TWO-SAMPLE TESTS FOR HIGH-DIMENSION, STRONGLY SPIKED EIGENVALUE MODELS

Makoto Aoshima and Kazuyoshi Yata

University of Tsukuba

Abstract: We consider two-sample tests for high-dimensional data under two disjoint models: the strongly spiked eigenvalue (SSE) model and the non-SSE (NSSE) model. We provide a general test statistic as a function of a positive-semidefinite matrix. We give sufficient conditions for the test statistic to satisfy a consistency property and to be asymptotically normal. We discuss an optimality of the test statistic under the NSSE model. We also investigate the test statistic under the SSE model by considering strongly spiked eigenstructures and create a new effective test procedure for the SSE model. Finally, we discuss the performance of the classifiers numerically.

Key words and phrases: Asymptotic normality, eigenstructure estimation, large $p$ small $n$, noise reduction methodology, spiked model.

1. Introduction

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. This is the so-called “HDLSS” or “large $p$, small $n$” data, where $p$ is the data dimension, $n$ is the sample size and $p/n \to \infty$. Statistical inference on this type of data is becoming increasingly relevant, especially in the areas of medical diagnostics, engineering, and other big data. Suppose we have independent samples of $p$-variate random variables from populations $\pi_i$, $i = 1, 2$, with unknown mean vectors $\mu_i$ and unknown positive-definite covariance matrices $\Sigma_i$. We do not assume the normality of the population distributions. The eigen-decomposition of $\Sigma_i$ ($i = 1, 2$) is given by $\Sigma_i = H_i \Lambda_i H_i^T = \sum_{j=1}^p \lambda_{ij} h_{ij} h_{ij}^T$, where $\Lambda_i = \text{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$ is a diagonal matrix of eigenvalues, $\lambda_{i1} \geq \cdots \geq \lambda_{ip} > 0$, and $H_i = [h_{i1}, \ldots, h_{ip}]$ is an orthogonal matrix of the corresponding eigenvectors. Note that $\lambda_{i1}$ is the largest eigenvalue of $\Sigma_i$ for $i = 1, 2$. For the eigenvalues, we consider two disjoint models: the strongly spiked eigenvalue (SSE) model, which will be defined by (1.6), and the non-SSE (NSSE) model, which will be defined by (1.4).
In this paper, we consider the two-sample test:

\[ H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2. \]  

(1.1)

Having recorded i.i.d. samples, \( x_{ij}, \ j = 1, \ldots, n_i, \) of size \( n_i \) from each \( \pi_i, \) we define \( \bar{x}_{in} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \) and \( S_{in} = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{in})(x_{ij} - \bar{x}_{in})^T/(n_i - 1) \) for \( i = 1, 2. \) We assume \( n_i \geq 4 \) for \( i = 1, 2. \) Hotelling’s \( T^2 \)-statistic is

\[ T^2 = \left( n_1 + n_2 \right)^{-1} n_1 n_2 (\bar{x}_{1n} - \bar{x}_{2n})^T S^{-1} (\bar{x}_{1n} - \bar{x}_{2n}), \]

where \( S = \left\{ (n_1 - 1)S_{1n_1} + (n_2 - 1)S_{2n_2} \right\}/(n_1 + n_2 - 2). \) However, \( S^{-1} \) does not exist in such HDLSS contexts as \( p/n_i \to \infty, \ i = 1, 2. \) In such situations, \( \text{Dempster, 1958} \) and \( \text{Srivastava, 2007} \) considered the test when \( \pi_1 \) and \( \pi_2 \) are Gaussian. When \( \pi_1 \) and \( \pi_2 \) are non-Gaussian, \( \text{Bai and Saranadasa, 1996} \) and \( \text{Cai, Liu and Xia, 2014} \) considered the test under homoscedasticity, \( \Sigma_1 = \Sigma_2. \)

\( \text{Chen and Qin, 2010} \) and \( \text{Aoshima and Yata, 2011, 2015} \) considered the test under heteroscedasticity, \( \Sigma_1 \neq \Sigma_2. \)

In this paper, we first consider a test statistic with a positive-semidefinite matrix \( A \) of dimension \( p:\)

\[ T(A) = (\bar{x}_{1n} - \bar{x}_{2n})^T A(\bar{x}_{1n} - \bar{x}_{2n}) - \frac{2}{n_i} \text{tr}(S_{in} A) \]

\[ = 2 \sum_{i=1}^{2} \sum_{j<j'} x_{ij}^T A x_{ij'} / n_i(n_i - 1) - 2 \bar{x}_{1n}^T A \bar{x}_{2n}. \]  

(1.2)

Note that \( E\{T(A)\} = (\mu_1 - \mu_2)^T A (\mu_1 - \mu_2). \) Let \( I_p \) denote the identity matrix of dimension \( p. \) We note that \( T(I_p) \) is equivalent to the statistics given by \( \text{Chen and Qin, 2010} \) and \( \text{Aoshima and Yata, 2011}. \) We call the test with \( T(I_p) \) the “distance-based two-sample test”. In Section 3, we discuss a choice of \( A. \)

We consider the divergence condition \( p \to \infty, \ n_1 \to \infty \) and \( n_2 \to \infty, \) that is equivalent to

\[ m \to \infty, \quad \text{where} \quad m = \min\{p, n_{\min}\} \quad \text{with} \quad n_{\min} = \min\{n_1, n_2\}. \]

By using Theorem 1 in \( \text{Chen and Qin, 2010}, \) or Theorem 4 in \( \text{Aoshima and Yata, 2015}, \) we can claim that under \( H_0 \) in (1.1),

\[ T(I_p) \quad \{K_1(I_p)\}^{1/2} \Rightarrow N(0, 1) \quad \text{as} \quad m \to \infty \]  

(1.3)

if we assume (A-i), see Section 2, and the condition that

\[ \frac{\lambda_{ii}^2}{\text{tr}(\Sigma_i)} \to 0 \quad \text{as} \quad p \to \infty \quad \text{for} \ i = 1, 2. \]  

(1.4)
Here, $K_1(A)$ is defined in Section 2.1, $\Rightarrow$ denotes convergence in distribution and $N(0,1)$ denotes the standard normal. Thus, by using $T(I_p)$ and an estimate of $K_1(I_p)$, one can construct a test procedure of (1.1) for high-dimensional data. As discussed in Section 2 of Aoshima and Yata (2015), the distance-based two-sample test is quite flexible for high-dimension, non-Gaussian data. In Section 3, we investigate an optimality of the test statistic in (1.2) and discuss a choice of $A$.

**Remark 1.** If all $\lambda_{ij}$'s are bounded as $\limsup_{p \to \infty} \lambda_{ij} < \infty$ and $\liminf_{p \to \infty} \lambda_{ij} > 0$, (1.4) trivially holds. On the other hand, they often have a spiked model such as

$$
\lambda_{ij} = a_{ij}p^{\alpha_{ij}} \quad (j = 1, \ldots, t_i) \quad \text{and} \quad \lambda_{ij} = c_{ij} \quad (j = t_i + 1, \ldots, p),
$$

(1.5)

where the $a_{ij}$'s, $c_{ij}$'s and $\alpha_{ij}$'s are positive fixed constants and the $t_i$'s are positive fixed integers. If they satisfy (1.5), (1.4) holds when $\alpha_{i1} < 1/2$ for $i = 1, 2$. See Yata and Aoshima (2012) for the details.

