tree objects. Furthermore, we find meaningful subspaces and project these tree curves into these subspaces to study interesting properties of trees, for example, the relationship between tree structure and gender or age. The Dyck path representation, see Harris (1952), is such a bridge that makes a tree uniquely correspond to a curve, as shown in Figure 4.4.

The Dyck path representation can be understood as follows. Assume that an ant starts from the root and walks around the individual tree following a sequence of time points as shown in Panel (A) to Panel (F) of Figure 4.4. The tree and the ant are shown in the top half of each panel, together with a piecewise line showing the ant's progress at the time represented by each panel. We record the number of time steps that the ant passed as the x-coordinate value and the branch height (i.e. distance to the root) as the y-coordinate value. The curve, connecting these coordinate points (x, y), is the Dyck path of the tree, shown in the bottom half of each panel. This section just shows the Dyck path of one binary tree. The extension of the Dyck path idea to a set of trees will be developed in Section 4.5.

4.3.4 Branch Length Representation

The Dyck path representation built a bridge between tree space and curve space and thus makes it possible to use functional data analysis to analyze tree structured data objects. The Dyck path representation recorded the branch height above the root that the ant passes, as

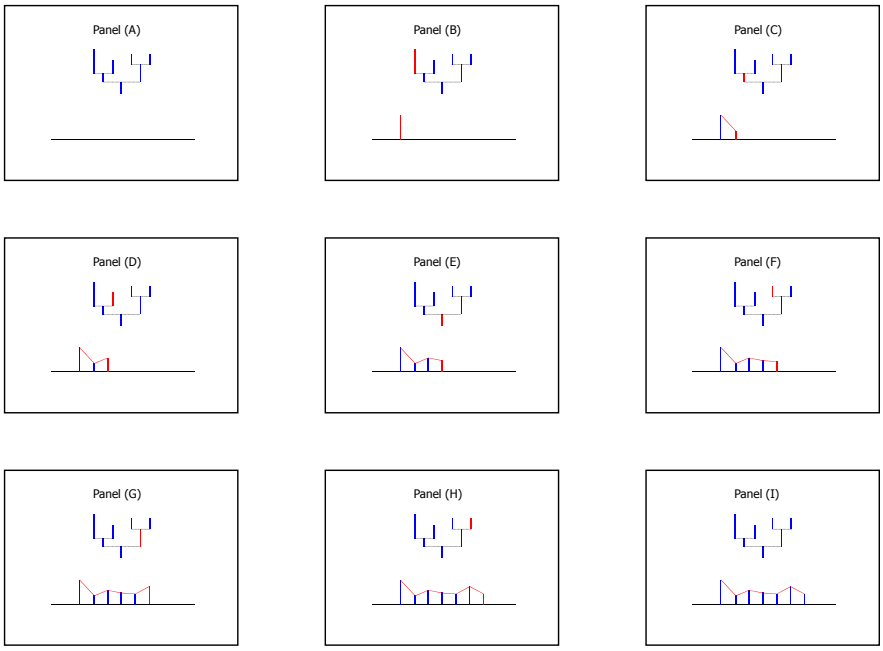

Figure 4.5: The branch length representation

shown in Figure 4.4. An alternative to recording the height of each branch is to record just the length of each branch. This generates an alternative bridge between tree space and curve space called the *branch length representation*.

This can be understood as follows. Assume that we start from the left most branch and record the branch lengths following a sequence of time points as shown in Panel (A) to Panel (I) of Figure 4.5. The tree and recorded (red) branch is shown in the top half of each panel, together with a piecewise line showing recording progress at the time represented by each panel. We record the number of time steps as the x-coordinate value and the branch length as the y-coordinate value. The piecewise linear curve, connecting these coordinate points (x, y), is the branch length representation of the tree. This section shows only the branch length representation of one binary tree. The extension of the branch length representation idea to a set of trees will be developed in Section 4.6.2.

4.4 Existing Data Analytic Methods

There are many different types of tree structured data objects, such as the brain artery trees in Figure 4.1, the phylogenetic trees in Holmes (1999); Li et al. (2000); Holmes (2003a), and the airway trees of Sonka et al. (1994). We will introduce some analysis methods and results on brain artery trees and phylogenetic tree analysis.

4.4.1 Combinational Analysis

The goal of exploring populations of tree structured data objects motivated Wang and Marron (2007) to develop a set of tree-population analogs of standard functional data analysis, such as PCA. The foundations were laid on some particular optimization problems and their solution generates the analysis method. One limitation of the work of Wang and Marron (2007) was the lack of a general solution for the optimization problem. Aydin et al. (2009) gave the complete solution of it, and developed a linear time computational method. Furthermore, Aydin et al. (2009) applied the developed method to 73 brain artery trees similar to that in Figure 4.1. That method provided some very interesting clinical findings. They found significant correlation between age and structure in the left sub-population tree in Figure 4.1.

4.4.2 Phylogenetic Methods

Holmes (1999) presents a large number of interesting statistical problems developed in the biological literature on estimating and evaluating phylogenetic trees. Holmes (1999) translates the problems in biology using statistical terms to help statisticians understand these problems. In addition, three main families of tree-building methods, such as maximum likelihood, distance-based methods, and maximum parsimony, are presented there. Holmes (1999) also outlined some relative strength and weakness of these methods. The bootstrap, described in Holmes (2003a), is a very popular method among biologists to evaluate a tree. Holmes (1999) discussed this method's feature and pointed out its drawback in phylogentic tree contexts. Furthermore, Holmes (2003b) uses distances and measures on a natural space of trees to represent relationship between phylogenetic trees in geometrical terms and suggests some

coherent ways to try to solve the problems of inference on the tree space.

4.5 Dyck Path Analysis

Until now, 98 brain artery trees of the type shown in Figure 4.1 have been transformed into

