Statistical Papers (2025) 66:63
https://doi.org/10.1007/s00362-025-01680-3

REGULAR ARTICLE

Spectral Clustering Algorithm for the Allometric Extension Model

Kohei Kawamoto[1] · **Yuichi Goto**[2] · **Koji Tsukuda**[2]

Received: 6 October 2023 / Revised: 19 December 2024
© The Author(s) 2025

Abstract
The spectral clustering algorithm is often used as a binary clustering method for unclassified data by applying the principal component analysis. When investigating the theoretical properties of the spectral clustering algorithm, existing studies have tended to invoke the assumption of conditional homoscedasticity. However, this assumption is restrictive and, in practice, often unrealistic. Therefore, in this paper, we consider the allometric extension model in which the directions of the first eigenvectors of two covariance matrices and the direction of the difference of two mean vectors coincide. We derive a non-asymptotic bound for the error probability of the spectral clustering algorithm under this allometric extension model. As a byproduct of this result, we demonstrate that the clustering method is consistent in high-dimensional settings.

Keywords High-dimension · Principal component analysis · Non-asymptotic bound

1 Introduction

1.1 Spectral Clustering Algorithm

The clustering of unclassified data is a typical problem in multivariate data analysis, and several clustering methods have been developed. Commonly used methods include the k-means clustering (Pollard 1981, 1982), the hierarchical clustering (Borysov et al. 2014), and the clustering based on linear discriminant functions (O'Neill 1978); for a review of clustering methods, see Amit et al. (2017) and references therein. Of these

✉ Kohei Kawamoto
kawamoto.kohei.532@s.kyushu-u.ac.jp

Yuichi Goto
yuichi.goto@math.kyushu-u.ac.jp

Koji Tsukuda
tsukuda@math.kyushu-u.ac.jp

[1] Joint Graduate School of Mathematics for Innovation, Kyushu University, 744 Motooka, Fukuoka, Fukuoka 819-0395, Japan

[2] Faculty of Mathematics, Kyushu University, 744 Motooka, Fukuoka, Fukuoka 819-0395, Japan

Published online: 17 March 2025

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

methods, the spectral clustering algorithm based on principal component analysis is popular because of its low computational load. Spectral clustering algorithm has various applications such as community detection and graph partitioning; see Löffler et al. (2021) for references and further applications. According to Sect. 4.7.1 of Vershynin (2018), the spectral clustering algorithm proceeds as follows. Let the dimension of the data and the sample size be n and m, respectively, and denote n-dimensional (centered) data points as $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$. Calculate the realized sample covariance matrix

$$\boldsymbol{s}_m = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i^\top$$

and the unit-length eigenvector $\boldsymbol{v}$ corresponding to the largest eigenvalue of $\boldsymbol{s}_m$, where $\top$ denotes the transpose of a vector. To classify the data points into two groups, we project them onto the space spanned by $\boldsymbol{v}$ and classify them according to the signs of their principal component scores. To review the theoretical properties of the spectral clustering algorithm, we define θ, indicating the group to which an individual belongs, as a symmetric Bernoulli variable (Rademacher variable), that is, $P(\theta = 1) = P(\theta = -1) = 1/2$, and let $\boldsymbol{g}^{(1)}$ and $\boldsymbol{g}^{(-1)}$ be n-dimensional centered normal variables with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively. Define an n-dimensional random variable $\boldsymbol{X}$ as

$$\boldsymbol{X} = \theta \boldsymbol{\mu} + \boldsymbol{g}^{(\theta)},$$

where $\boldsymbol{\mu}$ is an n-dimensional vector. Note that $E[\boldsymbol{X}] = \boldsymbol{0}_n$, where $\boldsymbol{0}_n$ is an n-dimensional zero vector. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be independent and identically distributed (iid) copies of $\boldsymbol{X}$, let

$$\boldsymbol{S}_m = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{X}_i \boldsymbol{X}_i^\top \tag{1}$$

be the sample covariance matrix, and let $\boldsymbol{\gamma}_1(\boldsymbol{S}_m)$ be the unit eigenvector corresponding to the largest eigenvalue of $\boldsymbol{S}_m$. Under the assumption that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the identity matrix of size n, Vershynin (2018, Sect. 4.7) evaluated

$$P\left(\{\text{The number of misclassifications of } X_1, \ldots, X_m\} \leq \varepsilon m\right), \tag{2}$$

where ε is a positive constant satisfying appropriate conditions. We will formulate "the number of misclassifications" in (12). Moreover, under the assumption that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{I}_n$ and $\|\boldsymbol{\mu}\|_2 \geq C_{\text{gap}} n/m$ for some large constant C_{gap}, Cai and Zhang (2018) derived an upper bound for the expectation of the misclustering rate that is a quantity of interest for clustering methods; "the misclustering rate" will be explained in (13). In this paper, we use the words "misclassification" and "misclustering" in different senses. A clustering method based on the moment method for a mixture distribution of spherical normal distributions was proposed by Hsu and Kakade (2013), while an improved method using the singular value decomposition of the Gram matrix

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

calculated with its diagonal components replaced by zero was developed by Abbe et al. (2022). Recently, for a derivative algorithm of Lloyd's iterative procedure in a two-component mixture of normal distributions, Ndaoud (2022) derived conditions under which asymptotically correct results are obtained. When $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, the first eigenvector of the covariance matrix $\boldsymbol{\Sigma}$ of the mixture distribution is generally different from that of $\boldsymbol{\Sigma}_1$/ However, if $\boldsymbol{\Sigma}_1$ is spherical, the direction of the first eigenvector of $\boldsymbol{\Sigma}$ is parallel to $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}_1 = \boldsymbol{\mu}$ and $\boldsymbol{\mu}_2 = -\boldsymbol{\mu}$. This serves as the basis for the spectral clustering algorithm. However, the assumptions that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_1$ is spherical are sometimes restrictive and unlikely to hold in practice. Section 5 of Abbe et al. (2022) evaluated eigenvalues and eigenvectors in heteroscedastic situations for each individual. Therefore, in this paper, we consider the weaker assumption that $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ follow the allometric extension relationship formulated by Flury (1997) and Bartoletti et al. (1999), that is, the leading eigenvector of $\boldsymbol{\Sigma}_1$ is parallel to that of $\boldsymbol{\Sigma}_2$ and is also parallel to $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. We then evaluate the misclassification probability of the spectral clustering algorithm. This result allows us to demonstrate the consistency of the spectral clustering algorithm in a high-dimensional setting.

According to Myers et al. (2020a), species delimitation has been studied based on morphological and molecular phylogenetics approaches, but the species classified through molecular phylogenetics conflict with those classified by phenotypic variation. For anole lizards, in particular, species lacking obvious visual differences classified by molecular phylogenetics present challenges in differentiation based on morphometric traits. This study explores the extent to which physical size differences caused by genetic or species variations affect successful spectral clustering. Recently, advances in measuring instruments have led to higher-dimensional data being obtained for living organisms. Therefore, this study focuses on the clustering in high-dimensional settings.

1.2 Notation

In this subsection, we introduce the notation used in this paper. For vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ that have the same dimension, $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^\top \boldsymbol{b}$ denotes the inner product of $\boldsymbol{a}$ and $\boldsymbol{b}$. For a vector $\boldsymbol{a}$, let $\|\boldsymbol{a}\|_2 = \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}$ be the ℓ^2-norm of $\boldsymbol{a}$. For a square matrix $\boldsymbol{A}$, the k-th largest eigenvalue is $\lambda_k(\boldsymbol{A})$, the corresponding eigenvector is $\boldsymbol{\gamma}_k(\boldsymbol{A})$, and the operator norm is $\|\boldsymbol{A}\|_{\mathrm{op}}$. For a random vector $\boldsymbol{X}$, $V[\boldsymbol{X}]$ denotes the covariance matrix of $\boldsymbol{X}$. For a positive integer n, $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the n-dimensional normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. Moreover, in general, $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. For a positive-definite covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{1/2}$ denotes the symmetric matrix satisfying $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$, and $\boldsymbol{\Sigma}^{-1/2}$ denotes the inverse of $\boldsymbol{\Sigma}^{1/2}$. For a random variable X, the L^2 norm $\left(E[X^2]\right)^{1/2}$ and the sub-Gaussian norm $\inf\{t > 0 : E[\exp(X^2/t^2)] \leq 2\}$ of X are denoted by $\|X\|_{L^2}$ and $\|X\|_{\psi_2}$, respectively. For a set A, #(A) denotes the number of elements contained in A.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

1.3 Organization of this Paper

In Sect. 2, the allometric extension model is introduced and some properties of this model are derived. An example of the application of the spectral clustering algorithm to real data is given in Sect. 3. Section 4 presents the main results of this paper. The proofs of these theoretical results are provided in Sect. 5. Finally, some concluding remarks are presented in Sect. 6.

2 Allometric Extension Model

In this section, we introduce the allometric extension model. Let n be a positive integer, and let $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ be n-dimensional vectors satisfying $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. Let $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ be $n \times n$ positive- definite symmetric matrices such that $\lambda_1(\boldsymbol{\Sigma}_i) > \lambda_2(\boldsymbol{\Sigma}_i)$ for $i = 1, 2$. We define the allometric extension relationship between two distributions $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ as follows: there exists some $\beta \in \mathbb{R}$ such that

$$\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1) = \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_2) = \beta(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \tag{3}$$

where the sign of $\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_2)$ is suitably chosen. Throughout the discussion, we assume that the two distributions are n-dimensional normal distributions. The allometric extension model formulated by Flury (1997) and Bartoletti et al. (1999) is used to express a typical relationship between two or more biotic groups. Kurata et al. (2008) also considered multi-group settings. For this model, Bartoletti et al. (1999) proposed a test procedure for determining whether two groups have an allometric extension relationship and analyzed carapace size data from turtles of different sexes, as discussed by Jolicoeur and Mosimann; for this data that will be analyzed in Sect. 3, we refer to Table 1.4 in Flury (1997). Moreover, Tsukuda and Matsuura (2023) proposed a test procedure for the allometric extension relationship when the observations consist of high-dimensional data. The properties of the mixtures of two or more distributions that form the allometric extension relationship are discussed in Flury (1997, Sect. 8.7), Kurata et al. (2008), and Matsuura and Kurata (2014). Note that Tarpey (2007) mentions the case of applying the k-means algorithm to the allometric extension model. We focus on the spectral clustering algorithm, which achieves greater computational efficiency than other traditional clustering algorithms through its dimensionality-reduction techniques. Let us derive some properties associated with the mixture distribution of two normal distributions forming the allometric extension relationship. Let n be a positive integer, and let $f_1(\cdot)$ and $f_2(\cdot)$ be the probability density functions of $\mathcal{N}_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}_n(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, respectively. We assume that $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ satisfy (3) and that the n-dimensional random variable $\boldsymbol{X}$ follows a mixture distribution whose probability density function is given by

$$f_X(\boldsymbol{x}) = \pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x}) \quad (\boldsymbol{x} \in \mathbb{R}^n),$$

where π_1 and π_2 are positive values satisfying $\pi_1 + \pi_2 = 1$. In this case, the following properties about the covariance matrix $\boldsymbol{\Sigma} = V[\boldsymbol{X}]$ hold.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 1 Inner products between the first PCs for male and female turtles in the turtle data and the unit vector of the mean difference between male and female turtles

Vectors of interest	Inner product
First PCs for male and female turtles	0.9998
First PC for female and mean difference between male and female turtles	0.9994
First PC for male and mean difference between male and female turtles	0.9987

Proposition 2.1 *Under the conditions stated above,*

$$\text{(i)} \quad \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}) = \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \tag{4}$$

$$\text{(ii)} \quad \lambda_1(\boldsymbol{\Sigma}) = \pi_1\lambda_1(\boldsymbol{\Sigma}_1) + \pi_2\lambda_1(\boldsymbol{\Sigma}_2) + \pi_1\pi_2\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2, \tag{5}$$

$$\text{(iii)} \quad \lambda_2(\boldsymbol{\Sigma}) \leq \pi_1\lambda_2(\boldsymbol{\Sigma}_1) + \pi_2\lambda_2(\boldsymbol{\Sigma}_2), \tag{6}$$

where the sign of $\boldsymbol{\gamma}_1(\boldsymbol{\Sigma})$ *is appropriately chosen in* (4).

