Inference on high-dimensional mean vectors under the strongly spiked eigenvalue model

Aki Ishii1 · Kazuyoshi Yata2 · Makoto Aoshima2

Received: 6 July 2018 / Accepted: 30 November 2018 / Published online: 13 December 2018
© The Author(s) 2021, corrected publication 2021

Abstract
In this paper, we discuss inference problems on high-dimensional mean vectors under the strongly spiked eigenvalue (SSE) model. First, we consider one-sample test. In order to avoid huge noise, we derive a new test statistic by using a data transformation technique. We show that the asymptotic normality can be established for the new test statistic. We give an asymptotic size and power of a new test procedure and investigate the performance theoretically and numerically. We apply the findings to the construction of confidence regions on the mean vector under the SSE model. We further discuss multi-sample problems under the SSE models. Finally, we demonstrate the new test procedure by using actual microarray data sets.

Keywords Asymptotic normality · Data transformation · Eigenstructure estimation · Large p small n · Noise reduction methodology · Spiked model
1 Introduction

In this paper, we consider statistical inference on mean vectors in the high-dimension, low-sample-size (HDLSS) context. Let \( x_1, \ldots, x_n \) be a random sample of size \( n \geq 4 \) from a \( p \)-variate distribution with an unknown mean vector \( \mu \) and unknown covariance matrix \( \Sigma \). In the HDLSS context, the data dimension \( p \) is very high and \( n \) is much smaller than \( p \). We define the eigen-decomposition of \( \Sigma \) by \( \Sigma = H \Lambda H^T \), where \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p) \) is a diagonal matrix of eigenvalues, \( \lambda_1 \geq \cdots \geq \lambda_p \geq 0 \), and \( H = [h_1, \ldots, h_p] \) is an orthogonal matrix of the corresponding eigenvectors. We write the sample mean vector and the sample covariance matrix as \( \bar{x} = \sum_{j=1}^{n} x_j / n \) and \( S = \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T / (n - 1) \).

In this paper, we discuss the one-sample test:

\[
H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0 , \tag{1}
\]

where \( \mu_0 \) is a candidate mean vector. We assume \( \mu_0 = 0 \) without loss of generality. One should note that Hotelling’s \( T^2 \)-statistic defined by

\[
T^2 = n^{-1} \bar{x}^T S^{-1} \bar{x}
\]

is not available because \( S^{-1} \) does not exist in the HDLSS context. Dempster (1958, 1960) and Srivastava (2007) considered the test when \( X \) is Gaussian. When \( X \) is non-Gaussian, Bai and Saranadasa (1996) considered the test. They considered a test statistic defined by

\[
T_{BS} = \| \bar{x} \|^2 - \text{tr} (S) / n , \tag{2}
\]

where \( \| \cdot \| \) denotes the Euclidean norm. Let \( \Delta = \| \mu \|^2 \). Note that \( E(T_{BS}) = \Delta \) and \( \text{Var} (T_{BS}) = K \), where \( K = K_1 + K_2 \) having

\[
K_1 = 2 \frac{\text{tr} (\Sigma^2)}{n(n - 1)} \quad \text{and} \quad K_2 = 4 \frac{\mu^T \Sigma \mu}{n} .
\]

Let us consider the following eigenvalue condition:

\[
\frac{\lambda_1^2}{\text{tr} (\Sigma^2)} \to 0 \quad \text{as} \quad p \to \infty . \tag{3}
\]

Let

\[
m = \min \{ p, n \} .
\]

Under (3), \( H_0 \) and some regularity conditions, Bai and Saranadasa (1996) and Aoshima and Yata (2015) showed the asymptotic normality as follows:

\[
T_{BS}/K_1^{1/2} \Rightarrow N(0, 1) \quad \text{as} \quad m \to \infty . \tag{4}
\]

Here, "\( \Rightarrow \)" denotes the convergence in distribution and \( N(0, 1) \) denotes a random variable distributed as the standard normal distribution.
Aoshima and Yata (2018a) called the eigenvalue condition (3) the “non-strongly spiked eigenvalue (NSSE) model” and drew attention that high-dimensional data do not fit the NSSE model on several occasions. In order to overcome this inconvenience, Aoshima and Yata (2018a) proposed the “strongly spiked eigenvalue (SSE) model” defined by

\[
\liminf_{p \to \infty} \left\{ \frac{\lambda_1^2}{\text{tr}(\Sigma^2)} \right\} > 0
\]

and gave a data transformation technique from the SSE model to the NSSE model.