For eigenvalues of high-dimensional data, Jung and Marron (2009), Yata and Aoshima (2012, 2013b), Onatski (2012), and Fan, Liao and Mincheva (2013) considered spiked models such that $\lambda_{ij} \to \infty$ as $p \to \infty$ for $j = 1, \ldots, k_i$, with some positive integer $k_i$. The above references show that spiked models are quite natural because the first several eigenvalues should be spiked for high-dimensional data. Hence, we consider the following situation as well:

$$
\liminf_{p \to \infty} \left\{ \frac{\lambda_{i1}^2}{\text{tr}(\Sigma_i^2)} \right\} > 0 \quad \text{for} \quad i = 1 \text{ or } 2.
$$

(1.6)

In (1.6), the first eigenvalue is more spiked than in (1.4). For example, (1.6) holds for the spiked model in (1.5) with $\alpha_{i1} \geq 1/2$. We call (1.6) the “strongly spiked eigenvalue (SSE) model”. We emphasize that the asymptotic normality in (1.3) is not satisfied under the SSE model. See Section 4.1. See also Katayama, Kano and Srivastava (2013) and Ma, Lan and Wang (2015). Recall that (1.3) holds under (1.4). We call (1.4) the “non-strongly spiked eigenvalue (NSSE) model”.

The organization of this paper is as follows. In Section 2, we give sufficient conditions for $T(A)$ to satisfy a consistency property and asymptotic normality. In Section 3, under the NSSE model, we give a test procedure with $T(A)$ and discuss the choice of $A$. In Section 4, under the SSE model, we investigate test procedures by considering strongly spiked eigenstructures. In Section 5, we create a new test procedure by estimating the eigenstructures for the SSE model. We show that the power of the new test procedure is much higher than the distance-based two-sample test for the SSE model. In Section 6, we discuss
the performance of the test procedures for the SSE model with simulations. In 
Section 7, we highlight the benefits of the new models. In the online supple-
mentary material, we give additional simulations, data analyses, and proofs of the 
theoretical results. We also provide a method to distinguish between the NSSE 
model and the SSE model, and estimate the required parameters.

2. Asymptotic Properties of $T(A)$

In this section, we give sufficient conditions for $T(A)$ to satisfy a consistency 
property and to be asymptotically normal. For a positive-semidefinite matrix 
$A$, we write the square root of $A$ as $A^{1/2}$. Let $x_{ij} = H_i A_i^{1/2} z_{ij} + \mu_i$, where 
$z_{ij} = (z_{i1j}, \ldots, z_{ipj})^T$ is considered as a sphered data vector having the zero mean 
vector and identity covariance matrix. We assume that the fourth moments of 
each variable in $z_{ij}$ are uniformly bounded. More specifically, we assume that 
$x_{ij} = \Gamma_i w_{ij} + \mu_i$ for $i = 1, 2; j = 1, \ldots, n_i$, (2.1) 
where $\Gamma_i$ is a $p \times r_i$ matrix for some $r_i \geq p$ such that $\Gamma_i \Gamma_i^T = \Sigma_i$, and $w_{ij}$, $j = 1, \ldots, n_i$, are i.i.d. random vectors having 
$E(w_{ij}) = 0$ and $Var(w_{ij}) = I_{r_i}$. Note 
that (2.1) includes the case that $\Gamma_i = H_i A_i^{1/2}$ and $w_{ij} = z_{ij}$. Refer to Bai and 
Saranadasa (1996), Chen and Qin (2010) and Aoshima and Yata (2015) for the 
details of the model. As for $w_{ij} = (w_{i1j}, \ldots, w_{ir_ij})^T$, we assume the following 
assumption for $\pi_i$, i, as necessary.

(A-i) The fourth moments of each variable in $w_{ij}$ are uniformly bounded, 
$E(w_{isj}^2 w_{itj}^2) = E(w_{isj}^2) E(w_{itj}^2)$ and $E(w_{isj} w_{itj} w_{iuj} w_{ivj}) = 0$ for all $s \neq t, u, v.$ 

When the $\pi_i$s are Gaussian, (A-i) naturally holds.

2.1. Consistency and asymptotic normality of $T(A)$

Let $\mu_A = A^{1/2}(\mu_1 - \mu_2)$, $\Sigma_{i,A} = A_i^{1/2} \Sigma_i A_i^{1/2}$, $i = 1, 2$, and $\Delta(A) = ||\mu_A||^2$, 
where $|| \cdot ||$ denotes the Euclidean norm. Let $K(A) = K_1(A) + K_2(A)$, where 
$K_1(A) = 2 \sum_{i=1}^2 \frac{tr(\Sigma_{i,A}^2)}{n_i(n_i - 1)} + 4 \frac{tr(\Sigma_{1,A} \Sigma_{2,A})}{n_1 n_2}$ and $K_2(A) = 4 \sum_{i=1}^2 \frac{\mu_{i,A}^T \Sigma_{i,A} \mu_A}{n_i}$. 

Here $E\{T(A)\} = \Delta(A)$ and $Var\{T(A)\} = K(A)$. Also, $\Delta(A) = 0$ under $H_0$ 
in (1.1). Let $\lambda_{max}(B)$ denote the largest eigenvalue of a positive-semidefinite 
matrix, $B$. We assume the following condition of the $\Sigma_{i,A}$’s, as necessary.

(A-ii) $\frac{\lambda_{max}(\Sigma_{i,A})^2}{tr(\Sigma_{i,A}^2)} \to 0$ as $p \to \infty$ for $i = 1, 2.$ 

When $A = I_p$, (A-ii) is [1.4]. We assume one of the following conditions, as necessary.

(A-iii) $\lim_{m \to \infty} \frac{K_1(A)}{\{\Delta(A)\}^2} = 0$;

(A-iv) $\limsup_{m \to \infty} \frac{\{\Delta(A)\}^2}{K_1(A)} < \infty$;

(A-v) $\frac{K_1(A)}{K_2(A)} \to 0$ as $m \to \infty$.

Note that (A-iv) holds under $H_0$ in (1.1). If $\Sigma_1 = \Sigma_2$ (or $\Sigma$, say), (A-iii) holds when $\text{tr}\{(\Sigma A)^2\}/\{n_{\min}\Delta(A)\}^2 \to 0$ as $m \to \infty$. On the other hand, (A-iv) holds when $\liminf_{m \to \infty} \text{tr}\{(\Sigma A)^2\}/\{n_{\min}\Delta(A)\}^2 > 0$. See Section 3.2 for the details of (A-v).

**Proposition 1.** (A-v) implies (A-iii).

**Theorem 1.** If (A-iii) holds, then $\frac{T(A)}{\Delta(A)} = 1 + o_p(1)$ as $m \to \infty$.

**Theorem 2.** If (A-i) and either (A-ii) and (A-iv) or (A-v) hold, then $\{T(A) - \Delta(A)\}/\{K(A)\}^{1/2} \Rightarrow N(0,1)$ as $m \to \infty$.

**Lemma 1.** If (A-ii) and (A-iv) hold, then $K_1(A)/K_1(A) = 1 + o(1)$ as $m \to \infty$.

By combining Theorem 2 with Lemmas 1 and 2, we have the following result.