Remark 1 Results (4) and (5) are given in Lemma 8.7.1 of Flury (1997), but (6) provides a better evaluation than the corresponding result in Flury (1997). Therefore, we include the proof of Proposition 2.1, although the procedure is similar.

Remark 2 In this study, we consider the situation in which $\lambda_1(\boldsymbol{\Sigma}_i) > \lambda_2(\boldsymbol{\Sigma}_i)$ for $i = 1, 2$. Therefore, we stress that the case where $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{I}_n$ is not included in our setting.

Remark 3 Although the allometric extension model has been studied for multi-group, we focus on two groups settings for simplicity.

3 Real Data Application of the Spectral Clustering Algorithm

This section demonstrates the superiority of the spectral clustering algorithm in two real data examples. The first contains the two-dimensional turtle carapace data (width and length of turtle shells) in Flury (1997), in which the number of individuals is 48 (24 males and 24 females). The second contains the 13-dimensional lizard data (snout-vent length, axilla-groin length, head width, head length, snout length, head depth, longitudinal ear opening, vertical length of ear opening, interorbital distance, shank length, femoral length, total foot length, fourth toe length) in Myers et al. (2020b), in which the number of individuals is 189, consisting of 100 genotype E (which includes "A. d. dominicensis – Central Hispaniola" and "A. d. ravitergum") and 89 genotype F ("A. d. dominicensis – Northern Hispaniola"). Those datasets can be found in Flury (1997) and Myers et al. (2020b). Figure 1 illustrates the scatter plot of male and female turtles data represented by white circles and black squares. We observe that all data points seem to be on a straight line. Figures 2 and 3 show the projections of the data onto the first and second principal components (PCs). Apparently data points from different groups are scattered along with the first PC direction. Tables 1 and 2

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

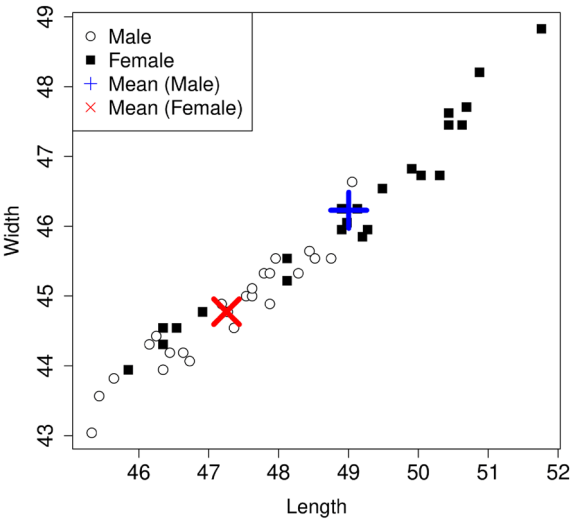

Fig. 1 The scatter plot of the turtle data (Flury 1997). The x- and y- axes correspond to width and length of turtle carapaces, respectively. White circles, black squares, blue plus sign, and red cross mark correspond to male data points, female data points, the mean of male data points, and the mean of female data points, respectively

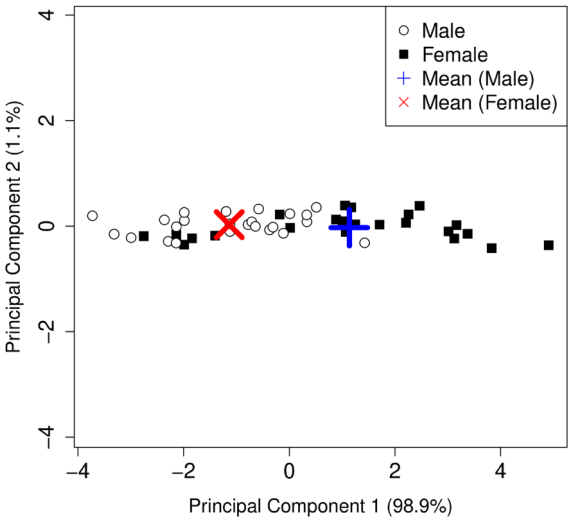

Fig. 2 Projection of the turtle data (Flury 1997) onto the space spanned by first and second principal component vectors. The x- and y- axes correspond to the first and second PCs of the turtle data, respectively. White circles, black squares, blue plus sign, and red cross mark correspond to male data points, female data points, the mean of male data points, and the mean of female data points, respectively

Table 2 Inner products between the first PCs for the genotypes E and F lizards in the lizard data and the unit vectors of the mean difference between genotypes E and F lizards

Vectors of interest	Inner product
First PCs for genotypes E and F lizards	0.9934
First PC for genotype E and mean difference between genotypes E and F lizards	0.9920
First PC for genotype F and mean difference between genotypes E and F lizards	0.9818

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

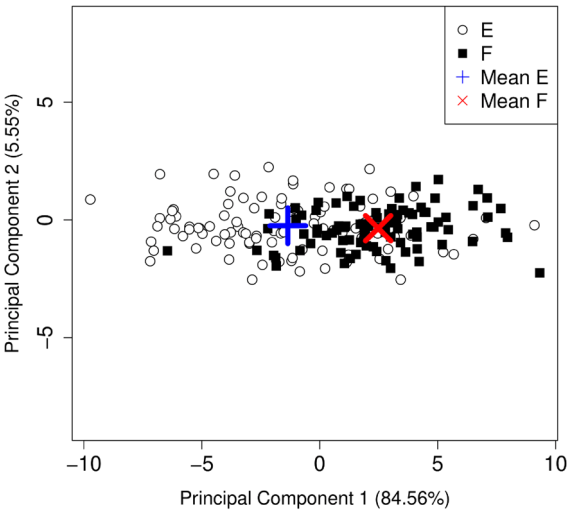

Fig. 3 Projection of the lizard data (Myers et al. 2020b) onto the space spanned by first and second principal component vectors. The x- and y- axes correspond to the first and second PCs of the lizard data, respectively. White circles, black squares, blue plus sign, and red cross mark correspond to genotype E data points, genotype F data points, the mean of genotype E data points, and the mean of genotype F data points, respectively

Table 3 Comparison of clustering methods for the turtle data (sample size: 48, dimension: 2)

Clustering method	Correctly clustered rate
Spectral clustering algorithm	0.7708
k-means	0.7708
EM algorithm for Gaussian mixture model	0.6875
Linear discriminant analysis	0.7500

Table 4 Comparison of clustering methods for the lizard data (sample size: 189, dimension: 13)

Clustering method	Correctly clustered rate
Spectral clustering algorithm	0.7407
k-means	0.7301
EM algorithm for Gaussian mixture model	0.6560
Linear discriminant analysis	0.7354

summarize the inner products between first PCs (for male and female turtles in the first dataset and genotypes E and F lizards in the second dataset) and the unit vectors of the mean differences between two groups. All values of the inner products are close to one, which suggests that datasets fit the allometric extension model. There are other examples, such as the 200-dimensional human data in Tsukuda and Matsuura (2023) and the five-dimensional crab data in Venables and Ripley (2002).

We apply the spectral clustering algorithm to the turtle and the lizard datasets. The results given by the spectral clustering algorithm are compared with those of the k-means algorithm, the expectation maximization (EM) algorithm for Gaussian mixture model, and the linear discriminant analysis in Tables 3 and 4. In these results, for the low-dimensional turtle dataset, the accuracy of the spectral clustering algorithm is

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

comparable to other methods. On the other hand, for the relatively high-dimensional lizard dataset, the spectral clustering algorithm provides higher accuracy than competitors. These examples motivate us to evaluate the accuracy of the spectral clustering algorithm for the allometric extension model.

Remark 4 For the applications described in this section, we used R with the packages `stats` (R Core Team 2023) and `mclust` (Scrucca et al. 2023). The k-means clustering (Hartigan–Wong and Lloyd algorithms) was conducted with the parameters *nstart* = 100 and *iter.max* = 1000, and the numbers of correctly clustered individuals for the both methods are the same. For the EM algorithm, the Gaussian mixture model was fitted using the VVV model, which means that there are no restrictions on covariance matrices (Scrucca et al. 2016). The linear discriminant analysis was performed using the mean vectors and covariance matrix estimated by the EM algorithm. Note that principal components are not used for these methods.

4 Spectral Clustering Algorithm for the Allometric Extension Model

In this section, we derive an upper bound on the misclassification probability of the spectral clustering algorithm for the allometric extension model. Let θ be a symmetric Bernoulli variable, let n be a positive integer, let $\boldsymbol{\mu}$ be an n-dimensional vector, and let $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ be $n \times n$ positive-definite symmetric matrices satisfying $\lambda_1(\boldsymbol{\Sigma}_i) > \lambda_2(\boldsymbol{\Sigma}_i)$ for $i = 1, 2$. Suppose that

$$\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1) = \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_2) = 2\beta\boldsymbol{\mu}$$

for some $\beta > 0$. We consider an n-dimensional random variable $\boldsymbol{X}$ defined as

$$\boldsymbol{X} = \theta\boldsymbol{\mu} + \boldsymbol{g}^{(\theta)},$$

where $\boldsymbol{g}^{(1)} \sim \mathcal{N}_n(\mathbf{0}_n, \boldsymbol{\Sigma}_1)$ and $\boldsymbol{g}^{(-1)} \sim \mathcal{N}_n(\mathbf{0}_n, \boldsymbol{\Sigma}_2)$ are independent of θ. Then, the probability density function $f_X(\cdot)$ of $\boldsymbol{X}$ is given by

$$f_X(\boldsymbol{x}) = \frac{1}{2} f_1(\boldsymbol{x}) + \frac{1}{2} f_2(\boldsymbol{x}) \quad (\boldsymbol{x} \in \mathbb{R}^n),$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are the probability density functions of $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}_n(-\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$, respectively. As shown in the following proposition, $\boldsymbol{X}$ is an $\mathbb{R}^n$-valued sub-Gaussian random variable.

Proposition 4.1 *There exists a constant $K \geq 1$ such that*

$$\|\langle \boldsymbol{X}, \boldsymbol{x} \rangle\|_{\psi_2} \leq K \|\langle \boldsymbol{X}, \boldsymbol{x} \rangle\|_{L^2} \tag{7}$$

for any $\boldsymbol{x} \in \mathbb{R}^n$. In particular, (7) holds for any $\boldsymbol{x} \in \mathbb{R}^n$ when $K = \sqrt{32/(4 - \mathrm{e})} (= 4.9966 \cdots)$.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

For a positive integer m, we observe random variables $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ following

$$\boldsymbol{X}_i = \theta_i \boldsymbol{\mu} + \boldsymbol{g}_i^{(\theta_i)} \quad (i = 1, \ldots, m),$$

where $\theta_1, \ldots, \theta_m$ are iid copies of θ and $\boldsymbol{g}_1^{(t)}, \ldots, \boldsymbol{g}_m^{(t)}$ are iid copies of $\boldsymbol{g}^{(t)}$ for $t = -1, 1$. The sample covariance matrix $\boldsymbol{S}_m$ is given in (1). Note that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ are iid copies of $\boldsymbol{X}$. In this setup, we presume that there are two groups, $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1)$ and $\mathcal{N}_n(-\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$, forming the allometric extension relationship, and θ_i indicates the group to which an individual $\boldsymbol{X}_i$ belongs for $i = 1, \ldots, m$. As an estimator of

$$\boldsymbol{\Sigma} = E[\boldsymbol{X}\boldsymbol{X}^\top] = \boldsymbol{\mu}\boldsymbol{\mu}^\top + \frac{1}{2}\boldsymbol{\Sigma}_1 + \frac{1}{2}\boldsymbol{\Sigma}_2, \tag{8}$$

the estimation error $\boldsymbol{S}_m - \boldsymbol{\Sigma}$ of $\boldsymbol{S}_m$ can be evaluated in the following proposition, which will be applied to prove our main result.

Proposition 4.2 *Let $K(\geq 1)$ be a constant satisfying (7) for any $\boldsymbol{x} \in \mathbb{R}^n$. For any $u \geq 0$,*

$$P\left(\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}} \leq CK^2\left(\sqrt{\frac{n+u}{m}} + \frac{n+u}{m}\right)\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2)}{2} + \|\boldsymbol{\mu}\|_2^2\right)\right) \geq 1 - 2\mathrm{e}^{-u},$$

where C is some positive absolute constant.