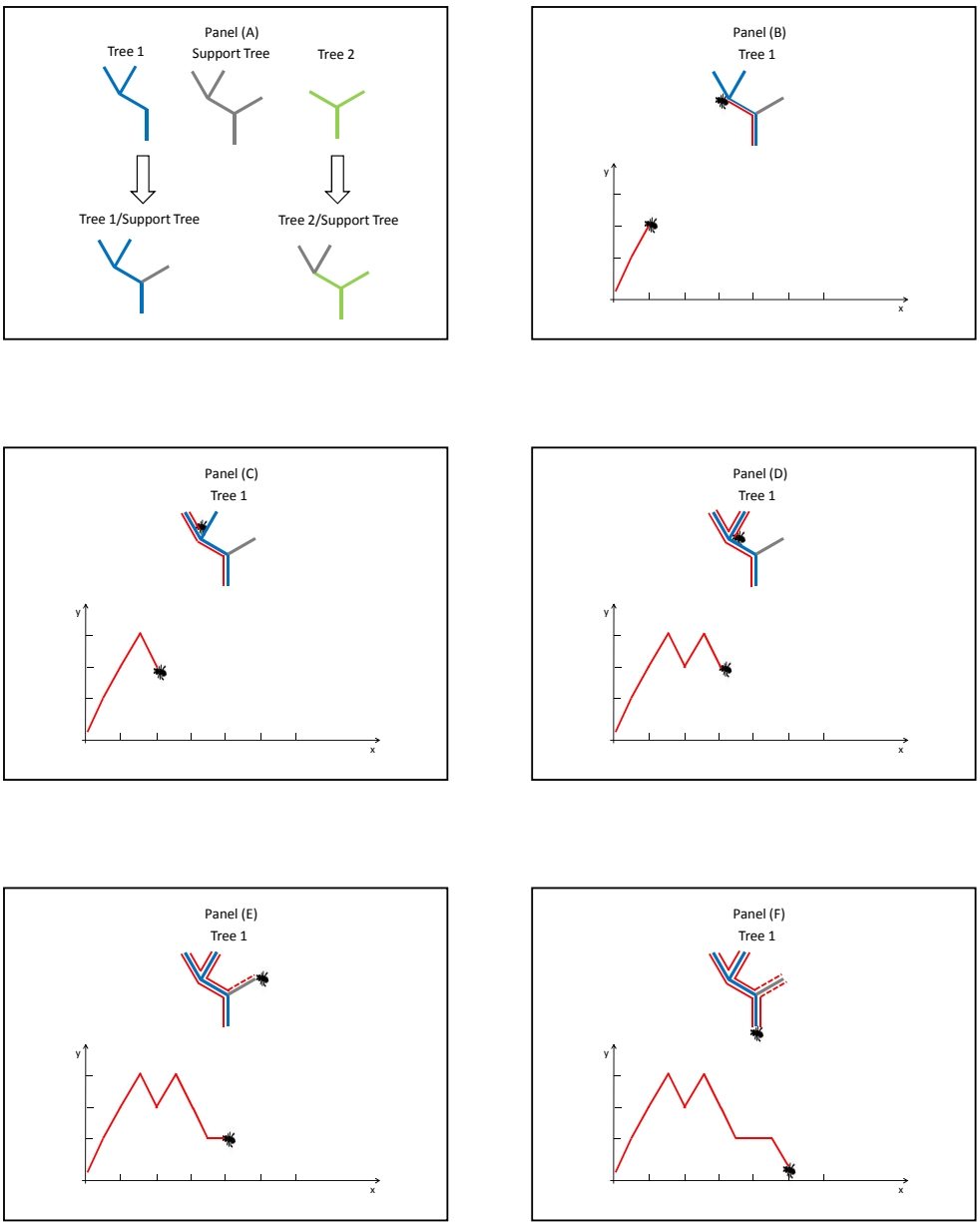

Figure 4.6: The support tree of individual

binary trees as shown in Figure 4.2. We want to explore the relationship between these binary

trees and to study population properties, such as gender and age. In order to compare the binary trees, we put them into a common structure context, called the *support tree*. The support tree is the union of individual trees' branches where branch length information is deliberately neglected. Panel (A) of Figure 4.6 is a toy example that illustrates how to define the support tree of a set of individual trees. We first neglect the branch length information of the individual trees, and then take the union of these trees' branches to get the support tree's branches. We set the branches of the support tree to have unit length. The support tree contains the total topological structure of the full set of individual trees. After finding the support tree, we can put each individual tree into a support tree structure context. The individual tree under the support structure context is the tree that keeps its original branches and sets the other branches of the support tree to be missing. This is illustrated in Panel (A) of Figure 4.6, where Tree 1, Tree 2 and the support tree are shown in the top half panel, and Tree 1 and Tree 2 under the support tree structure are shown in the bottom half panel, where the grey branches are missing branches. The length of these missing branches is taken to be zero. The rest of Figure 4.6 is discussed below.

The real individual binary tree, shown in Figure 4.2, is shown under the support tree structure in the top panel of Figure 4.7. The green flat parts are missing branches. To indicate variation in the population, two other trees are shown in the remaining panels. Each tree has many missing branches under the support tree structure and the missing branches' position is very different among different trees. This is a common feature of all 98 population trees.

Now, we will extend the method of the Dyck path from Section 4.3.3 to the case of individual binary trees under the support tree structure. The toy example in Figure 4.6 illustrates this process. The Dyck path of tree 1 is a piecewise curve, connecting the coordinate points (x, y) where the x-coordinate is the number of time steps that the ant walks around tree 1 and the y-coordinate is the corresponding branch height. Since the grey branch of tree 1 is missing, as shown in Panel (A) of Figure 4.6, the segment of the Dyck path corresponding to that branch is a flat curve as shown in Panel (E) and (F) of Figure 4.6. The collection of the Dyck paths for all 98 binary (back) trees is shown in Figure 4.8. The color of each the