**Remark 1** When we consider a spiked model such as

\[
\lambda_r = a_r p^{\alpha_r} \quad (r = 1, \ldots, t) \quad \text{and} \quad \lambda_r = c_r \quad (r = t + 1, \ldots, p)
\]

with positive and fixed constants, \(a_r, c_r, \alpha_r\), and a positive and fixed integer \(t\). Note that the NSSE condition (3) is met when \(\alpha_1 < 1/2\). On the other hand, the SSE condition (5) is met when \(\alpha_1 \geq 1/2\). We emphasize that high-dimensional data often have the SSE model. For instance, when we analyze a microarray data, we find several gene networks in which genes in the same network are highly correlated each other. The high correlation is one of the reasons why strongly spiked eigenvalues appear in high-dimensional data analysis.

Let us show a toy example about the asymptotic null distribution of \(T_{BS}\) in (4). We set \(p = 2000\) and \(n = 90\). We assumed that \(X\) is Gaussian as \(N_p(0, \Sigma)\) under \(H_0 : \mu = 0\). We set \(\Sigma = \text{diag}(p^{\alpha_1}, p^{\alpha_2}, 1, \ldots, 1)\) and considered two cases:

(i) \((\alpha_1, \alpha_2) = (1/3, 1/4)\) and (ii) \((\alpha_1, \alpha_2) = (2/3, 1/2)\).

Note that (i) is a NSSE model and (ii) is a SSE model. We generated independent pseudo-random observations for each case and calculated \(T_{BS}/K_1^{1/2}\) 1000 times. In Fig. 1, we gave histograms for (i) and (ii). One can observe that \(T_{BS}/K_1^{1/2}\) does not converge to \(N(0, 1)\) in case (ii).

![Fig. 1](image-url)
It is necessary to develop a new test statistic instead of $T_{BS}$ for the SSE model. Katayama et al. (2013) gave a test statistic and showed that it has an asymptotic null distribution as a $\chi^2$-distribution when $X$ is Gaussian. When $X$ is non-Gaussian, Ishii et al. (2016) gave a different test statistic by using the noise reduction (NR) method due to Yata and Aoshima (2012) and provided an asymptotic non-null distribution to discuss the size and power when $p \to \infty$ while $n$ is fixed. However, the performance of those test statistics is not always preferable because they are heavily influenced by strongly spiked eigenvalues and the variance of their statistics becomes very large because of the huge noise.

In this paper, we avoid the huge noise by using the data transformation technique and construct a new test procedure for the SSE model. The organization of this paper is as follows. In Sect. 2, we consider one-sample test and give a new test statistic by the data transformation. We show that the asymptotic normality can be established for the new test statistic. We give an asymptotic size and power of a new test procedure and investigate the performance theoretically and numerically. In Sect. 3, we apply the findings to the construction of confidence regions on the mean vector under the SSE model. In Sect. 4, we further discuss multi-sample problems under the SSE models. Finally, in Sect. 5, we demonstrate the new test procedure by using actual microarray data sets.

2 High-dimensional one-sample test by data transformation

In this section, we give a new test statistic by the data transformation and show that the asymptotic normality can be established for the new test statistic. We provide an asymptotic size and power of a new test procedure and investigate the performance theoretically and numerically.

2.1 Assumptions

We write the square root of $M$ as $M^{1/2}$ for any positive-semidefinite matrix $M$. Let

$$x_l = H\Lambda^{1/2}z_l + \mu,$$

where $z_l = (z_{1l}, \ldots, z_{pl})^T$ is considered as a sphered data vector having the zero mean vector and identity covariance matrix. Similar to Bai and Saranadasa (1996) and Chen and Qin (2010), we assume the following assumption as necessary:

(A-i) $\limsup_{p \to \infty} E(z_{rl}^4) < \infty$ for all $r$, $E(z_{rl}^2z_{sl}^2) = E(z_{rl}^2)E(z_{sl}^2) = 1$, $E(z_{rl}z_{sl}z_{rl}) = 0$ and $E(z_{rl}z_{sl}z_{st}z_{ul}) = 0$ for all $r \neq s, t, u$.

When $x_l$s are Gaussian, (A-i) naturally holds. Let

$$\Psi_r = \text{tr} (\Sigma^2) - \sum_{s=1}^{r-1} \lambda_s^2 = \sum_{s=r}^{p} \lambda_s^2 \quad \text{for} \quad r = 1, \ldots, p.$$