Remark 5 If we use the result of Exercise 4.7.3 of Vershynin (2018), Proposition 4.2 immediately follows from Propositions 2.1 and 4.1. The proof of Proposition 4.2 is included because a part of the proof is used in the proof of Theorem 4.3.

Remark 6 It also holds that

$$E[\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}}] \leq CK^2\left(\sqrt{\frac{n}{m}} + \frac{n}{m}\right)\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2)}{2} + \|\boldsymbol{\mu}\|_2^2\right),$$

which is a direct consequence of Theorem 4.7.1 of Vershynin (2018) and Propositions 2.1 and 4.1.

If we regard binary clustering as a binary classification problem for unlabeled data, the main objective is to classify individuals in a random sample into correct groups, where each individual belongs to one of the two groups. Let us assume that $\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{\gamma}_1(\boldsymbol{\Sigma})\rangle > 0$. In our problem setting, the spectral clustering algorithm classifies $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ into two clusters according to the signs of $\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i\rangle$ $(i = 1, \ldots, m)$. The misclassification probability of $\boldsymbol{X}_i$ can be expressed as

$$P(\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i\rangle < 0)$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

for $i = 1, \ldots, m$. As will be stated later, evaluating the misclassification probability enables us to evaluate the misclustering rate; see Remark 10. The following theorem provides a non-asymptotic upper bound of this misclassification probability.

Theorem 4.3 *Let C, K, K_g, and c be positive absolute constants that are independent of m, n, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$. Define*

$$c_1 = 1 + \frac{K_g^2}{\sqrt{c}}, \quad \eta = \frac{\|\boldsymbol{\mu}\|_2^2}{\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}$$

and

$$\delta = 2^{5/2} \cdot \min\left\{ c_1\sqrt{\frac{n}{m\eta}} + \frac{1}{\eta}\left(CK^2\left(\sqrt{\frac{n}{m}} + \frac{n}{m}\right) + \frac{1}{2}\right), \right.$$
$$\left. CK^2\left(\sqrt{\frac{n}{m}} + \frac{n}{m}\right)\left(\frac{1}{\eta} + 1\right)\right\}.$$

Suppose that δ, n, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ satisfy

$$\delta \leq \min\left\{ \frac{\alpha\|\boldsymbol{\mu}\|_2}{c_1\sqrt{(n-1)\max_{j=1,2}\{\lambda_2(\boldsymbol{\Sigma}_j)\}}}, \sqrt{2(1-\alpha)} \right\} \tag{9}$$

for some $\alpha \in (0, 1)$. Then, it holds that

$$P(\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0) \leq \Phi\left(-\frac{(2 - \delta^2 - 2\alpha)}{(2 + \delta^2)}\sqrt{\eta} \right) + 2(4 + \mathrm{e})\mathrm{e}^{-n}$$

for $i = 1, \ldots, m$, where $\Phi(\cdot)$ is the distribution function of $\mathcal{N}(0, 1)$.

Remark 7 Let us explain the constants C, K, K_g, and c in Theorem 4.3. The constants K and C are given in Propositions 4.1, and 4.2, respectively. Moreover, the constant K_g is the sub-Gaussian norm of a standard normal variable; in particular, $K_g = \sqrt{8/3}$. Letting $\boldsymbol{g} \sim \mathcal{N}_n(\boldsymbol{0}_n, \boldsymbol{I}_n)$, we have

$$P(\big|\|\boldsymbol{g}\|_2 - \sqrt{n}\big| \geq t) \leq 2\exp\left(-\frac{ct^2}{K_g^4}\right) \tag{10}$$

for all $t \geq 0$; see, e.g., Equation (3.3) in Vershynin (2018, p. 40). The constant c in Theorem 4.3 appears in the inequality (10).

The quantity η, which is a signal-to-noise ratio for our problem, plays an important role in evaluating performance. The numerator $\|\boldsymbol{\mu}\|_2^2$ and the denominator $\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}$ correspond to the strengths of the signal and the noise, respectively. If the noise is large compared to the signal, then the signal is hidden by the

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

noise, which makes detection difficult. In this sense, η can be interpreted as the difficulty of classification when the allometric extension model is considered. If η is small (large), it is difficult (easy) to classify individuals in a random sample into the correct groups. Roughly speaking, if either m or η is sufficiently large, then (9) holds.

Remark 8 This topic is discussed in a similar context by Abbe et al. (2022) and Ndaoud (2022).

Remark 9 By the Mills inequality, Theorem 4.3 leads to

$$\begin{aligned} &P(\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle \leq 0) \\ &\leq \frac{2+\delta^2}{(2-\delta^2-2\alpha)\sqrt{2\pi\eta}} \exp\left(-\left(\frac{2-\delta^2-2\alpha}{2+\delta^2}\right)^2 \frac{\eta}{2}\right) \\ &+ 2(4+\mathrm{e})\mathrm{e}^{-n} \quad (i=1,\ldots,m). \end{aligned} \tag{11}$$

The inequality (11) shows that the misclassification probability for an individual decays exponentially with the signal-to-noise ratio η (multiplied by a constant) and the dimension n.

Theorem 4.3 provides a non-asymptotic lower bound for the probability of the misclassification rate announced in Sect. 1. In our setting, (2) is formulated as

$$\begin{aligned} &P\left(\sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\} \leq \varepsilon m\right) \\ &= 1 - P\left(\sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\} > \varepsilon m\right). \end{aligned} \tag{12}$$

Applying the Markov inequality, we have

$$\begin{aligned} &P\left(\sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\} > \varepsilon m\right) \\ &\leq \frac{1}{\varepsilon m} \sum_{i=1}^{m} E\left[1\left\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\right\}\right] \\ &= \frac{1}{\varepsilon} P\left(\theta_1 \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_1 \rangle < 0\right). \end{aligned}$$

Thus, we obtain the following corollary to Theorem 4.3.

Corollary 4.4 *Consider the constants C, K, and c_1 in Theorem 4.3. Suppose that m, n, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ satisfy (9) for some $\alpha \in (0,1)$. Then, it holds that*

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

$$P\left(\sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\} \leq \varepsilon m\right) \geq 1 - \frac{1}{\varepsilon}\Phi\left(-\frac{(2-\delta^2-2\alpha)}{(2+\delta^2)}\sqrt{\eta}\right) - \frac{2}{\varepsilon}(4+\mathrm{e})\mathrm{e}^{-n}.$$

Remark 10 In our setting, the misclustering rate mentioned in Sect. 1 is expressed as

$$\frac{1}{m}\min\left\{\sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\}, \sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle > 0\}\right\}. \tag{13}$$

It holds that

$$P\left(\frac{1}{m}\min\left\{\sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\}, \sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle > 0\}\right\} \leq \varepsilon\right) \geq P\left(\sum_{i=1}^{m} 1\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\} \leq \varepsilon m\right).$$

We can evaluate the right-hand side using Corollary 4.4. In addition, the expectation of (13) may be evaluated similarly. Note that (13) is called the "misclassification rate" in Cai and Zhang (2018) and Abbe et al. (2022).

Finally, we consider the probability of the event

$$\left\{\bigcap_{i=1}^{m}\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle > 0\}\right\} \cup \left\{\bigcap_{i=1}^{m}\{\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0\}\right\},$$

indicating that all individuals $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ are clustered correctly, that is, the misclustering rate is equal to zero. Theorem 4.3 implies the consistency of clustering in the sense of (16) under a high-dimensional regime.

Corollary 4.5 *As $n \to \infty$ with*

$$\frac{\log m}{n} \to 0, \tag{14}$$

if

$$\frac{1}{\eta}\max\left\{\frac{n}{m}, \frac{n^2 \max_{j=1,2}\{\lambda_2(\boldsymbol{\Sigma}_j)\}}{m\eta \max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}, \frac{n \max_{j=1,2}\{\lambda_2(\boldsymbol{\Sigma}_j)\}}{\eta^2 \max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}, \log m\right\} \to 0, \tag{15}$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

then

$$P\left(\left\{\bigcap_{i=1}^{m}\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle>0\}\right\}\cup\left\{\bigcap_{i=1}^{m}\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle<0\}\right\}\right)\to 1. \quad (16)$$

Remark 11 Even when $n/m\to\infty$, the condition (15) holds if η is sufficiently large. For example, if $\eta\asymp n$, $m\to\infty$, or if $n/\eta\to 0$, then (15) holds.

Remark 12 Under the same assumption as Corollary 4.5, it can be shown that

$$E\left[\min\left\{\sum_{i=1}^{m}1\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle<0\},\sum_{i=1}^{m}1\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle>0\}\right\}\right]\to 0. \quad (17)$$

We compare our sufficient conditions for (17) when $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=I_n$ with those of Cai and Zhang (2018). Note that although $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=I_n$ is excluded to define $\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)$ and $\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_2)$ uniquely in our setting, the same method to prove (17) works. First, our sufficient conditions are summarized in

$$\frac{\log m}{n},\ \frac{n}{\|\boldsymbol{\mu}\|_2^2 m},\ \frac{n^2}{\|\boldsymbol{\mu}\|_2^4 m},\ \frac{n}{\|\boldsymbol{\mu}\|_2^6},\ \frac{\log m}{\|\boldsymbol{\mu}\|_2^2}\to 0. \quad (18)$$

On the other hand, the sufficient conditions of Cai and Zhang (2018) are

$$\frac{n}{\|\boldsymbol{\mu}\|_2^4},\ \frac{m}{\|\boldsymbol{\mu}\|_2^2}\to 0. \quad (19)$$

When $m\asymp n$, (18) is milder than (19), because (18) and (19) become $n/\|\boldsymbol{\mu}\|_2^4\to 0$ and $n/\|\boldsymbol{\mu}\|_2^2\to 0$, respectively. Moreover, when $n\to\infty$ with m fixed, (18) is more restricted than (19), because (18) and (19) are reduced to $n/\|\boldsymbol{\mu}\|_2^2\to 0$ and $n/\|\boldsymbol{\mu}\|_2^4\to 0$, respectively.

5 Proofs

5.1 Proof of Proposition 2.1

For simplicity, denote $\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)=\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_2)$ by $\boldsymbol{\gamma}_1$, where the sign of $\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_2)$ is appropriately chosen. Let us construct a random variable Y taking values in $\{1,2\}$ such that $P(Y=i)=\pi_i$ and $\boldsymbol{X}|\{Y=i\}\sim\mathcal{N}_n(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$ conditionally for $i=1,2$. Then, we have $E[\boldsymbol{X}|Y=i]=\mu_i$ and $V[\boldsymbol{X}|Y=i]=\boldsymbol{\Sigma}_i$ for $i=1,2$. These formulae give

$$E[V[\boldsymbol{X}|Y]]=\sum_{i=1}^{2}\pi_i\boldsymbol{\Sigma}_i,$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

$$V[E[\boldsymbol{X}|Y]] = \sum_{i=1}^{2} \pi_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)^\top,$$

where

$$\bar{\boldsymbol{\mu}} = \sum_{i=1}^{2} \pi_i \boldsymbol{\mu}_i = \boldsymbol{\mu}_1 - \frac{\pi_2}{\beta}\boldsymbol{\gamma}_1 = \boldsymbol{\mu}_2 + \frac{\pi_1}{\beta}\boldsymbol{\gamma}_1.$$