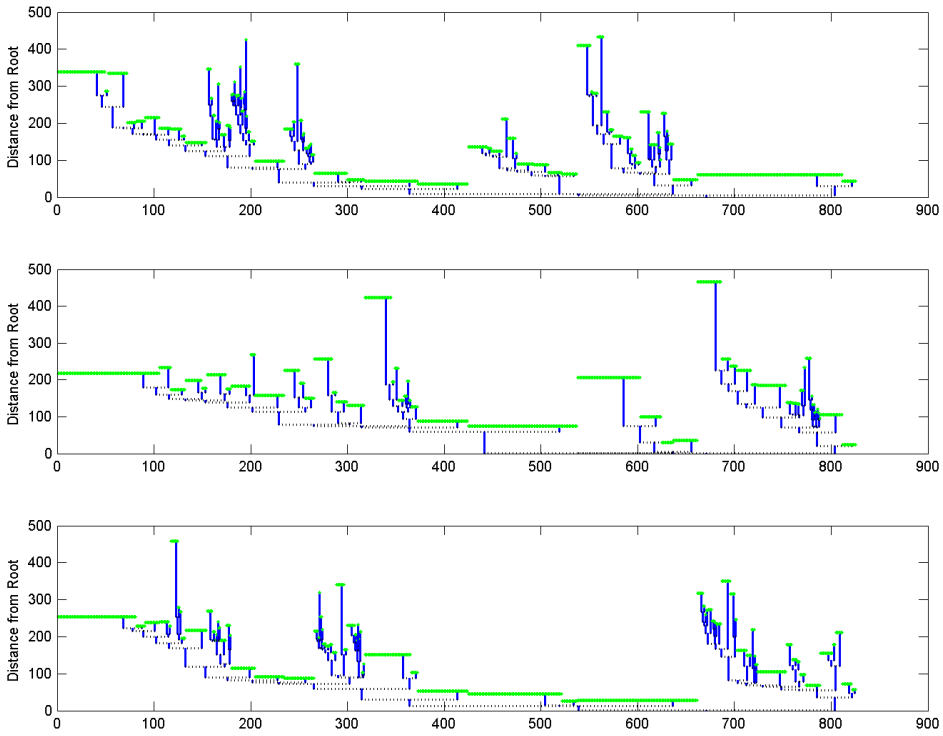

Figure 4.7: Three support individual binary (back) trees

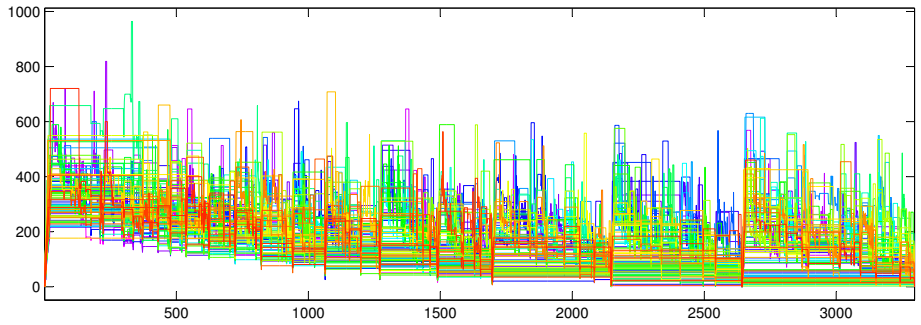

Figure 4.8: The Dyck path curves of the 98 support binary (back) trees

Dyck path corresponds to age, ranging from magenta (young) to red (old).

From the definition of the Dyck path, we found that every branch of the support tree should be passed twice, leading that the height of this branch appears twice in the Dyck path curves in Figure 4.8. The right bound of the x-coordinate range is twice of the number

of branches of the support tree. In addition, since there are many missing branches for the individual trees under the support tree structure context, as shown in Figure 4.7, there are many flat parts on these Dyck path curves, as shown in Figure 4.8. In addition, the population binary trees have some extreme long branches, shown in Figure 4.7, leading that the Dyck paths and their mean curve have jumps in Figure 4.8. The binary tree in Figure 4.2 is transformed from the back (gold) tree in Figure 4.1, using descendant correspondence. It makes the binary tree has more left descendant branches than right and such a property is kept by the binary trees under the support structure in Figure 4.7. It leads that the left part of the Dyck path curves in Figure 4.8 is taller than right.

4.5.1 Functional Data Analysis

Now, the functional data analysis can be applied to analyze this population of the Dyck path curves, representing back tree. First, we perform PCA of these Dyck path curves, as shown in Figure 4.9. The color of the curves corresponds to the age of the population, as in Figure 4.8. The plot in the first row and column in Figure 4.9 is the raw Dyck path curves, as in Figure 4.8. The plot in the first row and the second column in Figure 4.9 is the mean curve of the Dyck paths. Subtracting the mean from each Dyck path gives residual, i.e. the centered curves, shown in the first row and the third column plot. The bottom curve in the scree plot (Cattell, 1966; Jolliffe, 2002) in the first row and the fourth column shows empirical eigenvalues, and the top curve is the cumulative empirical eigenvalues. There is no clear low dimensional structure revealed by the scree plot. For these centered Dyck path curves, we get the PC1 direction through standard SVD (Jolliffe, 2002), and project these centered Dyck path curves on the PC1 direction to generate the plot in the second row and the first column in Figure 4.9. These PC1 projection curves, show that most of this dominant component of variation on the right side. This suggests that variation in the right part of trees drives the dominant component. We will further explore this variation in Figure 4.10. The second plot in the second row provides a different view of the projection in the first plot. This time they added to the sample mean and only the largest (i.e. component with maximal score, dashed) is shown, together with the smallest (dotted). Subtracting the PC1 projection data in the

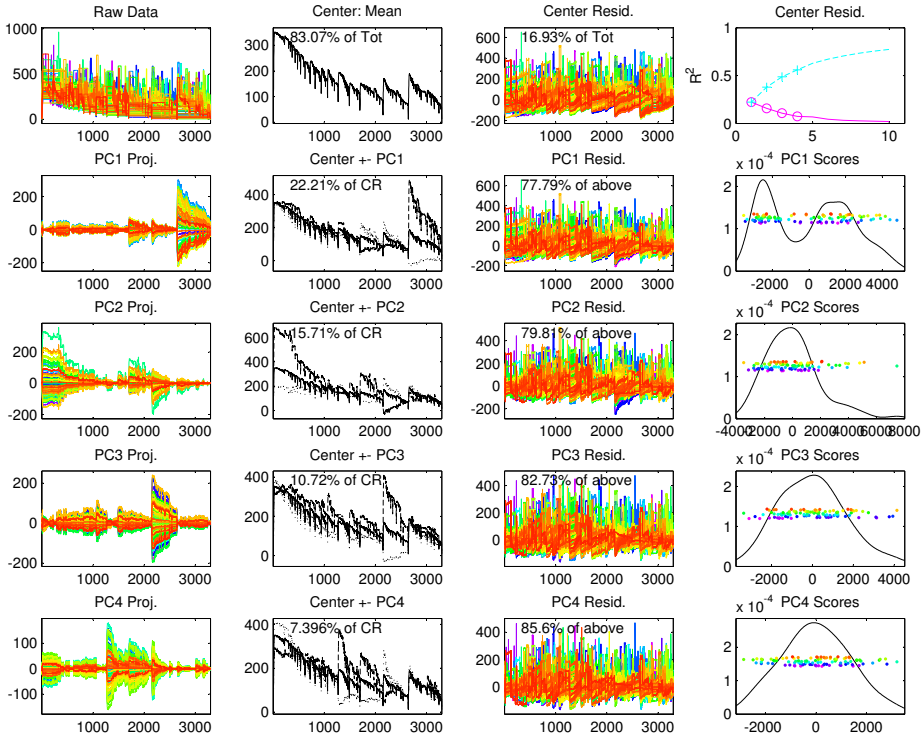

Figure 4.9: PCA of the Dyck path curves

first plot of the second row to the residual data in the third plot of the first row generates the PC1 residual curves in the third plot in the second row. Smooth histogram, i.e. kernel density distribution (Wand and Jones, 1995), of PC1 projection scores, the inner product between the centered Dyck path vectors and the PC1 direction vector, is shown in the fourth plot in the second row. The x-coordinate value corresponds to the projection score value and the y-coordinate is the jitter, corresponding to the age order (from young to old), to avoid overlapping of the points in the plot. The kernel distribution of PC1 scores seems a bimodal distribution.