**Corollary 1.** If (A-i), (A-ii), and (A-iv) hold, then $\{T(A) - \Delta(A)\}/\{K_1(A)\}^{1/2} \Rightarrow N(0,1)$ as $m \to \infty$.

### 2.2. Estimation of $\text{tr}(\Sigma_A^2)$

Throughout this section, we omit the population subscript. Chen, Zhang and Zhong [2010] considered an unbiased estimator of $\text{tr}(\Sigma^2)$, $W_n = \sum_{i \neq j}^{n} (x_i^T x_j)^2/nP_2 - 2\sum_{i \neq j}^{n} x_i^T x_j x_j^T x_i/nP_3 + \sum_{i \neq j \neq s \neq t}^{n} x_i^T x_j x_s x_t/nP_4$, where $nP_r = n!/(n-r)!$. Aoshima and Yata [2011] and Yata and Aoshima [2013a] gave a different unbiased estimator of $\text{tr}(\Sigma^2)$. From these backgrounds, we construct an unbiased
estimator of \( \text{tr}(\Sigma_A^2) \) as
\[
W_n(A) = \sum_{i \neq j}^n \frac{2}{n} x_i^T A x_j x_j^T A x_i - 2 \sum_{i \neq j \neq s}^n \frac{x_j^T A x_j x_j^T A x_s}{n P_3} + \sum_{i \neq j \neq s \neq t}^n \frac{x_j^T A x_j x_j^T A x_t}{n P_4}.
\]
(2.2)

Note that \( \mathbb{E}\{W_n(A)\} = \text{tr}(\Sigma_A^2) \) and \( W_n(I_p) = W_n \). In view of Chen, Zhang and Zhong (2010), one can claim that
\[
\text{Var}\{W_n(A)\} \xrightarrow{n \to \infty} 0 \quad (2.3)
\]
as \( p \to \infty \) and \( n \to \infty \) under (A-i), so that \( W_n(A) = \text{tr}(\Sigma_A^2)\{1 + o_P(1)\} \).

3. Test Procedures for Non-Strongly Spiked Eigenvalue Model

In this section, we consider test procedures given by \( T(A) \) when (A-ii) is met as in the NSSE model. With the help of asymptotic normality, we discuss an optimality of \( T(A) \) for high-dimensional data.

3.1. Test procedure by \( T(A) \)

Let \( z_c \) be a constant such that \( P\{N(0,1) > z_c\} = c \) for \( c \in (0,1) \). For given \( \alpha \in (0,1/2) \), from Corollary 1, we consider testing the hypothesis at (1.1) by
\[
\text{rejecting } H_0 \iff \frac{T(A)}{K_1(A)^{1/2}} > z_\alpha.
\]
(3.1)
The power of the test (3.1) depends on \( \Delta(A) \); we denote it by power(\( \Delta(A) \)).

**Theorem 3.** If (A-i) and (A-ii) hold, then the test (3.1) has, as \( m \to \infty \),
\[
\text{size} = \alpha + o(1) \quad \text{and} \quad \text{power}(\Delta(A)) = \Phi\left(\frac{\Delta(A)}{K_1(A)^{1/2}} - z_\alpha\right) = o(1),
\]
where \( \Phi(\cdot) \) denotes the cumulative distribution function (c.d.f.) of \( N(0,1) \).

**Corollary 2.** If (A-i) holds, then, under \( H_1 \), the test (3.1) has as, \( m \to \infty \),
\[
\text{power}(\Delta(A)) = 1 + o(1) \quad \text{under (A-iii)};
\]
\[
\text{power}(\Delta(A)) - \Phi\left(\frac{\Delta(A)}{K_1(A)^{1/2}} - z_\alpha\right) = o(1) \quad \text{under (A-ii) and (A-iv)};
\]
\[
\text{power}(\Delta(A)) - \Phi\left(\frac{\Delta(A)}{K_2(A)^{1/2}}\right) = o(1) \quad \text{under (A-v)}.
\]

3.2. Choice of \( A \) in (3.1)

Consider the case when (A-v) is met under \( H_1 \). From Corollary 2,
3.3 Simulations

non-Gaussian data. See Section 2 of Aoshima and Yata (2015) for the details. It is not necessary to estimate $A$ matrices. If (A-ii) holds, as in the NSSE model, we rather recommend choosing $\sigma$ of these reasons, we do not recommend using a test procedure based on the Mahalanobis distance, such as (3.1) with $\sigma$. From Corollary 2, in this case, $K_1(A_i) = 2c^2p\{1 + o(1)\}$ as $m \to \infty$. Also, note that $K_2(A_i) = 4c_\star \Delta(A_i)$. Thus, if (A-v) holds,

$$\frac{K_1(A_i)}{K_2(A_i)} = O\left(\frac{pc_\star}{\Delta(A_i)}\right) = O\left(\frac{p}{n_{\min}\Delta(A_i)}\right) \to 0 \quad \text{as} \quad m \to \infty.$$  

This is severe for high-dimensional data. For example, when $\Sigma_1 = \Sigma_2$ and the Mahalanobis distance is bounded as $\lim \sup_{p \to \infty} \Delta_{MD} < \infty$, the sample size should be large enough that $n_{\min}/p \to \infty$ because $\Delta(A_i) = \Delta_{MD}$. Hence, (A-v) is quite strict for high-dimensional data. From Proposition 1 and Corollary 2, for any choice of $A$ in (3.1), $\text{power}(\Delta(A)) = 1 + o(1)$ under (A-v). Hence, the optimal choice of $A$ does not make much improvement in the power if (A-v) is met. If (A-v) is not met (i.e., (A-iv) is met), the test (3.1) has

$$\text{power}(\Delta(A)) \approx \Phi\left(\frac{\Delta(A)}{\{K_2(A)\}^{1/2} - z_\alpha}\right).$$  

from Corollary 2. In this case, $A_\star$ is not the optimal choice any longer. Because of these reasons, we do not recommend using a test procedure based on the Mahalanobis distance, such as (3.1) with $A = A_\star$. In addition, it is difficult to estimate $A_\star$ for high-dimensional data unless the $\Sigma_i$'s are sparse. When they are sparse, see Bickel and Levina (2008).

Srivastava, Katayama and Kano (2013) considered a two-sample test using $A_{i(d)} = c_\star(\Sigma_1/n_1 + \Sigma_2/n_2)^{-1}$ for $A$, where $\Sigma_{i(d)} = \text{diag}(\sigma_{i(1)}, \ldots, \sigma_{i(p)})$ with $\sigma_{i(j)} > 0$ the $j$-th diagonal element of $\Sigma_i$ for $i = 1, 2$; $j = 1, \ldots, p$. We do not recommend choosing $A_{i(d)}$ unless (A-v) is met and the $\Sigma_i$'s are diagonal matrices. If (A-ii) holds, as in the NSSE model, we rather recommend choosing $A = I_p$ in (3.1), yielding the distance-based two-sample test. When $A = I_p$, it is not necessary to estimate $A$ and it is quite flexible for high-dimensional, non-Gaussian data. See Section 2 of Aoshima and Yata (2015) for the details.