It follows that

$$\begin{aligned}\boldsymbol{\Sigma} &= E[V[\boldsymbol{X}|Y]] + V[E[\boldsymbol{X}|Y]] = \sum_{i=1}^{2} \pi_i \boldsymbol{\Sigma}_i + \sum_{i=1}^{2} \pi_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)^\top \\ &= \sum_{i=1}^{2} \pi_i \boldsymbol{\Sigma}_i + \frac{\pi_1\pi_2}{\beta^2}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top,\end{aligned}$$

because

$$\begin{aligned}&\sum_{i=1}^{2} \pi_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)^\top = \pi_1 \left(\frac{\pi_2}{\beta}\boldsymbol{\gamma}_1\right)\left(\frac{\pi_2}{\beta}\boldsymbol{\gamma}_1\right)^\top + \pi_2 \left(-\frac{\pi_1}{\beta}\boldsymbol{\gamma}_1\right)\left(-\frac{\pi_1}{\beta}\boldsymbol{\gamma}_1\right)^\top \\ &= \frac{\pi_1\pi_2^2}{\beta^2}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top + \frac{\pi_1^2\pi_2}{\beta^2}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top = \frac{\pi_1\pi_2}{\beta^2}(\pi_1+\pi_2)\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top = \frac{\pi_1\pi_2}{\beta^2}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top,\end{aligned}$$

which comes from the definition of the allometric extension model. From

$$\boldsymbol{\Sigma}\boldsymbol{\gamma}_1 = \sum_{i=1}^{2} \pi_i \boldsymbol{\Sigma}_i \boldsymbol{\gamma}_1 + \frac{\pi_1\pi_2}{\beta^2}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top\boldsymbol{\gamma}_1 = \left(\sum_{i=1}^{2} \pi_i \lambda_1(\boldsymbol{\Sigma}_i) + \frac{\pi_1\pi_2}{\beta^2}\right)\boldsymbol{\gamma}_1,$$

we see that $\boldsymbol{\gamma}_1$ is an eigenvector of $\boldsymbol{\Sigma}$ corresponding to the eigenvalue

$$\sum_{i=1}^{2} \pi_i \lambda_1(\boldsymbol{\Sigma}_i) + \frac{\pi_1\pi_2}{\beta^2}.$$

Let $\boldsymbol{\xi}$ be another unit-length eigenvector of $\boldsymbol{\Sigma}$ orthogonal to $\boldsymbol{\gamma}_1$. As $\boldsymbol{\xi}$ and $\boldsymbol{\gamma}_1$ are orthogonal, it holds that

$$\boldsymbol{\xi}^\top\boldsymbol{\Sigma}\boldsymbol{\xi} = \sum_{i=1}^{2} \pi_i \boldsymbol{\xi}^\top\boldsymbol{\Sigma}_i\boldsymbol{\xi} + \frac{\pi_1\pi_2}{\beta^2}\boldsymbol{\xi}^\top\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top\boldsymbol{\xi} = \sum_{i=1}^{2} \pi_i \boldsymbol{\xi}^\top\boldsymbol{\Sigma}_i\boldsymbol{\xi}.$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Using the spectral decomposition

$$\boldsymbol{\Sigma}_i = \lambda_1(\boldsymbol{\Sigma}_i)\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^\top + \sum_{j=2}^{n} \lambda_j(\boldsymbol{\Sigma}_i)\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_i)\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_i)^\top \quad (i = 1, 2),$$

we have

$$\begin{aligned}\boldsymbol{\xi}^\top \boldsymbol{\Sigma}_i \boldsymbol{\xi} &= \boldsymbol{\xi}^\top \left(\sum_{j=2}^{n} \lambda_j(\boldsymbol{\Sigma}_i)\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_i)\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_i)^\top \right) \boldsymbol{\xi} = \sum_{j=2}^{n} \lambda_j(\boldsymbol{\Sigma}_i)\langle \boldsymbol{\xi}, \boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_i)\rangle^2 \\ &\leq \lambda_2(\boldsymbol{\Sigma}_i) \sum_{j=2}^{n} \langle \boldsymbol{\xi}, \boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_i)\rangle^2 = \lambda_2(\boldsymbol{\Sigma}_i)\end{aligned}$$

for $i = 1, 2$. Consequently, we deduce that

$$\boldsymbol{\xi}^\top \boldsymbol{\Sigma} \boldsymbol{\xi} \leq \sum_{i=1}^{2} \pi_i \lambda_2(\boldsymbol{\Sigma}_i) < \sum_{i=1}^{2} \pi_i \lambda_1(\boldsymbol{\Sigma}_i) + \frac{\pi_1 \pi_2}{\beta^2}$$

for any unit-length eigenvector $\boldsymbol{\xi}$ of $\boldsymbol{\Sigma}$ that is orthogonal to $\boldsymbol{\gamma}_1$. Hence,

$$\sum_{i=1}^{2} \pi_i \lambda_1(\boldsymbol{\Sigma}_i) + \frac{\pi_1 \pi_2}{\beta^2}$$

is the unique largest eigenvalue $\lambda_1(\boldsymbol{\Sigma})$ of $\boldsymbol{\Sigma}$ and

$$\lambda_2(\boldsymbol{\Sigma}) \leq \sum_{i=1}^{2} \pi_i \lambda_2(\boldsymbol{\Sigma}_i).$$

This completes the proof. □

5.2 Proof of Proposition 4.1

When $\boldsymbol{x} = \mathbf{0}_n$, (7) holds for any $K(\geq 1)$. Hereafter, we consider $\boldsymbol{x} \neq \mathbf{0}_n$. We first observe that

$$\|\langle \boldsymbol{X}, \boldsymbol{x}\rangle\|_{L^2}^2 = E[\langle \boldsymbol{X}, \boldsymbol{x}\rangle^2] = E[\boldsymbol{x}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{x}] = \boldsymbol{x}^\top E[\boldsymbol{X}\boldsymbol{X}^\top]\boldsymbol{x} = \langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle.$$

Next, we evaluate $\|\langle \boldsymbol{X}, \boldsymbol{x}\rangle\|_{\psi_2}$. It holds that

$$E\left[\exp\left(\frac{\langle \boldsymbol{X}, \boldsymbol{x}\rangle^2}{t^2}\right)\right] = E\left[\exp\left(\frac{\langle \theta\boldsymbol{\mu} + \boldsymbol{g}^{(\theta)}, \boldsymbol{x}\rangle^2}{t^2}\right)\right]$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

$$= E\left[E\left[\exp\left(\frac{\langle\theta\boldsymbol{\mu}+\boldsymbol{g}^{(\theta)},\boldsymbol{x}\rangle^2}{t^2}\right)\middle|\theta\right]\right]$$
$$= \frac{1}{2}E\left[\exp\left(\frac{\langle\boldsymbol{\mu}+\boldsymbol{g}^{(1)},\boldsymbol{x}\rangle^2}{t^2}\right)\right]+\frac{1}{2}E\left[\exp\left(\frac{\langle-\boldsymbol{\mu}+\boldsymbol{g}^{(-1)},\boldsymbol{x}\rangle^2}{t^2}\right)\right], \quad (20)$$

where t will be specified later. For the first term on the right-hand side of (20), it follows from $\langle\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}^{(1)}, \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}/\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2\rangle \sim \mathcal{N}(0,1)$ that

$$\begin{aligned}
&\frac{1}{2}E\left[\exp\left(\frac{\langle\boldsymbol{\mu}+\boldsymbol{g}^{(1)},\boldsymbol{x}\rangle^2}{t^2}\right)\right]\\
&\le \frac{1}{2}E\left[\exp\left(\frac{2\langle\boldsymbol{\mu},\boldsymbol{x}\rangle^2+2\langle\boldsymbol{g}^{(1)},\boldsymbol{x}\rangle^2}{t^2}\right)\right]\\
&= \frac{1}{2}\exp\left(\frac{2\langle\boldsymbol{\mu},\boldsymbol{x}\rangle^2}{t^2}\right)E\left[\exp\left(\frac{2\langle\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}^{(1)},\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\rangle^2}{t^2}\right)\right]\\
&= \frac{1}{2}\exp\left(\frac{2\langle\boldsymbol{\mu},\boldsymbol{x}\rangle^2}{t^2}\right)E\left[\exp\left(\frac{2\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2\left\langle\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}^{(1)},\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}/\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2\right\rangle^2}{t^2}\right)\right]\\
&= \frac{1}{2}\exp\left(\frac{2\langle\boldsymbol{\mu},\boldsymbol{x}\rangle^2}{t^2}\right)E\left[\exp\left(\frac{Z^2}{t^2/(2\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2)}\right)\right],
\end{aligned}$$

where Z is a standard normal random variable. The expectation on the right-hand side is finite when $t^2 > 4\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2$ and is given by

$$E\left[\exp\left(\frac{Z^2}{t^2/(2\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2)}\right)\right] = \frac{1}{\sqrt{1-4\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2/t^2}},$$

which yields

$$\frac{1}{2}E\left[\exp\left(\frac{\langle\boldsymbol{\mu}+\boldsymbol{g}^{(1)},\boldsymbol{x}\rangle^2}{t^2}\right)\right] \le \frac{1}{2}\exp\left(\frac{2\langle\boldsymbol{\mu},\boldsymbol{x}\rangle^2}{t^2}\right)\frac{1}{\sqrt{1-4\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2/t^2}}.$$

Similarly, for the second term on the right-hand side of (20), when $t^2 > 4\|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{x}\|_2^2$, it holds that

$$\frac{1}{2}E\left[\exp\left(\frac{\langle-\boldsymbol{\mu}+\boldsymbol{g}^{(-1)},\boldsymbol{x}\rangle^2}{t^2}\right)\right] \le \frac{1}{2}\exp\left(\frac{2\langle\boldsymbol{\mu},\boldsymbol{x}\rangle^2}{t^2}\right)\frac{1}{\sqrt{1-4\|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{x}\|_2^2/t^2}}.$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Letting $M = \max\{\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2, \|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{x}\|_2\}$, we have

$$E\left[\exp\left(\frac{\langle \boldsymbol{X}, \boldsymbol{x}\rangle^2}{t^2}\right)\right] \leq \exp\left(\frac{2\langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2}{t^2}\right)\frac{1}{\sqrt{1-4M^2/t^2}}$$

when $t^2 > 4M^2$. Hence, from the definition of the sub-Gaussian norm, it suffices to show that there exists a constant K satisfying this inequality with $t = K\sqrt{\langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle}$ for all $\boldsymbol{x}(\neq \boldsymbol{0}_n)$. It follows from (8) that

$$\langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle = \langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2 + \frac{1}{2}\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2 + \frac{1}{2}\|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{x}\|_2^2 \geq \langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2.$$

Moreover, substituting $t = K\sqrt{\langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle}$ into $\exp(2\langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2/t^2)$, we have

$$\exp\left(\frac{2\langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2}{t^2}\right) = \exp\left(\frac{2\langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2}{K^2\langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle}\right) \leq \exp\left(\frac{2\langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2}{K^2\langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2}\right) = \exp\left(\frac{2}{K^2}\right).$$

Furthermore, if $K \geq 2$, then $\exp(2/K^2) \leq \sqrt{\mathrm{e}}$. Therefore, it only remains to show that there exists a constant $K \geq 2$ such that

$$\frac{1}{\sqrt{1-4M^2/(K^2\langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle)}} \leq \frac{2}{\sqrt{\mathrm{e}}} \tag{21}$$

for all $\boldsymbol{x}(\neq 0)$. The inequality (21) is equivalent to

$$K^2 \geq \frac{16}{\mathrm{e}}\left(\frac{4}{\mathrm{e}} - 1\right)^{-1}\frac{M^2}{\langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle}.$$

From

$$\frac{M^2}{\langle \boldsymbol{\Sigma}\boldsymbol{x}, \boldsymbol{x}\rangle} = \frac{M^2}{\langle \boldsymbol{\mu}, \boldsymbol{x}\rangle^2 + \|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2/2 + \|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{x}\|_2^2/2} \leq \frac{2M^2}{\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{x}\|_2^2 + \|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{x}\|_2^2} \leq 2,$$

it follows that

$$K = \sqrt{\frac{32}{\mathrm{e}}\left(\frac{4}{\mathrm{e}} - 1\right)^{-1}} = \sqrt{\frac{32}{4-\mathrm{e}}}$$

satisfies (21) for all $\boldsymbol{x}(\neq \boldsymbol{0}_n)$. □

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

5.3 Proof of Proposition 4.2

Fix $u \geq 0$. Let

$$\boldsymbol{Z}_i = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}_i \quad (i = 1, \ldots, m), \quad \text{and} \quad \boldsymbol{R} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top - \boldsymbol{I}_n.$$

It follows from Equation (4.25) of Vershynin (2018, p. 94) that

$$\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}} \leq \|\boldsymbol{R}\|_{\mathrm{op}} \|\boldsymbol{\Sigma}\|_{\mathrm{op}}.$$