If we replace the PC1 direction with the PC2, PC3 and PC4 directions, we get the third, fourth, and fifth row plots. The first plot in the third row shows that the main PC2 variation is from left to middle, disjoint with the PC1 main variation part. The kernel density distribution of the PC2 projection scores in the last plot in the third row seems no obvious

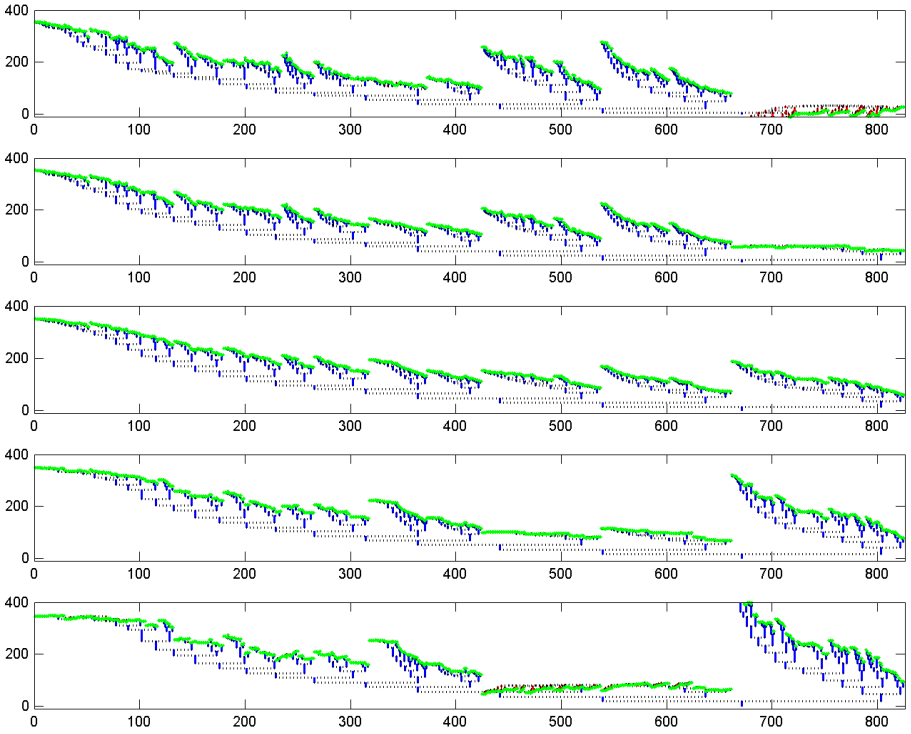

Figure 4.10: Five PC1 projection trees

bimodal distribution. The PC3 projection curve plot in the fourth row and the first column shows that the main variation part is between middle and right and seems disjoint with the PC1 main variation part. The kernel density distribution in the fourth row and column also have not obvious bimodal distribution. The PC4 projection curve plot in the last row and the first column shows that the variation is mainly due to the right and middle party. The kernel density distribution in the last row and column looks like the normal distribution.

In Figure 4.9, we explored the variation of the Dyck paths through functional PCA. Next we interpret this PC decompositional variation in terms of binary trees, as in Figure 4.10. These five binary trees from top to bottom correspond to the five PC1 projection points, which have -2, -1, 0, 1, and 2 standard deviation distance to the mean of the projection points respectively. Since the entries of the projection point vectors may have a negative value, it follows that some branches of the binary trees, indicated by the color red in Figure 4.10, have a negative length. This makes these trees invalid, i.e, this linear representation of the

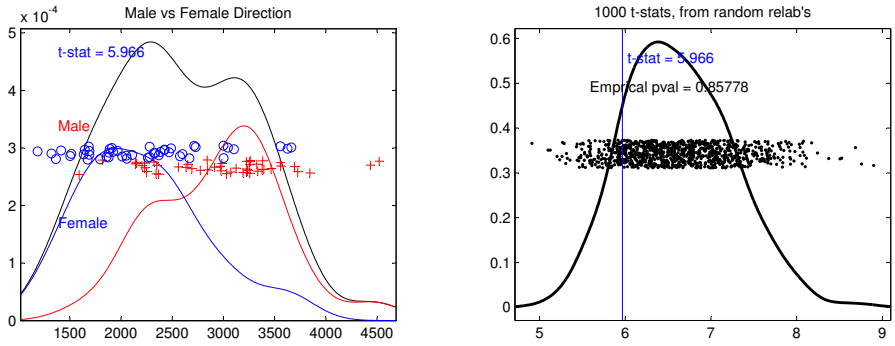

Figure 4.11: DiProPerm test on the DWD scores

data actually leaves tree space. In addition, these binary trees have essentially no flat parts (missing branches) which is very different from each population tree, as shown in Figure 4.7. Comparing these five trees, we found that the main difference is on the right part. Thus the most variation on the PC1 direction is due to branches in the right part of the binary trees. This is consistent with the result in the second row and the first column plot in Figure 4.9 that shows that the most variation is due to the right of the Dyck paths. We also did the same type of analysis on other PCs direction to study the variation of the binary trees. The results are also consistent with the results in the PC projection curve plot in the first column in Figure 4.9.

Besides exploring the variation of the binary trees, we are also interested in studying the relationship between tree structure and gender. Now we explore this relationship using the Dyck paths. Gender difference can be highlighted using two class discrimination methods. Distance weight discrimination (DWD) in Marron et al. (2007) and Qiao et al. (2010) is an efficient tool to study such problems, especially in high dimension. We project the Dyck paths on the DWD direction in the high dimensional space, and explore the projection scores in the following way. Similar with the functional PCA, we generate the kernel estimate of the projection scores in the left, panel (A) of Figure 4.11, along with the kernel estimate for the scores of the male, red curve, and the female, blue curve, respectively. In fact, the red and blue curves are rescaled kernel estimates of the male and female respectively and the rescaled