3.3 Simulations

We used computer simulations to study the performance of the test procedure
given by \( [3,1] \) when \( A = I_p, \ A = A_\ast, \ A = A_{\ast(d)} \) and \( A = \tilde{A}_{\ast(d)} \). Here, \( \tilde{A}_{\ast(d)} = c_s(S_{i1(d)}/n_1 + S_{i2(d)}/n_2)^{-1}, \) \( S_{i1(d)} = \text{diag}(s_{i1,1}, \ldots, s_{i1,p}), i = 1, 2, \) with \( s_{i,j} \) the \( j \)-th diagonal element of \( S_{i1} \). Srivastava, Katayama and Kano (2013) considered a test procedure given by \( T(A_{\ast(d)}) \). We set \( \alpha = 0.05 \). Independent pseudo-random observations were generated from \( \pi_i : N_p(\mu_i, \Sigma_i), i = 1, 2 \). We set \( p = 2^s, s = 4, \ldots, 10 \) and \( n_1 = n_2 = \lceil p^{1/2} \rceil \), where \( \lceil x \rceil \) denotes the smallest integer \( \geq x \). We set \( \mu_1 = 0 \) and \( \Sigma_1 = \Sigma_2 = C(0.3^{i-j^{1/2}})C, \) where \( C = \text{diag}[(0.5 + 1/(p + 1))^{1/2}, \ldots, (0.5 + p/(p + 1))^{1/2}] \). We considered three cases: (a) \( \mu_2 = 0, \) (b) \( \mu_2 = (1, \ldots, 1, 0, \ldots, 0)^T \) whose first ten elements are 1, and (c) \( \mu_2 = (0, \ldots, 0, 1, \ldots, 1)^T \) whose last ten elements are 1. When \( A = I_p, \ A = A_\ast, \) and \( A = A_{\ast(d)} \), we note that (A-ii) and (A-iv) are met for (a), (b), and (c).

We checked the performance of the test procedures given by \( [3,1] \) with (I) \( A = I_p, \) (II) \( A = A_\ast, \) (III) \( A = A_{\ast(d)} \), and (IV) \( A = \tilde{A}_{\ast(d)} \). The findings were obtained by averaging the outcomes from 2,000 (= \( R \), say) replications in each situation. We defined \( P_r = 1 \) (or 0) when \( H_0 \) was falsely rejected (or not) for \( r = 1, \ldots, 2,000 \) for (a) and defined \( \bar{\alpha} = \sum_{r=1}^{\alpha} P_r/R \) to estimate the size. We also defined \( P_r = 1 \) (or 0) when \( H_1 \) was falsely rejected (or not) for \( r = 1, \ldots, 2,000 \) for (b) and (c) and defined \( 1 - \bar{\beta} = 1 - \sum_{r=1}^{\alpha} P_r/R \) to estimate the power. Note that their standard deviations are less than 0.011. In Fig. 1, we plotted \( \bar{\alpha} \) for (a) and \( 1 - \bar{\beta} \) for (b) and (c). We also plotted the asymptotic power, \( \Phi(\Delta(A)/(K(A))^{1/2} - z_\alpha\{K_1(A)/K(A)\}^{1/2}, \) for (I) to (III) by using Theorem 3. As expected, we observe that the plots get close to the theoretical values. The test with (II) gave a better performance compared to (I) for (b); however, it gave quite a poor performance for (c). The test procedure based on the Mahalanobis distance does not always give a preferable performance for high-dimensional data even when the population distributions are Gaussian with a known and common covariance matrix. See Section 3.2 for the details. We observe that the test with (III) gives a good performance compared to (I) for (b); however, they trade places under (c), because \( \Delta(I_p) < \Delta(A_{\ast(d)}) \) for (b) and \( \Delta(I_p) > \Delta(A_{\ast(d)}) \) for (c) when \( p \) is sufficiently large. The test with (IV) gave quite a poor performance because the size for (IV) was much higher than \( \alpha \) even when \( p \) and the \( n_i \)'s are large. Hence, we do not recommend using the test procedures based on the Mahalanobis distance or the diagonal matrices unless the \( n_i \)'s are large enough to claim (A-v).

We also checked the performance of the test procedures by \( [3,1] \) for the multivariate skew normal (MSN) distribution. See Azzalini and Dalla Valle (1996).
Figure 1. Tests by (3.1) when (I) $A = I_p$, (II) $A = A_\star$, (III) $A = A_{\star(d)}$ and (IV) $A = \hat{A}_{\star(d)}$. The values of $\hat{\alpha}$ are denoted by the dashed lines in the top panel. The values of $1 - \tilde{\beta}$ are denoted by the dashed lines in the left panel for (b) and in the right panel for (c). The asymptotic powers were given by $\Phi(\Delta(A)/\{K(A)\}^{1/2} - z_\alpha\{K_1(A)/K(A)\}^{1/2})$ for (I) to (III) which are denoted by the solid lines both in the panels.

for the details of the MSN distribution. We observed performance similar to that in Fig. 1. The results are in Section S4.1 of the supplementary material.

4. Test Procedures for Strongly Spiked Eigenvalue Model

In this section, we consider test procedures when (A-ii) is not met, as in the SSE model. We emphasize that high-dimensional data often obey the SSE model. See Fig. 1 in Yata and Aoshima (2013b) or Section S3 of the supplementary material as well. In case of (A-iv), $T(A)$ does not satisfy the asymptotic normality in Theorem 2, so that one cannot use the test (3.1). For example, as for $T(I_p)$, we cannot claim either (1.3) or “size $= \alpha + o(1)$” under the SSE model. In such situations, we consider alternative test procedures.
4.1. Distance-based two-sample test

We write $T_I = T(I_p)$, $K_1(I) = K_1(I_p)$, and $\hat{K}_1(I) = \hat{K}_1(I_p)$ when $A = I_p$. For the SSE model, Katayama, Kano and Srivastava (2013) considered a one-sample test. Ma, Lan and Wang (2015) considered a two-sample test for a factor model which is a special case of the SSE model. Katayama, Kano and Srivastava (2013) showed that a test statistic is asymptotically distributed as a $\chi^2$ distribution under the Gaussian assumption. For the two-sample test in (1.1), we have the following result.

Theorem 4. Assume

$$|h_{11}^T h_{21}| = 1 + o(1) \quad \text{and} \quad \frac{\Psi_i(2)}{\lambda_{i1}^2} \to 0, \quad i = 1, 2, \text{ as } p \to \infty,$$

where

$$\Psi_i(s) = \sum_{j=s}^{p} \lambda_{ij}^2 \quad \text{for } i = 1, 2; \quad s = 1, \ldots, p.$$

Then, $(2/K_1(I))^{1/2} T_I + 1 \Rightarrow \chi_1^2$ as $m \to \infty$ under $H_0$, where $\chi_1^2$ denotes a random variable having a $\chi^2$ distribution with $\nu$ degrees of freedom.

We test (1.1) by

rejecting $H_0 \iff \left( \frac{2}{\hat{K}_1(I)} \right)^{1/2} T_I + 1 > \chi_1^2(\alpha),$

where $\chi_1^2(\alpha)$ denotes the $(1 - \alpha)$th quantile of $\chi_1^2$. Note that $\hat{K}_1(I)/K_1(I) = 1 + o_P(1)$ as $m \to \infty$ under (A-i). Then, from Theorem 4, the test (4.2) ensures that size = $\alpha + o(1)$ as $m \to \infty$ under (A-i).