Using Proposition 4.1 and Equation (4.22) of Vershynin (2018, p. 91) with $t = \sqrt{u}$, we have

$$P\left(\|\boldsymbol{R}\|_{\mathrm{op}} \leq K^2 \max\{\kappa, \kappa^2\}\right) \geq 1 - 2\mathrm{e}^{-u},$$

where $\kappa = \tilde{C}(\sqrt{n} + \sqrt{u})/\sqrt{m}$. Note that $\tilde{C}$ is the absolute constant C in Equation (4.22) of Vershynin (2018). We find that

$$\kappa = \tilde{C}\left(\frac{\sqrt{n} + \sqrt{u}}{\sqrt{m}}\right) \leq \tilde{C}\sqrt{\frac{2(n+u)}{m}}.$$

Letting $C = \max\{\sqrt{2}\tilde{C}, 2\tilde{C}^2\}$, we have

$$K^2 \max\{\kappa, \kappa^2\} \leq K^2(\kappa + \kappa^2) \leq CK^2\left(\sqrt{\frac{n+u}{m}} + \frac{n+u}{m}\right). \tag{22}$$

Consequently, it follows that

$$\begin{aligned} P\left(\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}} \leq CK^2\left(\sqrt{\frac{n+u}{m}} + \frac{n+u}{m}\right)\|\boldsymbol{\Sigma}\|_{\mathrm{op}}\right) \\ \geq P\left(\|\boldsymbol{R}\|_{\mathrm{op}} \leq CK^2\left(\sqrt{\frac{n+u}{m}} + \frac{n+u}{m}\right)\right) \\ \geq P\left(\|\boldsymbol{R}\|_{\mathrm{op}} \leq K^2 \max\{\kappa, \kappa^2\}\right) \geq 1 - 2\mathrm{e}^{-u}. \end{aligned}$$

Finally, (5) implies $\lambda_1(\boldsymbol{\Sigma}) = (\lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2))/2 + \|\boldsymbol{\mu}\|_2^2$. This completes the proof. □

5.4 Technical Lemmas

The following lemma will be used in the proof of Lemma 5.3.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Lemma 5.1 *It holds that*

$$P\left(\left\|\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{g}_i^{(\theta_i)}\right\|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\right) \le 2\mathrm{e}^{-n}.$$

Proof Consider the power set $\mathcal{A} = 2^{\{1,\ldots,m\}}$ of the set $\{1,\ldots,m\}$ and the random set $\mathrm{A}_\theta = \{i = 1,\ldots,m : \theta_i = 1\}$. For a set $\mathrm{A} \in \mathcal{A}$, define the event $\mathrm{E}_\mathrm{A} = \{\mathrm{A}_\theta = \mathrm{A}\}$. We have

$$\begin{aligned}
&P\left(\left|\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{g}^{(\theta_i)}\right|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\Big|\mathrm{E}_\mathrm{A}\right)\\
&= P\left(\left|\frac{1}{m}\left(\sum_{i\in\mathrm{A}}\boldsymbol{g}_i^{(\theta_i)} + \sum_{i\in\mathrm{A}^c}\boldsymbol{g}_i^{(\theta_i)}\right)\right|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\Big|\mathrm{E}_\mathrm{A}\right)\\
&\le P\Bigg(\frac{1}{m}\left\|\left(\#(\mathrm{A})\boldsymbol{\Sigma}_1 + \#(\mathrm{A}^c)\boldsymbol{\Sigma}_2\right)^{1/2}\right\|_{\mathrm{op}}\left\|\left(\#(\mathrm{A})\boldsymbol{\Sigma}_1 + \#(\mathrm{A}^c)\boldsymbol{\Sigma}_2\right)^{-1/2}\right.\\
&\cdot\left.\left(\sum_{i\in\mathrm{A}}\boldsymbol{g}_i^{(\theta_i)} + \sum_{i\in\mathrm{A}^c}\boldsymbol{g}_i^{(\theta_i)}\right)\right\|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\Big|\mathrm{E}_\mathrm{A}\Bigg)\\
&= P\left(\frac{1}{m}\left|\#(\mathrm{A})\boldsymbol{\Sigma}_1 + \#(\mathrm{A}^c)\boldsymbol{\Sigma}_2\right|_{\mathrm{op}}^{1/2}\|\boldsymbol{Z}\|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\Big|\mathrm{E}_\mathrm{A}\right)\\
&\le P\left(\frac{1}{m}\left(\#(\mathrm{A})\,\|\boldsymbol{\Sigma}_1\|_{\mathrm{op}} + \#(\mathrm{A}^c)\,|\boldsymbol{\Sigma}_2|_{\mathrm{op}}\right)^{1/2}|\boldsymbol{Z}|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\Big|\mathrm{E}_\mathrm{A}\right)\\
&\le P\left(\frac{1}{m}\left(\left(\#(\mathrm{A}) + \#(\mathrm{A}^c)\right)\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}\right)^{1/2}|\boldsymbol{Z}|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\Big|\mathrm{E}_\mathrm{A}\right)\\
&= P\left(\sqrt{\frac{\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\,|\boldsymbol{Z}|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\Big|\mathrm{E}_\mathrm{A}\right)\\
&= P\left(|\boldsymbol{Z}|_2 > c_1\sqrt{n}\Big|\mathrm{E}_\mathrm{A}\right),
\end{aligned}$$

where $\boldsymbol{Z} \sim \mathcal{N}_n(\boldsymbol{0}_n, \boldsymbol{I}_n)$. Recalling that $c_1 = 1 + K_g^2/\sqrt{c}$, we find that

$$\begin{aligned}
&P\left(|\boldsymbol{Z}|_2 > c_1\sqrt{n}\Big|\mathrm{E}_\mathrm{A}\right)\\
&\quad= P(|\,\|\boldsymbol{Z}\|_2 - \sqrt{n} + \sqrt{n}| > c_1\sqrt{n}|\mathrm{E}_\mathrm{A})\\
&\quad\le P(|\,\|\boldsymbol{Z}\|_2 - \sqrt{n}| > (c_1-1)\sqrt{n}|\mathrm{E}_\mathrm{A})\\
&\quad\le 2\mathrm{e}^{-n},
\end{aligned} \tag{23}$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

where (10) is used to obtain the last inequality. Then, we find that

$$
\begin{aligned}
&P\left(\left\|\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{g}_i^{(\theta_i)}\right\|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\right)\\
&=\sum_{\mathrm{A}\in\mathcal{A}} P\left(\left\|\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{g}_i^{(\theta_i)}\right\|_2 > c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\middle|\mathrm{E_A}\right) P\left(\mathrm{E_A}\right)\\
&\le 2\mathrm{e}^{-n}\sum_{\mathrm{A}\in\mathcal{A}} P\left(\mathrm{E_A}\right)\\
&=2\mathrm{e}^{-n}.
\end{aligned}
$$

□

The following Lemma 5.2 is used in the proof of Lemma 5.3.

Lemma 5.2 *Consider an n-dimensional random variable* $\boldsymbol{X}\sim(\boldsymbol{0}_n,\boldsymbol{\Sigma})$. *Suppose that there exists a constant* K *such that*

$$\|\langle\boldsymbol{X},\boldsymbol{x}\rangle\|_{\psi_2}\le K\,\|\langle\boldsymbol{X},\boldsymbol{x}\rangle\|_{L_2}$$

for any $\boldsymbol{x}\in\mathbb{R}^n$. *Then, it holds that*

$$\sup_{\boldsymbol{x}\in S^{n-1}}\left\|\langle\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X},\boldsymbol{x}\rangle\right\|_{\psi_2}\le K. \tag{24}$$

Proof We have

$$
\begin{aligned}
&\sup_{\boldsymbol{x}\in S^{n-1}}\left\|\langle\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X},\boldsymbol{x}\rangle\right\|_{\psi_2}\\
&=\sup_{\boldsymbol{x}\in S^{n-1}}\left\|\langle\boldsymbol{X},\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}\rangle\right\|_{\psi_2}\\
&\le\sup_{\boldsymbol{x}\in S^{n-1}}K\left\|\langle\boldsymbol{X},\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}\rangle\right\|_{L_2}\\
&=\sup_{\boldsymbol{x}\in S^{n-1}}K\left\|\langle\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X},\boldsymbol{x}\rangle\right\|_{L_2}\\
&=\sup_{\boldsymbol{x}\in S^{n-1}}K\|\boldsymbol{x}\|_2\\
&=K.
\end{aligned}
$$

This first inequality follows from the assumption of Lemma 5.2, and the second-last equality holds because $\left\|\langle\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X},\boldsymbol{x}\rangle\right\|_{L_2}=\|\boldsymbol{x}\|_2$. □

The following lemma will be used in the proof of Lemma 5.4.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Lemma 5.3 *It holds that*

$$P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma})\|_2 \le \frac{2^{3/2}}{\|\boldsymbol{\mu}\|_2^2}\left[2\|\boldsymbol{\mu}\|_2 c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}} + \left(\frac{\lambda_1(\boldsymbol{\Sigma}_1)+\lambda_1(\boldsymbol{\Sigma}_2)}{2}\right)\left\{CK^2\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)+1\right\}\right]\right) \le 4\mathrm{e}^{-n}.$$

Proof We have

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i\boldsymbol{X}_i^\top &= \frac{1}{m}\sum_{i=1}^m\left(\theta_i\boldsymbol{\mu}+\boldsymbol{g}_i^{(\theta_i)}\right)\left(\theta_i\boldsymbol{\mu}+\boldsymbol{g}_i^{(\theta_i)}\right)^\top\\ &= \frac{1}{m}\sum_{i=1}^m\left(\boldsymbol{\mu}\boldsymbol{\mu}^\top+\theta_i\boldsymbol{\mu}\boldsymbol{g}_i^{(\theta_i)\top}+\theta_i\boldsymbol{g}_i^{(\theta_i)}\boldsymbol{\mu}^\top+\boldsymbol{g}_i^{(\theta_i)}\boldsymbol{g}_i^{(\theta_i)\top}\right)\\ &= \boldsymbol{\mu}\boldsymbol{\mu}^\top+\boldsymbol{\mu}\left(\frac{1}{m}\sum_{i=1}^m\theta_i\boldsymbol{g}_i^{(\theta_i)}\right)^\top+\left(\frac{1}{m}\sum_{i=1}^m\theta_i\boldsymbol{g}_i^{(\theta_i)}\right)\boldsymbol{\mu}^\top\\ &\quad+\frac{1}{m}\sum_{i=1}^m\boldsymbol{g}_i^{(\theta_i)}\boldsymbol{g}_i^{(\theta_i)\top}.\end{aligned}$$

Then, by the triangle inequality, we have

$$\begin{aligned}&\left\|\frac{1}{m}\sum_{i=1}^m\boldsymbol{X}_i\boldsymbol{X}_i^\top-\boldsymbol{\mu}\boldsymbol{\mu}^\top\right\|_{\mathrm{op}}\\ &\le\left\|\boldsymbol{\mu}\left(\frac{1}{m}\sum_{i=1}^m\theta_i\boldsymbol{g}_i^{(\theta_i)}\right)^\top\right\|_{\mathrm{op}}+\left\|\left(\frac{1}{m}\sum_{i=1}^m\theta_i\boldsymbol{g}_i^{(\theta_i)}\right)\boldsymbol{\mu}^\top\right\|_{\mathrm{op}}\\ &\quad+\left\|\frac{1}{m}\sum_{i=1}^m\boldsymbol{g}_i^{(\theta_i)}\boldsymbol{g}_i^{(\theta_i)\top}\right\|_{\mathrm{op}}\\ &= 2\|\boldsymbol{\mu}\|_2\left\|\frac{1}{m}\sum_{i=1}^m\theta_i\boldsymbol{g}_i^{(\theta_i)}\right\|_2+\left\|\frac{1}{m}\sum_{i=1}^m\boldsymbol{g}_i^{(\theta_i)}\boldsymbol{g}_i^{(\theta_i)\top}\right\|_{\mathrm{op}},\end{aligned}$$

where we have used the fact that $\|\boldsymbol{a}\boldsymbol{b}^\top\|_{\mathrm{op}}=\|\boldsymbol{a}\|_2\|\boldsymbol{b}\|_2$ for n-dimensional vectors $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^n$. The first term on the right-hand side can be evaluated by Lemma 5.1, because the distributions of $\boldsymbol{g}_i^{(\theta_i)}$ and $\theta_i\boldsymbol{g}_i^{(\theta_i)}$ are the same. For $\boldsymbol{\Sigma}_*=(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2)/2$, we observe that

$$\left\|\frac{1}{m}\sum_{i=1}^m\boldsymbol{g}_i^{(\theta_i)}\boldsymbol{g}_i^{(\theta_i)\top}\right\|_{\mathrm{op}} \le \|\boldsymbol{\Sigma}_*\|_{\mathrm{op}}\left\|\frac{1}{m}\sum_{i=1}^m\left(\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right)\left(\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right)^\top\right\|_{\mathrm{op}}. \quad (25)$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Consider the $m \times n$ matrix $\boldsymbol{A}$ whose i-th row is $\left(\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right)^\top$ $(i = 1, \ldots, m)$. Then,

$$\frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right)\left(\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right)^\top = \frac{1}{m}\boldsymbol{A}^\top\boldsymbol{A}.$$