Similar to Aoshima and Yata (2018a), we assume the following model:
(A-ii) There exists a fixed integer \( k \geq 1 \) such that

\[
\begin{align*}
&\text{(i)} \quad \text{When } k \geq 2, \lambda_1, \ldots, \lambda_k \text{ are distinct in the sense that} \\
&\quad \lim_{p \to \infty} \inf (\lambda_r / \lambda_s - 1) > 0 \text{ for } 1 \leq r < s \leq k; \\
&\text{(ii)} \quad \lambda_k \text{ and } \lambda_{k+1} \text{ satisfy} \\
&\quad \lim_{p \to \infty} \inf \frac{\lambda_k^2}{\Psi_k} > 0 \text{ and } \frac{\lambda_{k+1}^2}{\Psi_{k+1}} \to 0 \text{ as } p \to \infty.
\end{align*}
\]

Note that (A-ii) is one of the SSE models. For example, (A-ii) holds when the spiked model in (6) has the constants such that

Remark 2 For instance, one may consider a \((k + 1)\)-mixture model as a SSE model having \( k \) strongly spiked eigenvalues \((\alpha_k \geq 1/2)\). See Sect. 5.1 in Aoshima and Yata (2018b) and Sect. 3 in Yata and Aoshima (2015) for the details.

2.2 Data transformation

According to Aoshima and Yata (2018a, b), we consider transforming the data from the SSE model to the NSSE model using the projection matrix

\[
A = I_p - \sum_{j=1}^{k} h_j h_j^T = \sum_{j=k+1}^{p} h_j h_j^T.
\]

We have that \( E(Ax_j) = A \mu = \mu_s, \text{ say} \) and

\[
\operatorname{Var}(Ax_j) = A \Sigma A = \sum_{j=k+1}^{p} \lambda_j h_j h_j^T = \Sigma_s, \text{ say}.
\]

Note that \( \text{tr} (\Sigma_s^2) = \Psi_{k+1} \) and \( \lambda_{\max}(\Sigma_s) = \lambda_{k+1} \), where \( \lambda_{\max}(\Sigma_s) \) denotes the largest eigenvalue of \( \Sigma_s \). Then, it holds that

\[
\frac{\lambda_{\max}^2(\Sigma_s)}{\text{tr}(\Sigma_s^2)} \to 0 \text{ as } p \to \infty \text{ under (A-ii)}.
\]

Thus, the transformed data has the NSSE model.

By using the transformed data, we consider the following quantity:

\[
T_{DT} = \|A\tilde{x}\|^2 - \frac{\text{tr} (AS)}{n} = 2 \sum_{i<j} x_i^T A x_j \frac{n}{n(n-1)}
\]

\[
= 2 \sum_{i<j} x_i^T (x_i - \sum_{j=1}^{k} x_j x_j)\frac{n}{n(n-1)},
\]
where

\[ x_{jl} = h_j^T x_l \] for all \( j, l \).

Let \( \Delta_s = \| \mu_s \|^2 \). Note that \( E(T_{DT}) = \Delta_s \) and \( \text{Var}(T_{DT}) = K_s \), where \( K_s = K_{1s} + K_{2s} \) having

\[ K_{1s} = 2 \frac{\text{tr}(\Sigma_s^2)}{n(n-1)} \quad \text{and} \quad K_{2s} = \frac{4 \mu_s^T \Sigma_s \mu_s}{n}. \]

Under (A-ii), we consider the following conditions as necessary:

(A-iii) \( \limsup_{m \to \infty} \frac{\Delta_s^2}{K_{1s}} < \infty \); \quad (A-iv) \( \frac{K_{1s}}{\Delta_s^2} \to 0 \) as \( m \to \infty \).

We note that (A-iii) is met under \( H_0 \) in (1). Also, note that \( K_{2s}/K_{1s} = o(1) \) under (A-ii) and (A-iii), and \( K_{2s}/\Delta_s^2 = o(1) \) under (A-iv) from the facts that \( K_{2s} \leq 4 \Delta_s \lambda_{k+1}/n = O(\Delta_s K_{1s}^{1/2}) \) and \( \Delta_s \lambda_{k+1}/n = o(K_{1s}) \) under (A-ii) and (A-iii).

Then, we have the following results.

**Proposition 1** Assume (A-i) to (A-iii). Then, it holds that as \( m \to \infty \)

\[ \frac{T_{DT}}{K_{1s}^{1/2}} - \Delta_s = \frac{T_{DT}}{K_{1s}^{1/2}} - \Delta_s + o_p(1) \Rightarrow N(0, 1). \]

**Proposition 2** Assume (A-ii) and (A-iv). Then, it holds that as \( m \to \infty \)

\[ \frac{T_{DT}}{\Delta_s} = 1 + o_p(1). \]

### 2.3 New test statistic and its asymptotic properties

Based on \( T_{DT} \), we define the test statistic as follows:

\[ \hat{T}_{DT} = 2 \sum_{l < p} \frac{x_l^T x_p - \sum_{j=1}^k \hat{x}_{jl} \hat{x}_{jp}}{n(n-1)} \]

\[ = \| \hat{x} \|^2 - \text{tr}(S)/n - 2 \frac{\sum_{l < p} \sum_{j=1}^k \hat{x}_{jl} \hat{x}_{jp}}{n(n-1)}, \] (7)

where \( \hat{x}_{jl} \) is a certain estimator of \( x_{jl} \). We give the definition of \( \hat{x}_{jl} \) in Appendix 1 since its derivation is somewhat technical. The number \( k \) in \( \hat{T}_{DT} \) should be determined before the data transformation. See Appendix 1 about a choice of \( k \).