We note that “$|h_{11}^T h_{21}| = 1 + o(1)$ as $p \to \infty$” in (4.1) is not a general condition for high-dimensional data, so that it is necessary to check the condition in data analyses. See Lemma 4.1 in Ishii, Yata and Aoshima (2016) for checking the condition. When (4.1) is not met, the test (4.2) cannot ensure accuracy.

4.2. Test statistics using eigenstructures

We consider the following model.

(A-\text{vi}) For $i = 1, 2$, there exists a positive fixed integer $k_i$ such that $\lambda_{i1}, \ldots, \lambda_{ik_i}$ are distinct in the sense that $\lim inf_{p \to \infty} (\lambda_{ij}/\lambda_{ij'}) > 0$ when $1 \leq j < j' \leq k_i$, and $\lambda_{ik_i}$ and $\lambda_{ik_i+1}$ satisfy

$$\lim inf_{p \to \infty} \frac{\lambda_{ik_i}^2}{\Psi_{i(k_i)}} > 0 \quad \text{and} \quad \frac{\lambda_{ik_i+1}^2}{\Psi_{i(k_i+1)}} \to 0 \text{ as } p \to \infty.$$
Note that (A-vi) implies (1.6); (A-vi) is one of the SSE models. (A-vi) is also a power spiked model given by Yata and Aoshima (2013b). For the spiked model in (1.5), (A-vi) holds under the conditions that \( \alpha_k \geq 1/2 \), \( a_{ij} \neq a_{ij'} \) for \( 1 \leq j < j' \leq k_i \), and \( \alpha_{ik} + 1 < 1/2 \) for \( i = 1, 2 \). We consider the following test statistic with positive-semidefinite matrices, \( A_i \), \( i = 1, 2 \), of dimension \( p \):

\[
T(A_1, A_2) = 2 \sum_{i=1}^{2} \sum_{j=1}^{k_i} \frac{n_{i,j}}{n_i(n_i - 1)} - 2 \bar{x}^T_{1n_1} A_1^{1/2} A_2^{1/2} \bar{x}_{2n_2}.
\]

We do not recommend choosing \( A_i = \Sigma_i^{-1}, i = 1, 2 \); see Section S1.2 in the supplementary material for the details. In addition, it is difficult to estimate \( \Sigma_i^{-1} \)'s for high-dimensional, non-sparse data. Here, we consider \( A_i \)'s as

\[
A_{i(k_i)} = I_p - \sum_{j=1}^{k_i} h_{ij} h_{ij}^T = \sum_{j=k_i+1}^{p} h_{ij} h_{ij}^T \quad \text{for } i = 1, 2.
\]

Note that \( A_{i(k_i)} = A_{i(k_i)}^{1/2} \). We write \( \mu_* = A_{1(k_1)} \mu_1 - A_{2(k_2)} \mu_2 \) and \( \Sigma_* = A_{i(k_i)^T} A_{i(k_i)} \) for \( i = 1, 2 \). Let \( T_* = T(A_{1(k_1)}, A_{2(k_2)}) \), \( \Delta_* = \| \mu_* \|^2 \), and \( K_* = K_{1*} + K_{2*} \), where

\[
K_{1*} = 2 \sum_{i=1}^{2} \frac{\text{tr}(\Sigma_{i*}^2)}{n_i(n_i - 1)} + 4 \frac{\text{tr}(\Sigma_{i1}, \Sigma_{i2})}{n_1 n_2} \quad \text{and} \quad K_{2*} = 4 \sum_{i=1}^{2} \frac{\mu_{i*}^T \Sigma_{i*} \mu_{i*}}{n_i}.
\]

Note that \( E(T_*) = \Delta_* \) and \( \text{Var}(T_*) = K_* \). Also, note that \( \text{tr}(\Sigma_{i*}^2) = \Psi_{i(k_i+1)} \) and \( \lambda_{\max}(\Sigma_{i*}) = \lambda_{k_i+1} \) for \( i = 1, 2 \), so that

\[
\frac{\lambda_{\max}(\Sigma_{i*})}{\text{tr}(\Sigma_{i*}^2)} \to 0 \quad \text{as } p \to \infty \quad \text{for } i = 1, 2, \quad \text{under (A-vi)}.
\]

From Theorem 2, we have the following result.

**Corollary 3.** If (A-i) holds and \( \limsup_{m \to \infty} \Delta_*^2 / K_{1*} < \infty \), then, under (A-vi), \( (T_* - \Delta_*) / K_*^{1/2} \Rightarrow N(0, 1) \) as \( m \to \infty \).

It does not always hold that \( \Delta_* = 0 \) under \( H_0 \) when \( A_{1(k_1)} \neq A_{2(k_2)} \). We assume the following.

(A-vii) \( \frac{\Delta_*^2}{K_{1*}} \to 0 \) as \( m \to \infty \) under \( H_0 \).

This is a mild condition because \( A_{1(k_1)} - A_{2(k_2)} = \sum_{j=1}^{k_2} h_{2j} h_{2j}^T - \sum_{j=1}^{k_1} h_{1j} h_{1j}^T \) is a low-rank matrix with rank \( k_1 + k_2 \) at most, and under \( H_0 \) \( \Delta_* = \| (A_{1(k_1)} - A_{2(k_2)}) \mu_1 \|^2 \) is small. From Corollary 3, under \( H_0 \), it follows that \( P(T_* / K_{1*}^{1/2} > z_\alpha) = \alpha + o(1) \). Similar to (3.1), one can construct a test procedure by using \( T_* \).
Let
\[ x_{ijl} = h_{ij}^T x_{il} = \lambda_{ij}^{1/2} z_{ijl} + \mu_{i(j)} \quad \text{for all } i, j, l, \text{ where } \mu_{i(j)} = h_{ij}^T \mu_i. \]

Then, we write that
\[ T_s = 2 \sum_{i=1}^{2} \sum_{l,l'=1}^{n_i} (x_{il}^T x_{il'} - \sum_{j=1}^{k_i} x_{ijl} x_{ijl'}) \]
\[ - 2 \sum_{l,l'=1}^{n_i} (x_{1l} - \sum_{j=1}^{k_i} x_{1j1} h_{ij})^T (x_{2l'} - \sum_{j=1}^{k_i} x_{2jl} h_{2j}) \]
\[ \frac{n_i(n_i - 1)}{n_1 n_2}. \]

In order to use \( T_s \), it is necessary to estimate the \( x_{ijl} \)'s and \( h_{ij} \)'s.

5. Test Procedure Using Eigenstructures for Strongly Spiked Eigenvalue Model

In this section, we assume (A-vi) and the following for the \( \pi_i \)'s:

\[ (A-viii) \quad E(z_{isj}^2 z_{itj}^2) = E(z_{isj}^2) E(z_{itj}^2), \quad E(z_{isj} z_{itj} z_{iuj} z_{ivj}) = 0 \quad \text{and} \quad E(z_{isj} z_{itj} z_{iuj} z_{ivj}) = 0 \quad \text{for all } s \neq t, u, v, \text{ with } z_{ij} \text{ defined in Section 2}. \]

Note that (A-viii) implies (A-i) because the \( E(z_{ijj}^2) \)'s are bounded and includes the case that \( \Gamma_i = H_i \Lambda_i^{1/2} \) and \( w_{ij} = z_{ij} \). When the \( \pi_i \)'s are Gaussian, (A-viii) naturally holds.