Using Proposition 4.1 and Lemma 5.2, we find that

$$\left\|\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right\|_{\psi_2} = \sup_{\boldsymbol{x}\in S^{n-1}}\left\|\langle\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{g}_i^{(\theta_i)}, \boldsymbol{x}\rangle\right\|_{\psi_2} \leq K.$$

Then, from the equation (4.22) of Vershynin (2018, p. 91) and (22), it follows that

$$P\left(\left\|\frac{1}{m}\boldsymbol{A}^\top\boldsymbol{A} - \boldsymbol{I}_n\right\|_{\mathrm{op}} \leq CK^2\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right)\right) \geq 1 - 2\mathrm{e}^{-n}. \tag{26}$$

Using (25) and (26), we have

$$\begin{aligned}
&P\left(\left\|\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{g}_i^{(\theta_i)}\boldsymbol{g}_i^{(\theta_i)\top}\right\|_{\mathrm{op}} \leq \left(\frac{\lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2)}{2}\right)\left(CK^2\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right) + 1\right)\right) \\
&\geq P\left(\|\boldsymbol{\Sigma}_*\|_{\mathrm{op}}\left\|\frac{1}{m}\boldsymbol{A}^\top\boldsymbol{A}\right\|_{\mathrm{op}} \leq \left(\frac{\lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2)}{2}\right)\left(CK^2\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right) + 1\right)\right) \\
&= P\left(\left\|\frac{1}{m}\boldsymbol{A}^\top\boldsymbol{A}\right\|_{\mathrm{op}} \leq CK^2\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right) + 1\right) \\
&\geq P\left(\left\|\frac{1}{m}\boldsymbol{A}^\top\boldsymbol{A} - \boldsymbol{I}_n\right\|_{\mathrm{op}} + \|\boldsymbol{I}_n\|_{\mathrm{op}} \leq CK^2\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right) + 1\right) \\
&= P\left(\left\|\frac{1}{m}\boldsymbol{A}^\top\boldsymbol{A} - \boldsymbol{I}_n\right\|_{\mathrm{op}} \leq CK^2\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right)\right) \\
&\geq 1 - 2\mathrm{e}^{-n}.
\end{aligned}$$

For simplicity, let us define

$$\begin{aligned}
D &= \frac{2^{3/2}}{\|\boldsymbol{\mu}\|_2^2}\left(2\|\boldsymbol{\mu}\|_2 c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}} + \left(\frac{\lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2)}{2}\right)\right. \\
&\quad \left.\cdot\left(CK^2\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right) + 1\right)\right), \\
\mathrm{E}_1 &= \left\{\left\|\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{g}_i^{(\theta_i)}\right\|_2 \leq c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\right\},
\end{aligned}$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

$$\mathrm{E}_2 = \left\{ \left\| \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{g}_i^{(\theta_i)} \boldsymbol{g}_i^{(\theta_i)\top} \right\|_{\mathrm{op}} \leq \left(\frac{\lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2)}{2} \right) \left(CK^2 \left(\sqrt{\frac{2n}{m}} + \frac{2n}{m} \right) + 1 \right) \right\}.$$

The Davis–Kahan theorem yields

$$\begin{aligned}
&P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma})\|_2 > D \right) \\
&= P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\mu}\boldsymbol{\mu}^\top)\|_2 > D \right) \\
&\leq P\left(\frac{2^{3/2}}{\|\boldsymbol{\mu}\|_2^2} \left\| \frac{1}{m} \sum_{i=i}^{m} \boldsymbol{X}_i \boldsymbol{X}_i^\top - \boldsymbol{\mu}\boldsymbol{\mu}^\top \right\|_{\mathrm{op}} > D \right) \\
&\leq P\left(\frac{2^{3/2}}{\|\boldsymbol{\mu}\|_2^2} \left(2\|\boldsymbol{\mu}\|_2 \left\| \frac{1}{m} \sum_{i=1}^{m} \theta_i \boldsymbol{g}_i^{(\theta_i)} \right\|_2 + \left\| \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{g}_i^{(\theta_i)} \boldsymbol{g}_i^{(\theta_i)\top} \right\|_{\mathrm{op}} \right) > D \right) \\
&\leq P\left(\left\{ \frac{2^{3/2}}{\|\boldsymbol{\mu}\|_2^2} \left(2\|\boldsymbol{\mu}\|_2 \left\| \frac{1}{m} \sum_{i=1}^{m} \theta_i \boldsymbol{g}_i^{(\theta_i)} \right\|_2 + \left\| \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{g}_i^{(\theta_i)} \boldsymbol{g}_i^{(\theta_i)\top} \right\|_{\mathrm{op}} \right) > D \right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \right) \\
&+ P\left((\mathrm{E}_1 \cup \mathrm{E}_2)^c \right) \\
&\leq P(D > D) + P\left(\mathrm{E}_1^c \right) + P\left(\mathrm{E}_2^c \right) \\
&= 1 - P(\mathrm{E}_1) + 1 - P(\mathrm{E}_2) \\
&\leq 4\mathrm{e}^{-n},
\end{aligned}$$

where $\boldsymbol{\gamma}_1(\boldsymbol{\mu}\boldsymbol{\mu}^\top) = \boldsymbol{\gamma}_1(\boldsymbol{\Sigma})$, $\lambda_1(\boldsymbol{\mu}\boldsymbol{\mu}^\top) = \|\boldsymbol{\mu}\|_2^2$, and $\lambda_2(\boldsymbol{\mu}\boldsymbol{\mu}^\top) = 0$ are used. □

5.5 Proof of Theorem 4.3

We see that

$$\begin{aligned}
&P\Big(\theta_i \langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0 \Big) \\
&= P\Big(\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0 \Big| \theta_i = 1 \Big) P\Big(\theta_i = 1 \Big) \\
&\quad + P\Big(\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle > 0 \Big| \theta_i = -1 \Big) P\Big(\theta_i = -1 \Big) \\
&= \frac{1}{2} P\Big(\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle < 0 \Big| \theta_i = 1 \Big) + \frac{1}{2} P\Big(\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i \rangle > 0 \Big| \theta_i = -1 \Big) \\
&\leq \frac{1}{2} \left(\Phi\left(-\frac{(2 - \delta^2 - 2\alpha)\|\boldsymbol{\mu}\|_2}{(2+\delta^2)\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}} \right) + \Phi\left(-\frac{(2 - \delta^2 - 2\alpha)\|\boldsymbol{\mu}\|_2}{(2+\delta^2)\sqrt{\lambda_1(\boldsymbol{\Sigma}_2)}} \right) \right) + 2(4 + \mathrm{e})\mathrm{e}^{-n} \\
&\leq \Phi\left(-\frac{(2 - \delta^2 - 2\alpha)\|\boldsymbol{\mu}\|_2}{(2+\delta^2)\sqrt{\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}} \right) + 2(4 + \mathrm{e})\mathrm{e}^{-n},
\end{aligned}$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

where the second-last inequality is a consequence of Lemma 5.4. □

Lemma 5.4 *Consider constants C, K, and c_1 in Theorem 4.3. Define*

$$\delta = \min\left\{2^{5/2}c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m\|\boldsymbol{\mu}\|_2^2}} + \sqrt{2}\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1)+\lambda_1(\boldsymbol{\Sigma}_2)}{\|\boldsymbol{\mu}\|_2^2}\right)\right.$$
$$\cdot\left(CK^2\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)+1\right),$$
$$\left.\frac{2^{3/2}CK^2\left(2\|\boldsymbol{\mu}\|_2^2+\lambda_1(\boldsymbol{\Sigma}_1)+\lambda_1(\boldsymbol{\Sigma}_2)\right)}{2\|\boldsymbol{\mu}\|_2^2+\lambda_1(\boldsymbol{\Sigma}_1)+\lambda_1(\boldsymbol{\Sigma}_2)-(\lambda_2(\boldsymbol{\Sigma}_1)+\lambda_2(\boldsymbol{\Sigma}_2))}\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)\right\}.$$

If δ, n, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}_1$ satisfy

$$\delta\frac{c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)}}{\|\boldsymbol{\mu}\|_2} \leq \alpha \leq 1-\frac{\delta^2}{2}$$

for some $\alpha \in (0,1)$, then it holds that

$$P\left(\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i\rangle < 0 \middle| \theta_i = 1\right)$$
$$\leq \Phi\left(-\frac{(2-\delta^2-2\alpha)|\boldsymbol{\mu}|_2}{\left(2+\delta^2\right)\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}}\right) + 2(4+\mathrm{e})\mathrm{e}^{-n} \tag{27}$$

for $i = 1,\ldots,m$. If δ, n, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}_2$ satisfy

$$\delta\frac{c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_2)}}{\|\boldsymbol{\mu}\|_2} \leq \alpha \leq 1-\frac{\delta^2}{2}$$

for some $\alpha \in (0,1)$, then it holds that

$$P\left(\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i\rangle > 0 \middle| \theta_i = -1\right)$$
$$\leq \Phi\left(-\frac{(2-\delta^2-2\alpha)|\boldsymbol{\mu}|_2}{\left(2+\delta^2\right)\sqrt{\lambda_1(\boldsymbol{\Sigma}_2)}}\right) + 2(4+\mathrm{e})\mathrm{e}^{-n} \tag{28}$$

for $i = 1,\ldots,m$.

This lemma provides an evaluation of the conditional misclassification probability, which is the key to proving Theorem 4.3. In the next subsection, we prove Lemma 5.4.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

5.6 Proof of Lemma 5.4

We only prove (27), because the proof of (28) is similar.