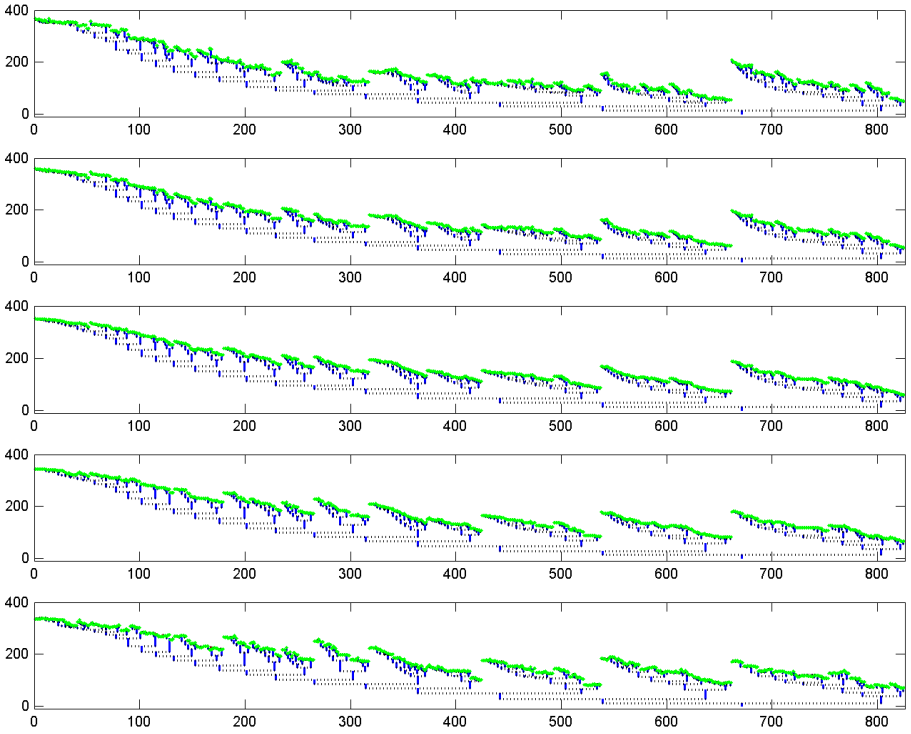

Figure 4.12: Five five DWD projection trees

coefficients are proportional to their population number. From these kernel density estimate curves, it seems that the male population has the bigger mean DWD scores than the female. The 2-sample t-test statistic of 5.97 for the average difference between the male and female DWD scores also suggests this might be statistically significant.

We did the Direction-Projection-Permutation (DiProPerm) test, described in Wichers et al. (2007), to rigorously explore this. The DiProPerm test uses a random permutation of the class labels to split the data into two groups, e.g the male and female group, and the DWD direction is recomputed for the relabeled data. For each random permutation, we calculated the 2-sample t-test statistic for the male and female DWD scores, and the statistic values are shown as the black dots in the left panel of Figure 4.11. The empirical p-value, which is essentially the percentage of random t-test statistic values greater than the real value, is shown in the left panel. The Gaussian fit p-value in the right panel is calculated based on the upper probability of the Gaussian distribution using the mean and standard deviation of

the simulated t-statistic. When the empirical p-value equal to 0, we need to use he Gaussian fit p-value to explore the statistical significance. The DiProPerm empirical p-value (0.398), shown in the right panel plot of Figure 4.11, implies that the DWD projection scores actually show no significant difference between male and female.

Similar as we did in Figure 4.10, we replace the PC1 direction with the DWD direction, to generate five binary trees in Figure 4.12. There does not appear to be a systematic visual difference among these five binary trees, which suggests that tree structure in the DWD direction is driven by artifacts of the sampling noise. This is consistent with the results implied by the empirical p-value of the DiProPerm test in Figure 4.11.

Furthermore, we have used PLS (with age) and CCA (with age) direction to generate similar binary trees to explore the age relationship, as we did in Figure 4.10 and 4.12. They are not shown here, because there were no meaningful results.

4.6 Pruned Trees

Dyck path and branch length representation build a bridge between tree space (non-Euclidean space) and curve space (Euclidean space), which makes it possible to apply functional data analysis to explore tree data. However, the binary trees under the support tree structure have a large number of missing branches, as in Figure 4.6, which brings many challenges to the statistical analysis. To address challenges, we present a novel idea, the *pruned tree*. For $n = 1, \cdots, 98$, the level n *pruned tree* is the union of the individual trees' branches, which appear in at least n individuals. The branch length information of the pruned tree is neglected and is set to be 1. The red tree in Figure 4.13 is the level 35 pruned tree and the whole tree is the support tree. In fact, the support tree is the level 1 pruned tree. For each pruning level, we defines *individual pruned trees*. Assume that each individual tree has been put into the support tree structure, as in Figure 4.7. Then, the individual pruned tree, e.g. the red tree in Figure 4.14, is the tree that keeps the branches of the level 35 global pruned tree, i.e. the red tree in Figure 4.13. The red tree in the top panel shown using only branches in the level 35 global pruned tree gives the red tree in the bottom panel. Note that many missing

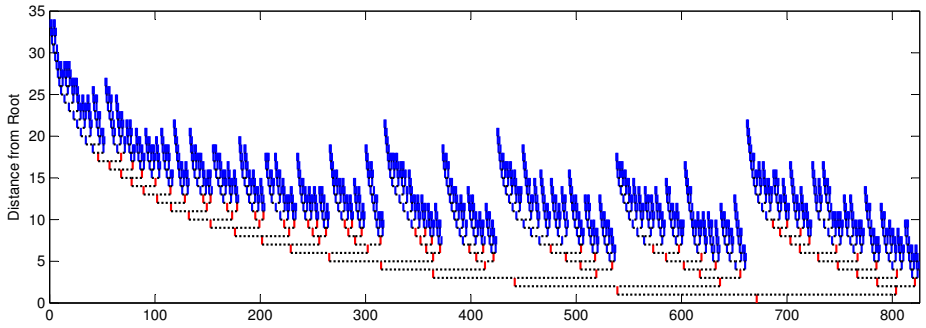

Figure 4.13: The 35 level pruned (Back) tree

branches in the top panel of Figure 4.14 disappear under the pruned tree structure.

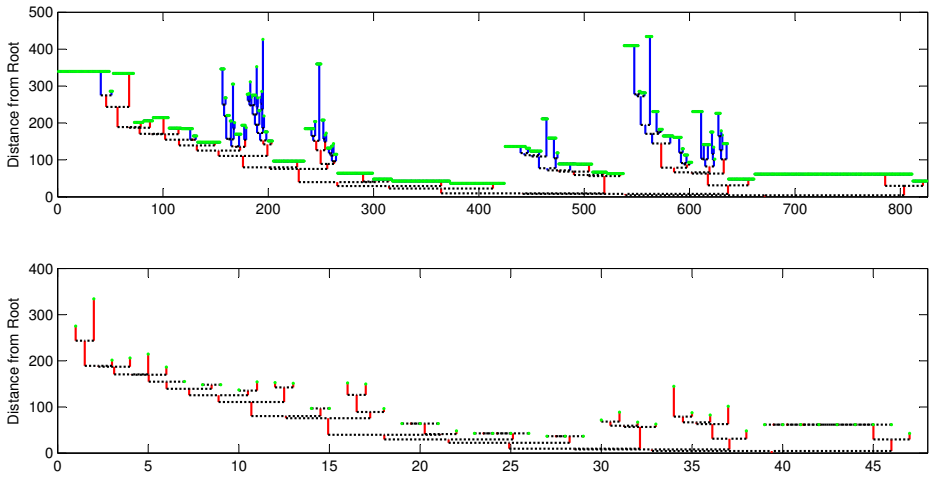

Figure 4.14: The individual (back) tree under the pruned tree structure.