5.1. Estimation of eigenvalues and eigenvectors

Throughout this section, we omit the population subscript for the sake of simplicity. Let \( \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0 \) be the eigenvalues of \( S_n \), and write the eigen-decomposition of \( S_n \) as \( S_n = \sum_{j=1}^p \hat{\lambda}_j \hat{h}_j \hat{h}_j^T \), where \( \hat{h}_j \) denotes a unit eigenvector corresponding to \( \hat{\lambda}_j \). We assume \( \hat{h}_j^T \hat{h}_j \geq 0 \) w.p.1 for all \( j \), without loss of generality. Let \( X = [x_1, \ldots, x_n] \) and \( \bar{X} = [\bar{x}_1, \ldots, \bar{x}_n] \). Then, we define the \( n \times n \) dual sample covariance matrix by \( S_D = (n-1)^{-1} (X - \bar{X})^T (X - \bar{X}) \). Note that \( S_n \) and \( S_D \) share non-zero eigenvalues. We write the eigen-decomposition of \( S_D \) as \( S_D = \sum_{j=1}^{n-1} \hat{\lambda}_j \hat{u}_j \hat{u}_j^T \), where \( \hat{u}_j = (\hat{u}_{j1}, \ldots, \hat{u}_{jn})^T \) denotes a unit eigenvector corresponding to \( \hat{\lambda}_j \). Note that \( \hat{h}_j \) can be calculated as \( \hat{h}_j = \left\{ (n-1)\hat{\lambda}_j \right\}^{-1/2} (X - \bar{X}) \hat{u}_j \). Let \( \delta_j = \lambda_j^{-1} \sum_{s=k+1}^p \lambda_s / (n-1), \) for \( j = 1, \ldots, k \). Let \( m_0 = \min\{p, n\} \).

**Proposition 2.** If (A-vi) and (A-viii) hold, then for \( j = 1, \ldots, k \), \( \hat{\lambda}_j / \lambda_j = 1 + \delta_j + O_P(n^{-1/2}) \) and \( (\hat{h}_j^T h_j)^2 = (1 + \delta_j)^{-1} + O_P(n^{-1/2}) \) as \( m_0 \to \infty \).

If \( \delta_j \to \infty \) as \( m_0 \to \infty \), \( \hat{\lambda}_j \) and \( \hat{h}_j \) are strongly inconsistent in the sense that \( \lambda_j / \hat{\lambda}_j = o_P(1) \) and \( (\hat{h}_j^T h_j)^2 = o_P(1) \). See Jung and Marron (2009) for the
concept of the strong inconsistency. Also, from Proposition 2, under (A-vi) and (A-viii), as \( m_0 \to \infty \),
\[
\| \tilde{h}_j - h_j \|^2 = 2\{1 - (1 + \delta_j)^{-1/2}\} + O_P(n^{-1/2}) \quad \text{for } j = 1, \ldots, k.
\] (5.1)

In order to overcome the curse of dimensionality, Yata and Aoshima (2012) proposed an eigenvalue estimation called the noise-reduction (NR) methodology, which was brought about by a geometric representation of \( S_D \). If one applies the NR methodology, the \( \lambda_j \)'s are estimated by
\[
\hat{\lambda}_j = \frac{\text{tr}(S_D) - \sum_{l=1}^{j} \hat{\lambda}_l}{n - 1 - j} \quad (j = 1, \ldots, n - 2).
\] (5.2)

Here \( \hat{\lambda}_j \geq 0 \) w.p.1 for \( j = 1, \ldots, n - 2 \), and the second term in (5.2) is an estimator of \( \lambda_j \delta_j \). When applying the NR methodology to the PC direction vector, one obtains
\[
\tilde{h}_j = \{ (n - 1)\hat{\lambda}_j \}^{-1/2}(X - \bar{X})\hat{u}_j
\] (5.3)
for \( j = 1, \ldots, n - 2 \).

**Proposition 3.** If (A-vi) and (A-viii) hold, then for \( j = 1, \ldots, k \), \( \hat{\lambda}_j/\lambda_j = 1 + O_P(n^{-1/2}) \) and \( (\tilde{h}_j^T h_j)^2 = 1 + O_P(n^{-1}) \) as \( m_0 \to \infty \).

Here \( \tilde{h}_j \) is not a unit vector because \( \| \tilde{h}_j \|^2 = \hat{\lambda}_j/\hat{\lambda}_j \). From Propositions 2 and 3, under (A-vi) and (A-viii), \( \| \tilde{h}_j - h_j \|^2 = \delta_j\{1 + o_P(1)\} + O_P(n^{-1/2}) \) as \( m_0 \to \infty \) for \( j = 1, \ldots, k \). We note that \( 2\{1 - (1 + \delta_j)^{-1/2}\} < \delta_j \). Thus, in view of (5.1), the norm loss of \( \tilde{h}_j \) is larger than that of \( \hat{h}_j \). However, \( \tilde{h}_j \) is a consistent estimator of \( h_j \) in terms of the inner product even when \( \delta_j \to \infty \) as \( m_0 \to \infty \).

We note that \( \hat{h}_j^T (x_l - \mu) = \lambda_j^{1/2}z_{jl} \) for all \( j, l \). For \( \hat{h}_j \) and \( \tilde{h}_j \), we have the following result.

**Proposition 4.** If (A-vi) and (A-viii) hold, then for \( j = 1, \ldots, k \) (\( l = 1, \ldots, n \)),
\[
\lambda_j^{-1/2}\hat{h}_j^T (x_l - \mu) = (1 + \delta_j)^{-1/2}[z_{jl} + (n - 1)^{1/2}\hat{u}_{jl}\delta_j\{1 + o_P(1)\}] + O_P(n^{-1/2}) \quad \text{and}
\lambda_j^{-1/2}\tilde{h}_j^T (x_l - \mu) = z_{jl} + (n - 1)^{1/2}\hat{u}_{jl}\delta_j\{1 + o_P(1)\} + O_P(n^{-1/2}) \quad \text{as } m_0 \to \infty.
\]

Consider the standard deviation of these quantities. Note that \( \{\sum_{l=1}^{n}\{(n - 1)^{1/2}\hat{u}_{jl}\delta_j\}^2/n\}^{1/2} = O(\delta_j) \) and \( \delta_j = O\{p/(n\lambda_j)\} \) for \( \lambda_{k+1} = O(1) \). Hence, in Proposition 4, the inner products are quite biased when \( p \) is large, as follows. Let \( P_n = I_n - 1_n 1_n^T/n \), where \( 1_n = (1, \ldots, 1)^T \). Then, \( 1_n^T \hat{u}_j = 0 \) and \( P_n \hat{u}_j = \hat{u}_j \) when \( \hat{\lambda}_j > 0 \), since \( 1_n^T S_D 1_n = 0 \). Also, when \( \hat{\lambda}_j > 0 \),
\[
\{(n - 1)\hat{\lambda}_j\}^{1/2}\hat{h}_j = (X - \bar{X})\hat{u}_j = (X - M)P_n \hat{u}_j = (X - M)\hat{u}_j,
\]
where \( M = [\mu, \ldots, \mu] \). Thus it holds that \( \{(n - 1)\hat{\lambda}_j\}^{1/2}\tilde{h}_j^T (x_l - \mu) = \hat{u}_j^T (X - \bar{X})\hat{u}_j = \hat{u}_j^T (X - \bar{X})\hat{u}_j \).
$M^T(x_l - \mu) = \hat{u}_{jl}||x_l - \mu||^2 + \sum_{s=1(s \neq l)}^{n} \hat{u}_{js}(x_s - \mu)^T(x_l - \mu)$, so that $\hat{u}_{jl}||x_l - \mu||^2$ is biased since $E(||x_l - \mu||^2)/\{(n - 1)^{1/2} \lambda_j\} \geq (n - 1)^{1/2} \delta_j$. Hence, one should not apply the $\hat{h}_{jl}$'s or the $\tilde{h}_{jl}$'s to the estimation of the inner product.