We assume the following conditions when (A-ii) is met.

(A-v) \( \frac{\lambda_{1i}^2}{n \text{ tr}(\Sigma_s^2)} \to 0 \) as \( m \to \infty \); \quad (A-vi) \( \liminf_{P \to \infty} \frac{\Delta_s}{\Delta} > 0 \) when \( \Delta \neq 0 \).
Note that (A-vi) is a mild condition when \( p \) is much larger than \( k \) because 
\[
\Delta_* = \Delta - \sum_{j=1}^{k}(h_j^{T}\mu)^2 = \sum_{j=k+1}^{p}(h_j^{T}\mu)^2.
\]
Also, note that \( \Delta_* \neq 0 \) when \( \Delta \neq 0 \) under (A-vi). Then, we have the following results.

**Theorem 1** Assume (A-i), (A-ii), (A-v) and (A-vi). It holds that
\[
\hat{T}_{DT} - T_{DT} = o_p(\max\{K_{1*}^{1/2}, \Delta_*\}) \quad \text{as} \quad m \to \infty.
\]

Furthermore, we also assume (A-iii). Then, it holds that
\[
\frac{\hat{T}_{DT} - \Delta_*}{\sqrt{K_{1*}^{1/2}}} = \frac{\hat{T}_{DT} - \Delta_*}{\sqrt{K_{1*}^{1/2}}} + o_p(1) \Rightarrow N(0, 1) \quad \text{as} \quad m \to \infty.
\]

**Corollary 1** Assume (A-i), (A-ii) and (A-iv) to (A-vi). Then, it holds that
\[
\frac{\hat{T}_{DT}}{\Delta_*} = 1 + o_p(1) \quad \text{as} \quad m \to \infty.
\]

From Theorem 1 and Lemma 1 in Appendix 1, under (A-i) to (A-iii), (A-v) and (A-vi), it holds that
\[
\frac{\hat{T}_{DT} - \Delta_*}{\sqrt{K_{1*}^{1/2}}} \Rightarrow N(0, 1) \quad \text{as} \quad m \to \infty,
\]

where \( \hat{K}_{1*} \) is given in Lemma 1 in Appendix 1. Thus, we can construct a test procedure of (1) by (8).

We checked the behavior of the asymptotic null distribution by using the toy example in (ii) of Fig. 1. We assumed \( N_{p}(\mathbf{0}, \Sigma) \) under \( H_0: \mu = \mathbf{0} \). We set \( p = 2000 \), \( n = 90 \) and \( \Sigma = \text{diag}(p^{\alpha_1}, p^{\alpha_2}, 1, \ldots, 1) \) with \( (\alpha_1, \alpha_2) = (2/3, 1/2) \). We calculated \( \hat{T}_{DT}/\sqrt{K_{1*}^{1/2}} \) 1000 times and constructed a histogram in Fig. 2. One can observe that \( \hat{T}_{DT}/\sqrt{K_{1*}^{1/2}} \) converges to \( N(0, 1) \) as expected theoretically.

### 2.4 Test procedure for (1) by \( \hat{T}_{DT} \) and its performance

We give a new test procedure for (1) by using \( \hat{T}_{DT} \). For a given \( \alpha \in (0, 1/2) \), we test the hypothesis (1) by

\[
\text{rejecting } H_0 \iff \hat{T}_{DT}/\sqrt{\hat{K}_{1*}^{1/2}} > z_{\alpha},
\]

where \( z_{\alpha} \) is a constant such that \( P(N(0, 1) > z_{\alpha}) = \alpha \). From Theorem 1, Corollary 1 and Lemma 1 in Appendix 1, we have the following results.
Theorem 2 Assume (A-i), (A-ii), (A-v) and (A-vi). The test procedure (9) holds that as \( m \to \infty \)

\[
\text{Size} = a + o(1) \quad \text{and} \quad \text{Power} = \Phi \left( \frac{\Delta_0 - z_0 \left( \frac{K_0}{K_1} \right)^{1/2}}{\frac{1}{2} \frac{1}{2}} \right) + o(1),
\]

where \( \Phi(\cdot) \) denotes the c.d.f. of \( \mathcal{N}(0, 1) \).