Fix $i \in \{1, \ldots, m\}$. The definition of δ yields

$$\delta \leq \frac{\alpha \|\boldsymbol{\mu}\|_2}{c_1 \sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)}}.$$

For simplicity, we define the events A_i, E_1, and E_2 as follows:

$$\mathrm{A}_i = \{\theta_i = 1\}, \ \mathrm{E}_1 = \left\{\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2 \leq \delta\right\}, \text{ and } \mathrm{E}_2 = \left\{\|\boldsymbol{r}_i\|_2 \leq c_1\sqrt{n-1}\right\},$$

where

$$\boldsymbol{r}_i = \left(\frac{\left\langle \boldsymbol{g}_i^{(1)}, \boldsymbol{\gamma}_2(\boldsymbol{\Sigma}_1)\right\rangle}{\sqrt{\lambda_2(\boldsymbol{\Sigma}_1)}}, \ldots, \frac{\left\langle \boldsymbol{g}_i^{(1)}, \boldsymbol{\gamma}_n(\boldsymbol{\Sigma}_1)\right\rangle}{\sqrt{\lambda_n(\boldsymbol{\Sigma}_1)}}\right)^\top.$$

Then, we have

$$\begin{aligned} &P\left(\left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i\right\rangle < 0 | \mathrm{A}_i\right) \\ &\quad \leq P\left(\left\{\left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i\right\rangle < 0\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 | \mathrm{A}_i\right) + P\left(\mathrm{E}_1^c | \mathrm{A}_i\right) + P\left(\mathrm{E}_2^c | \mathrm{A}_i\right). \end{aligned} \tag{29}$$

for the first term on the right-hand side of (29), we have

$$\begin{aligned} &P\left(\left\{\left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \boldsymbol{X}_i\right\rangle < 0\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \,\middle|\, \mathrm{A}_i\right) \\ &= P\left(\left\{\left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m), \theta_i \boldsymbol{\mu} + \boldsymbol{g}_i^{(\theta_i)}\right\rangle < 0\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \,\middle|\, \mathrm{A}_i\right) \\ &= P\left(\left\{\left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1) + \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{\mu} + \boldsymbol{g}_i^{(\theta_i)}\right\rangle < 0\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \,\middle|\, \mathrm{A}_i\right) \\ &= P\left(\left\{\left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{\mu}\right\rangle + \left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{g}_i^{(\theta_i)}\right\rangle + \left\langle \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{\mu}\right\rangle \right. \\ &\quad \left. + \left\langle \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{g}_i^{(\theta_i)}\right\rangle < 0\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \,\middle|\, \mathrm{A}_i\right) \\ &= P\left(\left\{\left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{\mu}\right\rangle + \left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \left\langle \boldsymbol{g}_i^{(\theta_i)}, \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\rangle \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\rangle \right. \right. \\ &\quad \left. + \left\langle \boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \sum_{j=2}^{n} \left\langle \boldsymbol{g}_i^{(\theta_i)}, \boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle \boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle + \left\langle \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{\mu}\right\rangle \right. \\ &\quad \left. \left. + \left\langle \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1), \boldsymbol{g}_i^{(\theta_i)}\right\rangle < 0\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \,\middle|\, \mathrm{A}_i\right) \end{aligned}$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

$$\leq P\Bigg(\Bigg\{-\frac{1}{2}\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2^2\|\boldsymbol{\mu}\|_2-\frac{1}{2}\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2^2\left|\left\langle\boldsymbol{g}_i^{(\theta_i)},\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\rangle\right| + \left\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\sum_{j=2}^{n}\left\langle\boldsymbol{g}_i^{(\theta_i)},\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle+\|\boldsymbol{\mu}\|_2 + \left\langle\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\boldsymbol{g}_i^{(\theta_i)}\right\rangle<0\Bigg\}\cap \mathrm{E}_1\cap\mathrm{E}_2\,\Bigg|\,\mathrm{A}_i\Bigg), \tag{30}$$

where $\boldsymbol{g}_i^{(\theta_i)}=\sum_{j=1}^{n}\left\langle\boldsymbol{g}_i^{(\theta_i)},\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)$, $\langle\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\boldsymbol{\mu}\rangle=\|\boldsymbol{\mu}\|_2$, and $\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\rangle=-\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2^2/2$ are used. From the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
&\left\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\sum_{j=2}^{n}\left\langle\boldsymbol{g}_i^{(1)},\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle\\
&\quad\geq-\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2\left\|\sum_{j=2}^{n}\left\langle\boldsymbol{g}_i^{(1)},\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\|_2\\
&\quad=-\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2\sqrt{\sum_{j=2}^{n}\left\langle\boldsymbol{g}_i^{(1)},\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle^2}\\
&\quad\geq-\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2\sqrt{\lambda_2(\boldsymbol{\Sigma}_1)}\sqrt{\sum_{j=2}^{n}\left(\frac{1}{\sqrt{\lambda_j(\boldsymbol{\Sigma}_1)}}\left\langle\boldsymbol{g}_i^{(1)},\boldsymbol{\gamma}_j(\boldsymbol{\Sigma}_1)\right\rangle\right)^2}\\
&\quad=-\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2\sqrt{\lambda_2(\boldsymbol{\Sigma}_1)}\|\boldsymbol{r}_i\|_2,
\end{aligned}$$

where the orthogonality of $\boldsymbol{\gamma}_2(\boldsymbol{\Sigma}_1),\ldots,\boldsymbol{\gamma}_n(\boldsymbol{\Sigma}_1)$ is used in the first equality. Consider the following $n\times n$ matrix $\boldsymbol{Q}$:

$$\boldsymbol{Q}=\left[\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\ \boldsymbol{\gamma}_2(\boldsymbol{\Sigma}_1)\ \ldots\ \boldsymbol{\gamma}_n(\boldsymbol{\Sigma}_1)\right].$$

Then,

$$\begin{pmatrix}\left\langle\boldsymbol{g}_i^{(1)},\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\rangle\\ \left\langle\boldsymbol{g}_i^{(1)},\boldsymbol{\gamma}_2(\boldsymbol{\Sigma}_1)\right\rangle\\ \vdots\\ \left\langle\boldsymbol{g}_i^{(1)},\boldsymbol{\gamma}_n(\boldsymbol{\Sigma}_1)\right\rangle\end{pmatrix}=\boldsymbol{Q}^\top\boldsymbol{g}_i^{(1)}\sim\mathcal{N}_n(\boldsymbol{0}_n,\boldsymbol{Q}^\top\boldsymbol{\Sigma}_1\boldsymbol{Q}),$$

where $\boldsymbol{Q}^\top\boldsymbol{\Sigma}_1\boldsymbol{Q}=\mathrm{diag}(\lambda_1(\boldsymbol{\Sigma}_1),\ldots,\lambda_n(\boldsymbol{\Sigma}_1))$. Therefore, each component of $\boldsymbol{Q}^\top\boldsymbol{g}_i^{(1)}$ is independent. Thus, $\boldsymbol{r}_i\sim\mathcal{N}_{n-1}(\boldsymbol{0}_{n-1},\boldsymbol{I}_{n-1})$. We evaluate the upper bound of (30)

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

as follows:

$$
\begin{aligned}
&P\Bigg(\Bigg\{-\frac{1}{2}\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2^2\|\boldsymbol{\mu}\|_2 \\
&\quad-\frac{1}{2}\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2^2\left|\left\langle \boldsymbol{g}_i^{(\theta_i)},\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\rangle\right| \\
&\quad-\sqrt{\lambda_2(\boldsymbol{\Sigma}_1)}\left\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\right\|_2\|\boldsymbol{r}_i\|_2 \\
&\quad+\|\boldsymbol{\mu}\|_2+\left\langle\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\boldsymbol{g}_i^{(\theta_i)}\right\rangle<0\Bigg\}\cap \mathrm{E}_1\cap \mathrm{E}_2\,\Big|\,\mathrm{A}_i\Bigg) \\
&\le P\Bigg(\Bigg\{-\frac{\delta^2}{2}\|\boldsymbol{\mu}\|_2-\frac{\delta^2}{2}\left|\left\langle\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\boldsymbol{g}_i^{(\theta_i)}\right\rangle\right|-\delta c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)}+\|\boldsymbol{\mu}\|_2 \\
&\quad+\left\langle\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1),\boldsymbol{g}_i^{(\theta_i)}\right\rangle<0\Bigg\}\cap \mathrm{E}_1\cap \mathrm{E}_2\,\Big|\,\mathrm{A}_i\Bigg) \\
&= P\Bigg(\Bigg\{-\frac{\delta^2}{2}\|\boldsymbol{\mu}\|_2-\frac{\delta^2}{2}\left|\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2\left\langle\frac{\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)}{\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2},\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right\rangle\right| \\
&\quad-\delta c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)}+\|\boldsymbol{\mu}\|_2 \\
&\quad+\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2\left\langle\frac{\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)}{\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2},\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right\rangle<0 \\
&\quad\Bigg\}\cap \mathrm{E}_1\cap \mathrm{E}_2\,\Big|\,\mathrm{A}_i\Bigg) \\
&= P\Bigg(\Bigg\{-\frac{\delta^2}{2}\|\boldsymbol{\mu}\|_2-\frac{\delta^2}{2}\left|\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}\left\langle\frac{\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)}{\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2},\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right\rangle\right| \\
&\quad-\delta c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)} \\
&\quad+\|\boldsymbol{\mu}\|_2+\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}\left\langle\frac{\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)}{\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2},\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\right\rangle<0\Bigg\}\cap \mathrm{E}_1\cap \mathrm{E}_2\,\Big|\,\mathrm{A}_i\Bigg),
\end{aligned}
\tag{31}
$$

where $\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2=\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}$ is used in the last equality. Let

$$
\boldsymbol{g}_i=\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{g}_i^{(\theta_i)}\ \text{and}\ Z_i=\left\langle\frac{\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)}{\|\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2},\boldsymbol{g}_i\right\rangle.
$$

We have $\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}Z_i|\mathrm{A}_i\sim\mathcal{N}(0,\lambda_1(\boldsymbol{\Sigma}_1))$ because $\boldsymbol{g}_i|\mathrm{A}_i\sim\mathcal{N}_n(\boldsymbol{0}_n,\boldsymbol{I}_n)$. The right-hand side of (31) can be evaluated as follows:

$$
P\left(\left\{\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}(Z_i-\frac{\delta^2}{2}|Z_i|)<-(1-\frac{\delta^2}{2})\|\boldsymbol{\mu}\|_2+\delta c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)}\right\}\cap \mathrm{E}_1\cap \mathrm{E}_2\,\Big|\,\mathrm{A}_i\right)
$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

$$
\begin{aligned}
&\le P\Bigg(\Bigg\{\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}(Z_i - \frac{\delta^2}{2}|Z_i|) < -(1-\frac{\delta^2}{2})\|\boldsymbol{\mu}\|_2 + \frac{\alpha\|\boldsymbol{\mu}\|_2}{c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)}} \\
&\cdot c_1\sqrt{(n-1)\lambda_2(\boldsymbol{\Sigma}_1)}\Bigg\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \,\Big|\, \mathrm{A}_i\Bigg) \\
&= P\left(\left\{\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}(Z_i - \frac{\delta^2}{2}|Z_i|) < -(1-\frac{\delta^2}{2}-\alpha)\|\boldsymbol{\mu}\|_2\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \,\Big|\, \mathrm{A}_i\right) \\
&= P\left(\left\{\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}(Z_i + \frac{\delta^2}{2}Z_i) < -(1-\frac{\delta^2}{2}-\alpha)\|\boldsymbol{\mu}\|_2\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \cap \{Z_i < 0\} \,\Big|\, \mathrm{A}_i\right) \\
&\quad + P\left(\left\{\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}(Z_i - \frac{\delta^2}{2}Z_i) < -(1-\frac{\delta^2}{2}-\alpha)\|\boldsymbol{\mu}\|_2\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \cap \{Z_i \ge 0\} \,\Big|\, \mathrm{A}_i\right) \\
&= P\left(\left\{\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}(1+\frac{\delta^2}{2})Z_i < -(1-\frac{\delta^2}{2}-\alpha)\|\boldsymbol{\mu}\|_2\right\} \cap \mathrm{E}_1 \cap \mathrm{E}_2 \cap \{Z_i < 0\} \,\Big|\, \mathrm{A}_i\right) \\
&\le P\left(Z_i < -\frac{(2-\delta^2-2\alpha)\|\boldsymbol{\mu}\|_2}{(2+\delta^2)\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}} \,\Big|\, \mathrm{A}_i\right), \qquad (32)
\end{aligned}
$$

where $1-\delta^2/2-\alpha > 0$ is used in the last equality. The second term on the right-hand side of (29) is evaluated separately for the following two cases.