The descendant correspondence method puts the branch that has more descendant branches to the left. This makes the binary tree in Figure 4.2 and 4.8 have the descendant correspondence property. We now show that individual pruned trees also have such a property. Since individual pruned trees have the same branches as the pruned tree, we just need to show that the level n pruned tree, e.g. the red tree in Figure 4.13, has the descendant correspondence property.

Theorem 4.6.1. *Each level n pruned tree keeps the descendant correspondence property.*

Proof. For every branch, we consider its first generation descendant branches. Assume that the first generation branch on the left is pruned out. Then the first generation right branch, having the same parent branch, is also pruned out. If not, some individual tree has more right branches than left. This is conflicts with the descendant correspondence property of the individual trees under support structure. Repeat this for the next generation branches until the top leaf branch. This finishes the proof of theorem 4.6.1. □

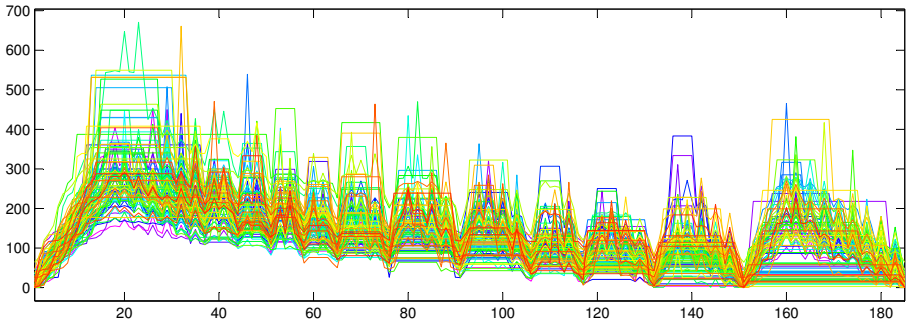

Figure 4.15: The Dyck paths of the level 35 pruned binary (back) trees

4.6.1 Dyck Path Analysis

The pruned tree structure decreases the number of missing branches of the binary trees, as in Figure 4.7, and can help improve the efficiency of the statistical analysis. As we did in Section 4.5, we first transform the individual trees under the different level pruned tree structure to the Dyck paths, as in Figure 4.15. The color of the Dyck paths corresponds to age, ranging from magenta (for young) to red (for old), and it is same as in Figure 4.8 and 4.9. The Dyck path curves under the pruned tree structure still have the flat parts because the individual pruned trees still have missing branches. The left part of the Dyck paths is taller than right because of the descendant correspondence property of the individual pruned trees.

As in Section 4.5.1, we use the DiProPerm test to study the relationship between the Dyck paths and gender. The kernel density estimates of the male, red curve, female, blue curve and the union of the DWD scores, black curve, are shown in the left panel of Figure 4.16.

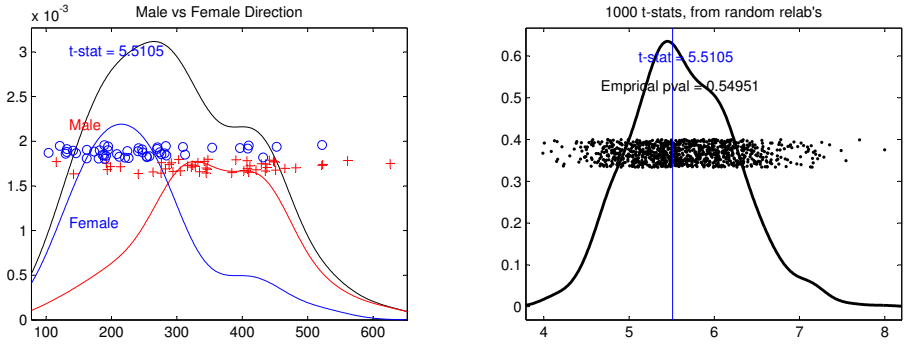

Figure 4.16: DiProPerm test on the DWD scores

It seems that the male population has the bigger mean DWD scores than the female from these kernel estimates. The 2-sample t-test statistic of 5.51 also suggests that the statistical significant difference of mean DWD scores exists between the male and female. Then, we used DiProPerm test to rigorously explore it. DiProPerm empirical p-value (0.55), shown in the right panel of Figure 4.11, implies that the DWD projection scores actually show no significant difference between male and female. Comparing the DiProPerm test in Figure 4.16 and 4.11, we found the result now to be slightly closer significant ($p = 0.55$) than before ($p = 0.86$) empirical p-value decreases a little.

This suggests that the pruned tree structure decreases the influence of the missing branches and slightly clarifies the gender difference. In addition, We just found that Dyck paths under pruning level 21 for right tree have significant difference in statistics (p-value is less than 0.05) between male and female.

4.6.2 Branch Length Analysis

The branch length representation in Section 4.3.4 is another bridge, connecting tree space and curve space, and will be applied to a set of individual pruned trees. Since the length of missing branches is set to be 0, the branch length representation curve of each individual pruned tree should go down to 0 when recording the missing branches, as in Figure 4.17. The plot in Figure 4.17 is the branch length curves of 98 individual trees under level 35 pruning. The color corresponds to age, ranging from magenta (for young) to red (for old), which is

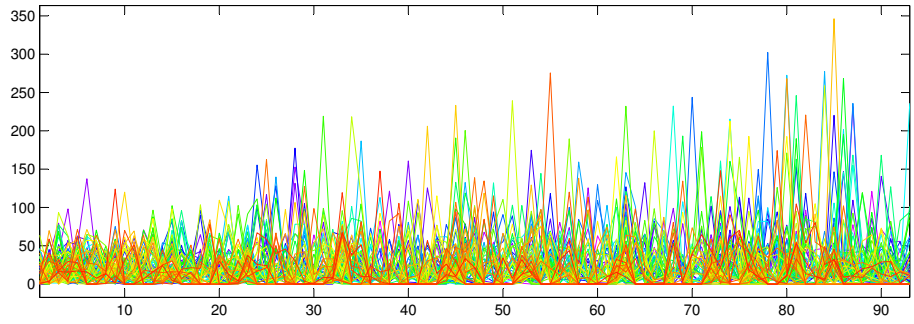

Figure 4.17: The branch length curves of the (level 35) pruned binary (back) trees