Corollary 2 Assume (A-i), (A-ii) and (A-iv) to (A-vi). Under \( H_1 \), the test procedure (9) holds that as \( m \to \infty \)

\[
\text{Power} = 1 + o(1).
\]

Remark 3 The number \( k \) in \( \hat{T}_{DT} / \hat{K}_{1*}^{1/2} \) should be determined before the data transformation. See Appendix 1 about a choice of \( k \).
Katayama et al. (2013) for $T_{BS}$ and $T_{SD}$ and denoted them by $\widetilde{T}_{BS}$ and $\widetilde{T}_{SD}$. We set $\mu = 0$ under $H_0$ in (1). We set $\alpha = 0.05$. Let $\Omega_s$ be a $s \times s$ matrix as $\Omega_s = (0.3^{[i-j]^{1/2}})$. We considered three cases. For (i) and (ii), we set
\[
\Sigma = \begin{pmatrix}
\Gamma_{p_1} & 0 & 0 \\
0 & \Gamma_{p_2} & 0 \\
0 & 0 & \Omega_{p-p_1-p_2}
\end{pmatrix},
\]
where $\Gamma_s = (I_s + 1_s 1_s^T)/2$, $1_s = (1, \ldots, 1)^T$ and $O$ is a zero matrix. As for $\Gamma_s$, note that the largest eigenvalue is $(s + 1)/2$ and the other ones are $1/2$.

(i) $x_i$ is Gaussian, $N_p(0, \Sigma)$, $p = 2^s$ ($s = 6, \ldots, 12$). We set $n = 2[p^{1/2}]$ and $(p_1, p_2) = ([p^{2/3}], [p^{1/2}])$, where $[x]$ denotes the smallest integer $\geq x$.

(ii) $x_i = H_A^{1/2}(z_{11}, \ldots, z_{1p})^T + \mu$, where $H$ and $A$ are given by $\Sigma = HAH^T$ having $(p_1, p_2) = ([p^{3/4}], [p^{1/2}])$, and $z_{1r} = (v_{1r} - 5)/10^{1/2} (r = 1, \ldots, p)$ with $v_{1r}$s i.i.d. as the chi-squared distribution with 5 degrees of freedom. We set $n = 50$ and $p = 100s$ for $s = 1, \ldots, 10$.

(iii) $x_i$ is of a Gaussian mixture of $N_p(\mu_1, \Omega_p)$ and $N_p(\mu_2, \Omega_p)$ with a mixture proportion $1/2$, where $\mu_1 = (1, \ldots, 1, 0, \ldots, 0)^T$ and $\mu_2 = (-1, \ldots, -1, 0, \ldots, 0)^T$ whose first $p_1$ elements are 1 and $-1$. We set $n = 10s$ ($s = 4, \ldots, 12$), $p = 1000$ and $p_1 = [p^{2/3}]$. Note that $\Sigma = \Omega_p + (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T/4$ and $\lambda_1 \approx p^{2/3}$.

For (i) and (ii), (A-i) and (A-ii) with $k = 2$ are met. For (iii), (A-ii) with $k = 1$ is met while (A-i) is not met. See Sect. 6 in Yata and Aoshima (2010) for the details. Note that (A-v) is met for (i). As for the alternative hypothesis, we considered two cases:

(a) $\mu = (1, 1, 1, 0, \ldots, 0)^T$ whose first 3 elements are 1 and the remaining elements are 0; and

(b) $\mu = (0, \ldots, 0, 1, 1, 1)^T$ whose last 3 elements are 1 and the remaining elements are 0.