Consider a bias-reduced estimation of the inner product. Write

$$\hat{u}_{jl} = \left(\hat{u}_{j1}, \ldots, \hat{u}_{j(l-1)}, \frac{-\hat{u}_{jl}}{n-1}, \hat{u}_{jl+1}, \ldots, \hat{u}_{jn}\right)^T$$

where the $\tilde{h}_{jl}$'s or the $\hat{h}_{jl}$'s are defined by (5.4). From Propositions 3 and 5, we consider the test statistic for (1.1),

$$T = \sum_{l=1}^{n} \{ \hat{u}_{js} + (n - 1)^{-1} \hat{u}_{jl}\}(x_s - \mu)^T(x_l - \mu),$$

so that the large biased term, $||x_l - \mu||^2$, has vanished.

**Proposition 5.** If (A-vi) and (A-viii) hold, then for $j = 1, \ldots, k$ ($l = 1, \ldots, n$),

$$\lambda_j^{-1/2} \hat{h}_{jl}(x_l - \mu) = z_{jl} + \hat{u}_{jl} \times \text{OP}\{(n^{1/2} \lambda_j)^{-1} \lambda_1\} + \text{OP}(n^{-1/2}) \text{ as } m_0 \rightarrow \infty.$$

As $\sum_{l=1}^{n} \{ \hat{u}_{jl} (1/(n^{1/2} \lambda_j)\}^2/n^{1/2} = \lambda_1/(\lambda_j n)$, the bias term is small when $\lambda_1/\lambda_j$ is not large.

### 5.2. Test procedure using eigenstructures

Let $\tilde{x}_{ijl} = \tilde{h}_{ijl} x_{il}$ for all $i, j, l$, where the $\tilde{h}_{ijl}$'s are defined by (5.4). From Propositions 3 and 5, we consider the test statistic for (1.1),

$$\tilde{T}_s = 2 \sum_{l=1}^{2} \sum_{i=1}^{n_{ll'}}(x_l^T x_{il'} - \sum_{j=1}^{k_{ll'}} \tilde{x}_{ijl} \tilde{x}_{ijl'})^{n_{l}(n_{i} - 1)} - 2 \sum_{l=1}^{n_{1}} \sum_{l'=1}^{n_{2}}(x_{1l} - \sum_{j=1}^{k_{1l'}} \tilde{x}_{1jl} \tilde{h}_{1lj})^T(x_{2l'} - \sum_{j=1}^{k_{2l'}} \tilde{x}_{2jl} \tilde{h}_{2lj})/n_{1n_{2}},$$

where the $\tilde{h}_{ijl}$'s are defined by (5.3). We assume the following conditions when (A-vi) is met.
(A-ix) \( \frac{\lambda_{i1}^2}{n_i \Psi_{i(k_i+1)}} \to 0 \) as \( m \to \infty \) for \( i = 1, 2; \)

(A-x) \( \frac{\mu_{1s}^T \Sigma_{is} \mu_{1s} + \mu_{2s}^T \Sigma_{is} \mu_{2s}}{\Psi_{i(k_i+1)}} \to 0 \) as \( p \to \infty \) and \n
\[
\limsup_{m \to \infty} \frac{n_i \{\mu_{i(i)}^2 + (h^T_{ij} \mu_{i*})^2\}}{\lambda_{ij}} < 0 \quad (i' \neq i) \quad \text{for} \quad i = 1, 2; \quad j = 1, \ldots, k_i.
\]

Then, we have the following result.

**Theorem 5.** If (A-vi) and (A-viii) to (A-x) hold, then \( \hat{T}^* - T^* = o_P(K_1^{1/2}) \) as \( m \to \infty \). If also \( \limsup_{m \to \infty} \Delta^2_s / K_1^{1/2} \to 0 \) and \( \limsup_{m \to \infty} \Delta^2_s / K_1^{1/2} \to 0 \), then \( (\hat{T}^* - \Delta^*_s) / K_1^{1/2} \to N(0, 1) \) as \( m \to \infty \).

By using Lemma 1, \( K_1^{1/2} / K_* = 1 + o(1) \) as \( m \to \infty \) under (A-vi) and \( \limsup_{m \to \infty} \Delta^2_s / K_1^{1/2} < \infty \). Thus, we consider estimating \( K_1^{1/2} \). Let \( \hat{A}_{i(k_i)} = I_p - \sum_{j=1}^{k_i} h_{ij}^T h_{ij} \) for \( i = 1, 2 \). We estimate \( K_1^{1/2} \) by \( \hat{K}_1^{1/2} = 2 \sum_{i=1}^{2} \frac{\hat{\Psi}_{i(k_i+1)}}{n_i(n_i - 1)} + 4 \frac{\text{tr}(S_{1n_1} \hat{A}_{1(k_1)} S_{2n_2} \hat{A}_{2(k_2)})}{n_1 n_2} \), where \( \hat{\Psi}_{i(k_i+1)} \) is defined by (S2.1) of the supplementary material.

**Lemma 3.** If (A-vi), (A-viii) and (A-ix) hold, then \( \hat{K}_1^{1/2} / K_1^{1/2} = 1 + o(1) \) as \( m \to \infty \).

Now, we test (1.1) by

\[
\text{rejecting } H_0 \iff \frac{\hat{T}^*}{\hat{K}_1^{1/2}} > z_\alpha. \tag{5.5}
\]

Let power(\( \Delta^*_s \)) denote the power of the test (5.5). Then, from Theorem 5 and Lemma 3, we have the following result.

**Theorem 6.** If (A-vi) and (A-vii) to (A-x) hold, then the test (5.5) has, as \( m \to \infty \),

\[
\text{size} = \alpha + o(1) \quad \text{and} \quad \text{power}(\Delta^*_s) - \Phi\left( \frac{\Delta^*_s}{K_1^{1/2}} - z_\alpha \left( \frac{K_1^{1/2}}{K_*} \right)^{1/2} \right) = o(1).
\]

In general, the \( k_i \)'s are unknown in \( \hat{T}^* \) and \( \hat{K}_1^{1/2} \). See Section S2.2 in the supplementary material for estimation of the \( k_i \)'s. If (4.1) is met, one may use the test (4.2). However, under (4.1), (A-vi) and \( \limsup_{m \to \infty} \Delta^2_s / K_1^{1/2} < \infty \), we note that \( \text{Var}(T^*_s) / \text{Var}(T_I) = O(K_1^{1/2} / K_1) \to 0 \) as \( m \to \infty \), so that the power of
5.2 must be lower than that of (5.5). See Section 6 for numerical comparisons. We recommend the use of the test (5.5) for the SSE model in general.