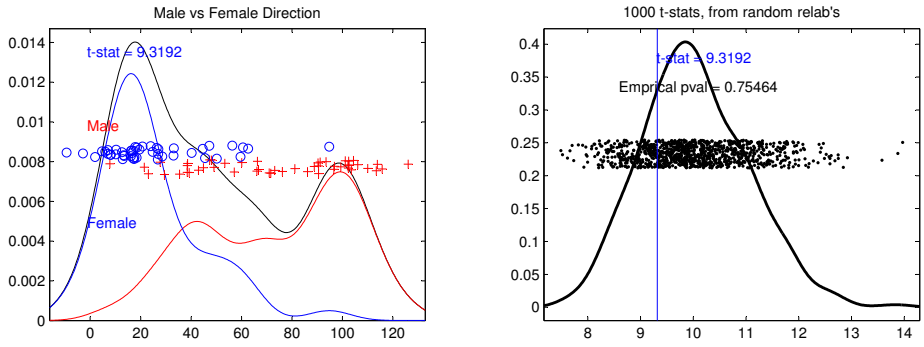

Figure 4.18: DiProPerm test on the DWD scores

the same as in Figure 4.8, 4.9 and 4.15. The $y = 0$ flat part of the branch length curves in Figure 4.17 is due to missing branches. The right bound of the x-coordinate range is the number of branches of the individual pruned trees. It is half of the right bound of the x-coordinate range in the Dyck paths in Figure 4.15 because every branch is passed twice in the Dyck paths. The y-coordinate is the branch length in these curves, as in Figure 4.17, in contrast to the branch height shown in the Dyck path curves, as in Figure 4.15. It follows that the top bound of y-coordinate range is smaller in Figure 4.17 than in Figure 4.15.

We also use the DiProPerm test to study the relationship between the branch length curves and gender. The left panel plot in Figure 4.18 is the kernel density estimate of the male, red curve, female, blue curve, and the union of the DWD scores, black curve, and it is the same as in Figures 4.11 and 4.16. The male population seems to have the bigger mean

DWD scores than the female The 2-sample t-test statistic of 9.32 also suggests that the DWD scores have significant difference in statistics between the male and female. However, the empirical p-value (0.75), shown in the right panel plot of Figure 4.11, implies that such a difference is not significant in statistics.

4.7 Conclusion and Discussion

Chapter 4 presented the Dyck path and branch length representation methods that build a bridge between tree space (non-Euclidean space) and curve space (Euclidean space). These two methods make it possible to apply the functional data analysis to tree data. Since the binary trees, as in Figure 4.7, have many missing branches that bring many challenges to the statistical analysis, this thesis presents a novel idea, the pruned tree, which helps to decrease the number of missing branches and increase the efficiency of the statistical analysis.

PCA helps us explore the variation of the Dyck paths, as in Figure 4.9. However, the curve PC view is hard to interpret Another interpretation view of the variation in each PC direction, in terms of binary trees, is shown in Figure 4.10. The same method is also applied to the DWD direction, as in Figure 4.12. The DiProPerm test of the Dyck path curves (right tree) under the level 21 pruned tree shows that the empirical p-value is smaller than 0.05. This suggests that there may exist significant differences between male and female on the DWD projection scores. We will further explore different pruning levels.

4.8 Future Work

There are several interesting problems that we will explore in the future. The first one is to compare the performance of the Dyck path and branch length representation under different pruning levels. In addition, we will study the relationship between the projection scores, based on all of PCs, PLS and CCA, and age.

The second one is to combine individual Front, Back (i.e the binary trees in Figure 4.7), Left and Right binary trees together, and then do Dyck path and branch length analysis. Finally, for individual pruned trees, we will calculate their branch number (the number of

all non-missing branches), total branch length (the total length of all branches), and average branch length (the ratio between the total branch length and the branch number). We will explore whether these three indices have some difference between male and female and their relationship with age under the different pruning levels.

Bibliography

Acker, A. (1974), "Absolute continuity of eigenvectors of time-varying operators," *Proceedings of the American Mathematical Society*, 42, 198–201.

Ahn, J., Lee, M. H., and Yoon, Y. J. (2012, forthcoming), "Clustering high dimension, low sample size data using the maximal data piling distance," *Statistica Sinica*.

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

Amini, A. and Wainwright, M. (2009), "High-dimensional analysis of semidefinite relaxations for sparse principal components," *Annals of Statistics*, 37, 2877–2921.

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

— (1984), *An introduction to multivariate statistical analysis*, John Willey & Sons, New York.

Aydin, B., Pataki, G., Wang, H., Bullitt, E., and Marron, J. (2009), "A Principal Component Analysis for Trees," *The Annals of Applied Statistics*, 3, 1597–1615.

Aylward, S. and Bullitt, E. (2002), "Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction," *IEEE Transactions on Medical Imaging*, 21, 61–75.

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

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

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

Bickel, P. and Levina, E. (2008a), "Covariance regularization by thresholding," *Annals of Statistics*, 36, 2577–2604.

— (2008b), "Regularized estimation of large covariance matrices," *Annals of Statistics*, 36, 199–227.

Bickel, P., Ritov, Y., and Tsybakov, A. (2009), "Simultaneous analysis of Lasso and Dantzig selector," *Annals of Statistics*, 37, 1705–1732.

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

Bosq, D. (2000), *Linear processes in function spaces: theory and applications*, vol. 149, Springer Verlag.

Bühlmann, P. (2006), "Boosting for high-dimensional linear models," *Annals of Statistics*, 34, 559–583.

Bunea, F., Tsybakov, A., Wegkamp, M., and Barbu, A. (2010), "Spades and mixture models," *Annals of Statistics*, 38, 2525–2558.

Candes, E. and Tao, T. (2007), "The Dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, 35, 2313–2351.

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

Cattell, R. (1966), "The scree test for the number of factors," *Multivariate Behavioral Research*, 1, 245–276.

d'Aspremont, A., El Ghaoui, L., Jordan, M., and Lanckriet, G. (2007), "A direct formulation for sparse PCA using semidefinite programming," *SIAM Review*, 49, 434–448.

Dauxois, J., Pousse, A., and Romain, Y. (1982), "Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference," *Journal of Multivariate Analysis*, 12, 136–154.

Donoho, D. and Johnstone, J. (1994), "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, 81, 425–455.

Dryden, I. and Mardia, K. (1998), *Statistical Shape Analysis*, vol. 4, Wiley New York.

Dumoulin, C. and Hart, H. (1986), "Magnetic Resonance Angiography." *Radiology*, 161, 717–720.

Eckart, C. and Young, G. (1936), "The approximation of one matrix by another of lower rank," *Psychometrika*, 1, 211–218.