In each case of (i) to (iii), we checked the performance by 2000 replications. We defined $P_r = 1$ (or 0) when $H_0$ was falsely rejected (or not) for $r = 1, \ldots, 2000$, and defined $\bar{\alpha} = \sum_{r=1}^{2000} P_r/2000$ to estimate the size. We also defined $P_r = 1$ (or 0) when $H_1$ was falsely rejected (or not) for $r = 1, \ldots, 2000$, and defined $1 - \bar{\beta} = 1 - \sum_{r=1}^{2000} P_r/2000$ to estimate the power for (a) and (b). Note that their standard deviations are less than 0.011. In Fig. 3, we plotted $\bar{\alpha}$ (the left panel) and $1 - \bar{\beta}$ (the right upper panel for (a) and right lower panel for (b)) in each case of (i) to (iii). One can observe that $\hat{T}_{DT}$ gave preferable performances both for the size and power in (i) and (iii). $\hat{T}_{DT}$ gave a moderate performance for (ii); however, it seems that the sample size was not large enough to guarantee accuracy. $T_{BS}$ and $T_{SD}$ gave bad performances for the size and power. They are quite sensitive about the NSSE condition (3). On the other hand, $\hat{T}_{BS}$ and $\hat{T}_{SD}$ gave good performances for (i) to (iii) regarding the size, while they gave much worse performances regarding the power compared to $\hat{T}_{DT}$. This is probably because the variance of those statistics becomes very large since they are heavily influenced by strongly spiked eigenvalues. In fact, it holds that $Var(T_{DT})/Var(T_{BS}) = K_s/K = o(1)$ for (i) to
Fig. 3 The performance of the test procedures given by Bai and Saranadasa (1996) (denoted by $T_{BS}$), Srivastava and Du (2008) (denoted by $T_{SD}$), Katayama et al. (2013) (denoted by $T_{BS}$ and $T_{SD}$) and (9) (denoted by $T_{DT}$). The value of $\alpha$ is denoted by the dashed line in the left panel, and the value of $1-\alpha$ is denoted by the dashed line in the right upper panel for (a) and in the right lower panel for (b) in each case of (i) to (iii). The solid line denotes the asymptotic power of $T_{DT}$ given in Theorem 2.
(iii). We emphasize that it is quite important to select a suitable test procedure depending on the eigenstructure. In conclusion, we recommend using $\hat{T}_{DT}$ when the data fits the SSE model.

### 3 Applications

In this section, we give several applications of the findings in Sect. 2.

#### 3.1 Confidence region for $\mu$

We first consider the following quantity:

$$ T_\mu = \|A(\bar{x} - \mu)\|^2 - \text{tr}(AS)/n $$

$$ = \|\bar{x} - \mu\|^2 - \text{tr}(S)/n - 2 \sum_{l<l'} \sum_{j=1}^k \tilde{h}_j^T(x_l - \mu)\tilde{h}_j^T(x_{l'} - \mu) $$

Note that $E(T_\mu) = 0$ and $\text{Var}(T_\mu) = K_{1*}$. Then, from Proposition 1, we have the following result.

**Proposition 3** Asssume (A-i) and (A-ii). It holds that as $m \to \infty$

$$ T_\mu/K_{1*}^{1/2} \Rightarrow N(0, 1). $$

Let

$$ R = \left\{ \mu \in \mathbb{R}^p : \text{tr}(AS)/n - z_{a/2}K_{1*}^{1/2} \leq \|A(\bar{x} - \mu)\|^2 \leq \text{tr}(AS)/n + z_{a/2}K_{1*}^{1/2} \right\}. $$

From Proposition 3 it holds that under (A-i) and (A-ii),

$$ P(\mu \in R) = 1 - \alpha + o(1) \quad \text{as} \quad m \to \infty. $$

We note that $\text{tr}(AS)/n = \text{tr}(\Sigma_\nu)/n + o_p(K_{1*}^{1/2})$ under (A-i) and (A-ii). Also, note that $z_{a/2}K_{1*}^{1/2} = o(\text{tr}(\Sigma_\nu)/n)$ under (A-ii) from the fact that $\text{tr}(\Sigma_\nu^2)/[\text{tr}(\Sigma_\nu)]^2 = o(1)$ as $p \to \infty$ under (A-ii). One can claim that $\text{tr}(AS)/n - z_{a/2}K_{1*}^{1/2} > 0$ with probability tending to 1. Thus, $R$ indicates that $A\mu (= \mu_*)$ is included in the region sandwiched by the two $p$-dimensional spheres with radii of $(\text{tr}(\Sigma_\nu)/n - z_{a/2}K_{1*}^{1/2})^{1/2}$ and $(\text{tr}(\Sigma_\nu)/n + z_{a/2}K_{1*}^{1/2})^{1/2}$ from the center $A\bar{x}$.

Now, we construct a confidence region for $\mu$ by estimating the eigenstructures. Let

$$ \eta(\mu) = 2\sum_{l<l'} \sum_{j=1}^k \tilde{h}_j^T(x_l - \mu)\tilde{h}_j^T(x_{l'} - \mu) $$

where $\tilde{h}_j$s are given in Appendix 1. From (7) we consider the following quantity:
By using Theorem 1, we have the following result.