5.3. How to check SSE models and estimate parameters

We provide a method to distinguish between the NSSE model at (1.4) and the SSE model at (1.6). We also give a method to estimate the parameters required in the test procedure (5.5). We summarize the results in Section S2 of the supplementary material.

5.4. Demonstration

We introduce two high-dimensional data sets that obey the SSE model. We illustrate the proposed test procedure at (5.5) by using microarray data sets. We summarize the results in Section S3 of the supplementary material.

6. Simulations for Strongly Spiked Eigenvalue Model

We used computer simulations to study the performance of the test procedures at (4.2) and (5.5) for the SSE model. In general, the $k_i$'s are unknown for (5.5). Hence, we estimated $k_i$ by $\hat{k}_i$, where $\hat{k}_i$ is given in Section S2.2 of the supplementary material. We set $\kappa(n_i) = (n_i-1)\log n_i^{1/2}$ in (S2.2) of the supplementary material. We checked the performance of the test procedure at (5.5) with $k_i = \hat{k}_i$, $i = 1, 2$. We considered a naive estimator of $T_*$ as $T(\hat{A}_{1(k_1)}, \hat{A}_{2(k_2)})$ and checked the performance of the test procedure given by

$$\text{rejecting } H_0 \iff T(\hat{A}_{1(k_1)}, \hat{A}_{2(k_2)}) \left\rangle \frac{\hat{K}^{1/2}}{\hat{K}^{1/2}} > z_\alpha. \right.$$  \hspace{1cm} (6.1)

We also checked the performance of the test procedure at (3.1) with $A = I_p$. We set $\alpha = 0.05$, $\mu_1 = 0$, and

$$\Sigma_i = \begin{pmatrix} \Sigma_{(1)} & O_{2p-2} \\ O_{p-2,2} & c_i \Sigma_{(2)} \end{pmatrix}$$

with $\Sigma_{(1)} = \text{diag}(p^{2/3}, p^{1/2})$ and $\Sigma_{(2)} = (0.3)^{i-j}1^{1/2}$ for $i = 1, 2$, where $O_{l,l'}$ is the $l \times l'$ zero matrix and $(c_1, c_2) = (1, 1.5)$. Here (4.1) and (A-vi) with $k_1 = k_2 = 2$ are met. When considering the alternative hypothesis, we set $\mu_2 = (0, \ldots, 0, 1, 1, 1, 1)^T$ with last four elements 1. We considered three cases:

(a) $\pi_i : N_p(\mu_i, \Sigma_i)$, $p = 2^s$, $n_1 = 3[p^{1/2}]$ and $n_2 = 4[p^{1/2}]$ for $s = 4, \ldots, 10$;
(b) The $z_{ij}$'s are i.i.d. as the $p$-variate $t$-distribution, $t_p(\nu)$, with mean zero, covariance matrix $I_p$, and degrees of freedom $\nu = 15$, $(n_1, n_2) = (40, 60)$, and...
Figure 2. The performances of five tests: (I) from (3.1) with $A = I_p$, (II) from (4.2), (III) from (5.5), (IV) from (5.5) with $\hat{k}_i = k_i$, $i = 1, 2$, and (V) from (6.1). For (a) to (c), the values of $\alpha$ are denoted by the dashed lines in the left panel and the values of $1 - \beta$ are denoted by the dashed lines in the right panel. The asymptotic power of (III) was given by $\Phi(\Delta_*/K_1^{1/2} - z_\alpha(K_{1s}/K_*)^{1/2})$ which is denoted by the solid line in the right panels. When $n_i$s are small or $p$ is large, $\bar{\alpha}$ for (V) was too high to describe.
\[ p = 50 + 100(s - 1) \text{ for } s = 1, \ldots, 7; \]
(c) \[ z_{itj} = (v_{itj} - 5)/10^{1/2} \text{ ( } t = 1, \ldots, p \text{ ) in which the } v_{itj} \text{'s are i.i.d. as } \chi^2_5, p = 500, n_1 = 10s \text{ and } n_2 = 1.5n_1 \text{ for } s = 2, \ldots, 8. \]

Here (A-viii) is met both for (a) and (c). However, (A-viii) (or (A-i)) is not met for (b). Similar to Section 3.3, we calculated \( \bar{\alpha} \) and \( 1 - \bar{\beta} \) with 2000 replications for five test procedures: (I) from \((3.1)\) with \( A = I_p \), (II) from \((4.2)\), (III) from \((5.5)\), (IV) from \((5.5)\) with \( k_i = \hat{k}_i, i = 1, 2 \), and (V) from \((6.1)\). Their standard deviations are less than 0.011. In Fig. 2, for (a) to (c), we plotted \( \bar{\alpha} \) in the left panel and \( 1 - \bar{\beta} \) in the right panel. From Theorem 6, we plotted the asymptotic power, \( \Phi(\Delta_\ast / K_1^{1/2} - z_\alpha (K_1 / K_\ast)^{1/2}) \), for (III).

We observe that (II) gives better performances compared to (I) regarding size. The size of (I) did not get close to \( \alpha \), probably because \( T_I \) does not satisfy the asymptotic normality given in Theorem 2 when \((1.4)\) is not met. On the other hand, (II) (or (I)) gave quite poor performances compared to (III) and (IV) regarding power, probably because \( \text{Var}(T_I) / \text{Var}(T_\ast) \to \infty \) as \( p \to \infty \) in the current setting. The size of (V) was much higher than \( \alpha \), probably because of the bias of \( T(\hat{A}_1(\hat{k}_1), \hat{A}_2(\hat{k}_2)) \). See Section 5.1 for the details. We observe that (III) and (IV) gave adequate performances even in the non-Gaussian cases. The performances of (III) and (IV) were similar to each other in almost all cases. When \( p \) and the \( n_i \)'s are not small, the plots of (IV) were close to the theoretical values. Hence, we recommend the use of the test procedure at \((5.5)\) with \( k_i = \hat{k}_i, i = 1, 2 \), when \((1.6)\) holds.

We also checked the performance of the test procedures for the MSN distribution and the multivariate skew t (MST) distribution. See Azzalini and Capitanio (2003) and Gupta (2003) for the details of the MST distribution. We give the results in Section S4.2 of the supplementary material.

### 7. Conclusion

By classifying eigenstructures into two classes, the SSE and NSSE models, and then selecting a suitable test procedure depending on the eigenstructure, we can quickly obtain a much more accurate result at lower computational cost. These benefits are vital in groundbreaking research of medical diagnostics, engineering, big data analysis, etc.

### Supplementary Materials

We give data analyses and proofs of the theoretical results, together with ad-
ditional simulations, in the online supplementary material. We also give methods to distinguish between the NSSE model and the SSE model, and estimate the parameters required in the test procedure [5.5].

Acknowledgment

We thank an associate editor and two anonymous referees for their constructive comments. The authors are grateful to the Co-Editor, Hsin-Cheng Huang, for his helpful remarks. Research of the first author was partially supported by Grants-in-Aid for Scientific Research (A) and Challenging Exploratory Research, Japan Society for the Promotion of Science (JSPS), under Contract Numbers 15H01678 and 26540010. Research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), JSPS, under Contract Number 26800078.

References


Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan.

E-mail: aoshima@math.tsukuba.ac.jp

E-mail: yata@math.tsukuba.ac.jp

(Received February 2016; accepted November 2016)