El Karoui, N. (2008), "Operator norm consistent estimation of large-dimensional sparse covariance matrices," *Annals of Statistics*, 36, 2717–2756.

Fan, J. and Li, R. (2001), "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, 96, 1348–1360.

Geman, S. (1980), "A limit theorem for the norm of random matrices," *The Annals of Probability*, 8, 252–261.

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

Hall, P. and Hosseini-Nasab, M. (2006), "On properties of functional principal components analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 109–126.

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

Handle (2008), *Tree Data, "http://hdl.handle.net/1926/594"*.

Härdle, W. and Simar, L. (2007), *Applied Multivariate Statistical Analysis*, Springer Verlag.

Harris, T. (1952), "First passage and recurrence distributions," *Transactions of the American Mathematical Society*, 73, 471–486.

Holmes, S. (1999), "Phylogenies: an overview," *Statistics and Genetics*, 112, 81–118.

— (2003a), "Bootstrapping phylogenetic trees: theory and methods," *Statistical Science*, 18, 241–255.

— (2003b), "Statistics for phylogenetic trees," *Theoretical Population Biology*, 63, 17–32.

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

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

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

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

Jolliffe, I. (2002), *Principal Component Analysis*, Springer Verlag.

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

Jung, S., Sen, A., and Marron, J. (2012), "Boundary behavior in high dimension, low sample size asymptotics of PCA," *Technical Report, UNC-CH.*

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

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983), *Extremes and related properties of random sequences and processes*, Springer-Verlag.

Lee, M., Shen, H., Huang, J., and Marron, J. (2010a), "Biclustering via sparse singular value decomposition," *Biometrics*, 66, 1087–1095.

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

Lee, Y., Lee, E., and Park, B. (2012, forthcoming), "Principal component analysis in very high-dimensional spaces," *Statistica Sinica*.

Leng, C. and Wang, H. (2009), "On general adaptive sparse principal component analysis," *Journal of Computational and Graphical Statistics*, 18, 201–215.

Li, S., Pearl, D., and Doss, H. (2000), "Phylogenetic Tree Construction Using Markov chain Monte Carlo," *Journal of the American Statistical Association*, 95, 493–508.

Ma, Z. (2012), "Sparse principal component analysis and iterative thresholding," *Technical Report, UPenn*.

Marron, J., Todd, M., and Ahn, J. (2007), "Distance-weighted discrimination," *Journal of the American Statistical Association*, 102, 1267–1271.

Meier, L., van de Geer, S., and Bühlmann, P. (2009), "High-dimensional additive modeling," *Annals of Statistics*, 37, 3779–3821.

Meinshausen, N. and Bühlmann, P. (2006), "High-dimensional graphs and variable selection with the lasso," *Annals of Statistics*, 34, 1436–1462.

Muirhead, R. (1982), *Aspects of multivariate statistical theory*, John Willey & Sons, New York.

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

Obozinski, G., Wainwright, M., and Jordan, M. (2011), "Support union recovery in high-dimensional multivariate regression," *Annals of Statistics*, 39, 1–47.

Omidiran, D. and Wainwright, M. (2010), "High-dimensional variable selection with sparse random projections: measurement sparsity and statistical efficiency," *Journal of Machine Learning Research*, 99, 2361–2386.

Onatski, A. (2006), "Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak," *Manuscript, Columbia University*.

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

Paul, D. and Johnstone, I. (2007a), "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Technical Report, UC Davis*.

— (2007b), "Augmented sparse principal component analysis for high-dimensional data," *Technical Report, UC Davis*.

Qiao, X., Zhang, H., Liu, Y., Todd, M., and Marron, J. (2010), "Weighted Distance Weighted Discrimination and Its Asymptotic Properties," *Journal of the American Statistical Association*, 105, 401–414.

Ramsay, J. and Silverman, B. (2002), *Applied Functional Data Analysis: Methods and Case Studies*, Springer Verlag.

— (2005), *Functional Data Analysis*, Springer Verlag.

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

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

Schwarz, G. (1978), "Estimating the dimension of a model," *Annals of Statistics*, 6, 461–464.

Shen, D., Shen, H., and Marron, J. (2012a), "Consistency of sparse PCA in high dimension and low sample size contexts," *Submitted to Journal of Multivariate Analysis*.

Shen, D., Shen, H., and Marron, J. S. (2012b), "Consistency of sparse PCA in high dimension, low sample size contexts: additional simulation results," *Available online at http://www.unc.edu/∼dshen/RSPCA/SupplementB.html*.

Shen, H. and Huang, J. (2008), "Sparse principal component analysis via regularized low rank matrix approximation," *Journal of Multivariate Analysis*, 99, 1015–1034.

Silverstein, J. (1985), "The smallest eigenvalue of a large dimensional Wishart matrix," *The Annals of Probability*, 13, 1364–1368.

Sonka, M., Sundaramoorthy, G., and Hoffman, E. (1994), "Knowledge-based segmentation of intrathoracic airways from multidimensional high resolution CT images," *Medical Imaging*, 2168, 73–85.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, 58, 267–288.

van de Geer, S. (2008), "High-dimensional generalized linear models and the lasso," *Annals of Statistics*, 36, 614–645.

Wand, M. and Jones, M. (1995), *Kernel smoothing*, Chapman & Hall/CRC.

Wang, H. (2005), "Object Oriented Data Analysis: Sets of Trees, available at "`http://www.stat.colostate.edu/~wanghn/Tree_files/OODAtree2.pdf`"," .

Wang, H. and Marron, J. (2007), "Object oriented data analysis: Sets of trees," *Annals of Statistics*, 35, 1849–1873.

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

Wichers, L., Lee, C., Costa, D., Watkinson, W., and Marron, J. (2007), "A functional data analysis approach for evaluation temporal physiologic responses to particualate matter," *Tech. Rep. 5, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill.*

Wikipedia (2011a), *Equivalence Class, Available at "`http://en.wikipedia.org/wiki/Equivalence_class`".*

— (2011b), *Equivalence Relation, Available at "http://en.wikipedia.org/wiki/Equivalence_relation"*.

Witten, D., Tibshirani, R., and Hastie, T. (2009), "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, 10, 515–534.

Wold, S., Ruhe, A., Wold, H., and Dunn III, W. (1984), "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses," *SIAM Journal on Scientific and Statistical Computing*, 5, 735–743.

Yata, K. and Aoshima, M. (2012), "Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations," *Journal of Multivariate Analysis*, 105, 193–215.

Zou, H., Hastie, T., and Tibshirani, R. (2006), "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, 15, 265–286.

— (2007), "On the "degrees of freedom" of the lasso," *Annals of Statistics*, 35, 2173–2192.