(i) First, we consider the case where

$$
\delta = \frac{2^{3/2}CK^2\left(2\|\boldsymbol{\mu}\|_2^2 + \lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2)\right)}{2\|\boldsymbol{\mu}\|_2^2 + \lambda_1(\boldsymbol{\Sigma}_1) + \lambda_1(\boldsymbol{\Sigma}_2) - (\lambda_2(\boldsymbol{\Sigma}_1) + \lambda_2(\boldsymbol{\Sigma}_2))}\left(\sqrt{\frac{2n}{m}} + \frac{2n}{m}\right).
$$

Using (4), (5), (6), and the Davis–Kahan theorem, we have

$$
\begin{aligned}
&P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2 > \delta \,\Big|\, \mathrm{A}_i\right) \\
&= P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m) - \boldsymbol{\gamma}_1(\boldsymbol{\Sigma})\|_2 > \delta \,\Big|\, \mathrm{A}_i\right) \\
&\le P\left(\frac{2^{3/2}\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}}}{\lambda_1(\boldsymbol{\Sigma}) - \lambda_2(\boldsymbol{\Sigma})} > \delta \,\Big|\, \mathrm{A}_i\right) \\
&\le P\left(\frac{2^{3/2}\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}}}{\|\boldsymbol{\mu}\|_2^2 + \frac{1}{2}\lambda_1(\boldsymbol{\Sigma}_1) - \frac{1}{2}\lambda_2(\boldsymbol{\Sigma}_1) + \frac{1}{2}\lambda_1(\boldsymbol{\Sigma}_2) - \frac{1}{2}\lambda_2(\boldsymbol{\Sigma}_2)} > \delta \,\Big|\, \mathrm{A}_i\right) \\
&= P\left(\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}} > \frac{\delta}{2^{3/2}}\left(\|\boldsymbol{\mu}\|_2^2 + \frac{1}{2}\lambda_1(\boldsymbol{\Sigma}_1) - \frac{1}{2}\lambda_2(\boldsymbol{\Sigma}_1)\right.\right. \\
&\quad \left.\left. + \frac{1}{2}\lambda_1(\boldsymbol{\Sigma}_2) - \frac{1}{2}\lambda_2(\boldsymbol{\Sigma}_2)\right) \,\Big|\, \mathrm{A}_i\right).
\end{aligned}
$$

The definition of δ and Proposition 4.2 with $u = n$ yield

$$
P\left(\|\boldsymbol{S}_m - \boldsymbol{\Sigma}\|_{\mathrm{op}} > \frac{\delta}{2^{3/2}}\left(\|\boldsymbol{\mu}\|_2^2 + \frac{1}{2}\lambda_1(\boldsymbol{\Sigma}_1) - \frac{1}{2}\lambda_2(\boldsymbol{\Sigma}_1)\right.\right.
$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

$$
\begin{aligned}
&+\frac{1}{2}\lambda_1(\boldsymbol{\Sigma}_2)-\frac{1}{2}\lambda_2(\boldsymbol{\Sigma}_2)\Big)\,\Big|\,\mathrm{A}_i\Big)\\
&=P\Big(\|\boldsymbol{S}_m-\boldsymbol{\Sigma}\|_{\mathrm{op}}>CK^2\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)\\
&\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1)}{2}+\frac{\lambda_1(\boldsymbol{\Sigma}_2)}{2}+\|\boldsymbol{\mu}\|_2^2\right)\Big|\,\mathrm{A}_i\Big)\\
&=\frac{1}{P\left(\mathrm{A}_i\right)}P\left(\|\boldsymbol{S}_m-\boldsymbol{\Sigma}\|_{\mathrm{op}}>CK^2\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)\right.\\
&\left.\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1)}{2}+\frac{\lambda_1(\boldsymbol{\Sigma}_2)}{2}+\|\boldsymbol{\mu}\|_2^2\right)\cap\mathrm{A}_i\right)\\
&\le 2P\left(\|\boldsymbol{S}_m-\boldsymbol{\Sigma}\|_{\mathrm{op}}>CK^2\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1)}{2}+\frac{\lambda_1(\boldsymbol{\Sigma}_2)}{2}+\|\boldsymbol{\mu}\|_2^2\right)\right)\\
&\le 4\mathrm{e}^{-n}.
\end{aligned}
$$

Consequently, it follows that

$$
P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2>\delta\,\Big|\,\mathrm{A}_i\right)\le 4\mathrm{e}^{-n}. \tag{33}
$$

(ii) Second, we consider the case where

$$
\begin{aligned}
\delta=\frac{2^{3/2}}{\|\boldsymbol{\mu}\|_2^2}\Bigg(&2\|\boldsymbol{\mu}\|_2c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}+\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1)+\lambda_1(\boldsymbol{\Sigma}_2)}{2}\right)\\
&\cdot\left(CK^2\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)+1\right)\Bigg).
\end{aligned}
$$

By Lemma 5.3, we have

$$
\begin{aligned}
&P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2>\delta\,\Big|\,\mathrm{A}_i\right)\\
&=P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma})\|_2>\delta\,\Big|\,\mathrm{A}_i\right)\\
&=\frac{1}{P\left(\mathrm{A}_i\right)}P\left(\left\{\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma})\|_2>\delta\right\}\cap\mathrm{A}_i\right)\\
&\le 2P\Bigg(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma})\|_2>\frac{2^{\frac{3}{2}}}{\|\boldsymbol{\mu}\|_2^2}\Bigg(2\|\boldsymbol{\mu}\|_2c_1\sqrt{\frac{n\max_{j=1,2}\{\lambda_1(\boldsymbol{\Sigma}_j)\}}{m}}\\
&\quad+\left(\frac{\lambda_1(\boldsymbol{\Sigma}_1)+\lambda_1(\boldsymbol{\Sigma}_2)}{2}\right)\left(CK^2\left(\sqrt{\frac{2n}{m}}+\frac{2n}{m}\right)+1\right)\Bigg)\Bigg)\\
&\le 8\mathrm{e}^{-n},
\end{aligned}
\tag{34}
$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

where (4) is used to obtain the first equality.

Therefore, using (33) and (34), we have

$$P\left(\|\boldsymbol{\gamma}_1(\boldsymbol{S}_m)-\boldsymbol{\gamma}_1(\boldsymbol{\Sigma}_1)\|_2>\delta\,\middle|\,\mathrm{A}_i\right)\le 8\mathrm{e}^{-n}. \tag{35}$$

For the third term on the right-hand side of (29), recalling that $c_1=1+K_g^2/\sqrt{c}$, we have

$$\begin{aligned}P(\mathrm{E}_2^c|\mathrm{A}_i)&=P(|\|\boldsymbol{r}_i\|_2-\sqrt{n-1}+\sqrt{n-1}|>c_1\sqrt{n-1}|\mathrm{A}_i)\\&\le P(|\|\boldsymbol{r}_i\|_2-\sqrt{n-1}|>(c_1-1)\sqrt{n-1}|\mathrm{A}_i)\\&\le 2\mathrm{e}^{-(n-1)}.\end{aligned} \tag{36}$$

Therefore, using (32), (35), and (36), we have

$$\begin{aligned}&P\left(\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle<0\,\middle|\,\mathrm{A}_i\right)\\&\le P\left(Z_i<-\frac{(2-\delta^2-2\alpha)|\boldsymbol{\mu}|_2}{(2+\delta^2)\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}}\,\middle|\,\mathrm{A}_i\right)+2P(\mathrm{E}_1^c)+P(\mathrm{E}_2^c)\\&\le\Phi\left(-\frac{(2-\delta^2-2\alpha)\|\boldsymbol{\mu}\|_2}{(2+\delta^2)\sqrt{\lambda_1(\boldsymbol{\Sigma}_1)}}\right)+8\mathrm{e}^{-n}+2\mathrm{e}^{-(n-1)}.\end{aligned}$$

This completes the proof. □

5.7 Proof of Corollary 4.5

Consider m, n, and η satisfying (9). Then, by using the Bonferroni inequality and (11), we have

$$\begin{aligned}&P\left(\left\{\bigcap_{i=1}^m\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle>0\}\right\}\cup\left\{\bigcap_{i=1}^m\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle<0\}\right\}\right)\\&\ge P\left(\bigcap_{i=1}^m\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle>0\}\right)=1-P\left(\bigcup_{i=1}^m\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle\le 0\}\right)\\&\ge 1-\sum_{i=1}^m P\left(\{\theta_i\langle\boldsymbol{\gamma}_1(\boldsymbol{S}_m),\boldsymbol{X}_i\rangle\le 0\}\right)\\&\ge 1-\frac{(2+\delta^2)m}{(2-\delta^2-2\alpha)\sqrt{2\pi\eta}}\exp\left(-\left(\frac{2-\delta^2-2\alpha}{2+\delta^2}\right)^2\frac{\eta}{2}\right)-2(4+\mathrm{e})m\mathrm{e}^{-n}.\end{aligned} \tag{37}$$

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

The second and third terms on the right-hand side of (37) converge to 0 as $n \to \infty$ with (14) and (15). □

6 Concluding Remarks

In this paper, we have derived non-asymptotic bounds for the error probability of the spectral clustering algorithm when considering the mixture distribution of two multivariate normal distributions that form the allometric extension relationship. In future research, we would like to relax the assumption of the normal distribution to the sub-Gaussian distribution and to consider weights other than $\pi_1 = \pi_2 = 1/2$ in the mixture distribution. We are also interested in developing a clustering method for multi-group settings based on the spectral clustering algorithm, and discussing it theoretically. In addition, it would be interesting to discuss the sufficient conditions for the consistency of clustering in detail because there exists a case where the sufficient conditions in existing studies are milder; see Remark 12.

Acknowledgements This study was supported in part by Japan Society for the Promotion of Science KAKENHI Grant Numbers 21K13836 (KT) and 23K16851 (YG). The authors would like to thank anonymous referees who provided comments that improve the manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Abbe E, Fan J, Wang K (2022) An ℓ_p theory of PCA and spectral clustering. Ann Stat 50(4):2359–2385

Amit S, Mukesh P, Akshansh G, Neha B, Om PP, Aruna T, Meng JE, Weiping D, Chin-Teng L (2017) A review of clustering techniques and developments. Neurocomputing 267:664–681

Bartoletti S, Flury BD, Nel DG (1999) Allometric extension. Biometrics 55(4):1210–1214

Borysov P, Hannig J, Marron JS (2014) Asymptotics of hierarchical clustering for growing dimension. J Multivar Anal 124:465–479

Cai TT, Zhang A (2018) Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. Ann Stat 46(1):60–89

Flury B (1997) A first course in multivariate statistics. Springer, New York

Hills M (2006) Allometry. In: Kotz S, Read CB, Balakrishnan N, Vidakovic B, Johnson NL (eds) Encyclopedia of statistical sciences. https://doi.org/10.1002/0471667196.ess0033.pub2. Accessed 1 June 2023

Hsu D, Kakade S (2013) Learning mixtures of spherical Gaussians: moment methods and spectral decompositions. In: ITCS'13—proceedings of the 2013 ACM conference on innovations in theoretical computer science. ACM, New York, pp 11–19

Kurata H, Hoshino T, Fujikoshi Y (2008) Allometric extension model for conditional distributions. J Multivar Anal 99(9):1985–1998

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Löffler M, Zhang AY, Zhou HH (2021) Optimality of spectral clustering in the Gaussian mixture model. Ann Stat 49(5):2506–2530
Matsuura S, Kurata H (2014) Principal points for an allometric extension model. Stat Pap 55(3):853–870
Myers TC, de Mello P, Glor RE (2020a) A morphometric assessment of species boundaries in a widespread anole lizard (Squamata: Dactyloidae). Biol J Linn Soc 130(4):813–825
Myers TC, de Mello P, Glor RE (2020b) A morphometric assessment of species boundaries in a widespread anole lizard (Squamata: Dactyloidae) [Dataset]. Dryad. https://doi.org/10.5061/dryad.qfttdz0dq. Accessed 14 Dec 2024
Ndaoud M (2022) Sharp optimal recovery in the two component Gaussian mixture model. Ann Stat 50(4):2096–2126
O'Neill TJ (1978) Normal discrimination with unclassified observations. J Am Stat Assoc 73(364):821–826
Pollard D (1981) Strong consistency of k-means clustering. Ann Stat 9(1):135–140
Pollard D (1982) A central limit theorem for k-means clustering. Ann Probab 10(4):919–926
R Core Team (2023) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Scrucca L, Fop M, Murphy TB, Raftery AE (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J 8(1):289–317
Scrucca L, Fraley C, Murphy TB, Raftery AE (2023) Model-based clustering, classification, and density estimation using mclust in R. Chapman and Hall/CRC, Boca Raton
Tarpey T (2007) Linear transformations and the k-means clustering algorithm. Am. Stat. 61(1):34–40
Tsukuda K, Matsuura S (2023) High-dimensional hypothesis testing for allometric extension model. J. Multivar. Anal. 197:105208
Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York
Vershynin R (2018) High-dimensional probability. An introduction with applications in data science. Cambridge University Press, Cambridge

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com