**Proposition 4** Assume (A-i), (A-ii) and (A-v). It holds that as $m \to \infty$

$$
\hat{T}_\mu - T_\mu = o_p(K_1^{1/2}).
$$

Let

$$
\hat{R} = \left\{ \mu \in \mathbb{R}^p : \text{tr} (S)/n - z_{a/2} \hat{K}_{1*}^{1/2} \leq \|\hat{x} - \mu\|^2 - \eta(\mu) \leq \text{tr} (S)/n + z_{a/2} \hat{K}_{1*}^{1/2} \right\}. \quad (11)
$$

From Lemma 1 in Appendix 1, Propositions 3 and 4, it holds that under (A-i), (A-ii) and (A-v),

$$
P(\mu \in \hat{R}) = 1 - \alpha + o(1) \text{ as } m \to \infty.
$$

For example, if $0 \in \hat{R}$, we can consider $0$ as a likely candidate of $\mu$.

**Remark 4** From Lemma 1 in Appendix 1, the length of the interval for $\|\hat{x} - \mu\|^2 - \eta(\mu)$ in $\hat{R}$ is $2z_{a/2} \hat{K}_{1*}^{1/2} \{1 + o_p(1)\}$. When $\text{tr} (\Sigma^{2*}_e) = O(p)$ and $p^{1/2}/n = o(1)$, it holds that $K_{1*} = o(1)$ so that the length of the interval tends to 0.

### 3.2 Confidence lower bound for $\Delta$

We construct a confidence lower bound for $\Delta$ by

$$
I_L = [\max \{ \hat{T}_D - z_{a/2} \hat{K}_{1*}^{1/2}, 0 \}, \infty),
$$

where $\alpha \in (0, 1/2)$. By using Theorem 1 and Lemma 1, it holds that

$$
P(\Delta \in I_L) \geq 1 - \alpha + o(1) \text{ as } m \to \infty
$$

under (A-i) to (A-iii), (A-v) and (A-vi) from the fact that $\Delta \geq \Delta_*$. For example, one can apply $I_L$ to the high-dimensional discriminant analysis in which a lower bound for $\Delta$ is the key to discuss misclassification rates of any classifier. See Sect. 4 in Aoshima and Yata (2011) and Sects. 3 and 4 in Aoshima and Yata (2014).

### 3.3 Estimation of $\Delta$

We consider the following conditions when (A-ii) is met.

(A-vii) $\frac{\hat{\lambda}_1^2}{n^3 \Delta^2} \to 0$ and $\frac{\hat{\lambda}_1}{\hat{\lambda}_k} = O(n^{1/2})$ as $m \to \infty$;

(A-viii) $\Delta_*/\Delta \to 1$ as $p \to \infty$. 

\( \odot \) Springer
We note that the second condition \( \lambda_1 / \lambda_k = O(n^{1/2}) \) in (A-vii) naturally holds when \( k = 1 \). Also, note that (A-viii) is a mild condition when \( p \) is much larger than \( k \) because \( \Delta_s = \Delta - \sum_{j=1}^{k} (h_j^T \mu)^2 = \sum_{j=k+1}^{p} (h_j^T \mu)^2 \).

We have the following result.

**Theorem 3** Assume (A-i), (A-ii), (A-iv), (A-vii) and (A-viii). It holds that

\[
\hat{T}_{DT} \rightarrow 1 + o_p(1) \text{ as } m \rightarrow \infty.
\]

One can estimate \( \Delta \) by \( \hat{T}_{DT} \). On the other hand, by noting that \( K_2 \leq 4 \lambda_1 / n \leq 4 \Delta K_1^{1/2} \), it holds that \( \text{Var}(T_{BS})/\Delta^2 \rightarrow 0 \) as \( m \rightarrow \infty \) under

\[
K_1 / \Delta \rightarrow 0 \text{ as } m \rightarrow \infty. \tag{12}
\]

Hence, we have that

\[
T_{BS}/\Delta = 1 + o_p(1).
\]

We note that (12) is equivalent to

\[
\frac{\lambda_1^2}{n^2 \Delta^2} \rightarrow 0 \text{ as } m \rightarrow \infty
\]

under (A-ii). If \( \lambda_1 \) is quite large enough to claim \( \Delta / \lambda_1 \rightarrow 0 \) as \( p \rightarrow \infty \), the first condition in (A-vii) is much milder than (12). Thus, we recommend using \( \hat{T}_{DT} \) as an estimation of \( \Delta \) under the SSE model. We demonstrate the estimation of \( \Delta \) by using a microarray data set in Sect. 5.2.

### 4 Multi-sample problems

Suppose we have independent and \( p \)-variate populations, \( \pi_i, i = 1, \ldots, g \), having unknown mean vector \( \mu_i \) and unknown covariance matrix \( \Sigma_i \) for each \( \pi_i \). We have i.i.d. observations, \( x_{ij}, \ldots, x_{in_i} \), from each \( \pi_i \). We assume \( n_1 = \min\{n_1, \ldots, n_g\} \) without loss of generality.

#### 4.1 Two-sample test with balanced samples

In this section, we set \( g = 2 \) and consider the two-sample test:

\[
H_0 : \mu_1 = \mu_2 \text{ vs. } H_1 : \mu_1 \neq \mu_2. \tag{13}
\]

Aoshima and Yata (2018a) gave a two-sample test procedure under the SSE model by the data transformation technique. We consider testing (13) by using the one-sample test procedure (9) when the sample sizes are asymptotically balanced such that \( n_1/n_2 \rightarrow 1 \) as \( n_1, n_2 \rightarrow \infty \).

Let

\[
x_j = x_{1j} - x_{2j} \text{ for } j = 1, \ldots, n_1 = n, \text{ say }, \tag{14}
\]
and let $\mu = \mu_1 - \mu_2$ and $\Sigma = \Sigma_1 + \Sigma_2$. Note that $x_j$ has the SSE model in general if one of the $\pi_i$ has the SSE model. For example, if $\text{tr} (\Sigma^2_1) \geq \text{tr} (\Sigma^2_2)$ and $\liminf_{p \to \infty} \lambda^2_{\max} (\Sigma_1) / \text{tr} (\Sigma^2_1) > 0$, it holds that

$$\liminf_{p \to \infty} \frac{\lambda^2_{\max} (\Sigma)}{\text{tr} (\Sigma^2)} \geq \liminf_{p \to \infty} \frac{\lambda^2_{\max} (\Sigma_1)}{4 \text{tr} (\Sigma^2_1)}. $$

Now, we consider using the test procedure (9) with $x_j$.

When we apply the test procedure (9) to (13), we do not use the remaining samples, $x_{2n_1+1}, \ldots, x_{2n_2}$. Let us check the influence of the remaining samples. Let $\Sigma_* = \Sigma_{1*} + \Sigma_{2*}$. From Theorem 2, the asymptotic power is given by

$$\Phi(\Delta_*/K_*^{1/2} - z_{\alpha} K^1_{1*}^{1/2} / K_*^{1/2}),$$

where

$$K_{1*} = 2 \text{tr} (\Sigma^2_1)/n^2 = 2 \left( \text{tr} (\Sigma^2_{1*}) + \text{tr} (\Sigma^2_{2*}) \right)/n^2 + 4 \text{tr} (\Sigma_{1*} \Sigma_{2*})/n^2$$

and $K_*= K_{1*} + 4 \sum_{i=1}^2 \mu^*_i \Sigma_{is} \mu_*/n$.

In the balanced setting such as $n_1/n_2 \to 1$ when $n_1 \to \infty$, the above asymptotic power is equivalent to that of the two-sample test procedure by Aoshima and Yata (2018a). See Theorem 6 in Aoshima and Yata (2018a). Thus, one may use the test procedure (9) when the sample sizes are balanced. We give a demonstration by using microarray data sets in Sections 5.1 and 5.2.

### 4.2 Multi-sample test for a linear function of mean vectors

In this section, we consider the following test for a linear function of mean vectors:

$$H_0 : \sum_{i=1}^k b_i \mu_i = 0 \quad \text{vs.} \quad H_1 : \sum_{i=1}^k b_i \mu_i \neq 0, \quad (15)$$

where $b_i$s are known and nonzero scalars (not depending on $p$). We consider a transformation given by Bennett (1951) and Anderson (2003). Let

$$x_j = b_1 x_{1j} + \sum_{i=2}^{g} b_i \sqrt{\frac{n_1}{n_i}} \left( x_{ij} - \frac{1}{n_1} \sum_{j=1}^{n_1} x_{ij} + \frac{1}{\sqrt{n_1n_i}} \sum_{j'=1}^{n_1} x_{ij'} \right) \quad \text{for} \quad j = 1, \ldots, n_1. \quad (16)$$

Then, it holds that

$$E(x_j) = \sum_{i=1}^{g} b_i \mu_i \quad \text{and} \quad \text{Var} (x_j) = \sum_{i=1}^{g} b_i^2 (n_1/n_i) \Sigma_i.$$
Now, we assume \( \pi _{i} : N_{p}(\mu _{i} , \Sigma _{i}) \) for \( i = 1 , \ldots , g \). Then, \( x_1 , \ldots , x_n \) are i.i.d. as \( N_{p}(\sum _{i=1}^{g} b_{i} \mu _{i} , \sum _{i=1}^{g} b_{i}^{2}(n_{i} / n) \Sigma _{i}) \). Nishiyama et al. (2013) proposed a test procedure for (15) by the transformation (16) under a NSSE model. In this section, we consider testing (15) under the SSE model. Let \( n = n_{1} , \mu = \sum _{i=1}^{g} b_{i} \mu _{i} \) and \( \Sigma = \sum _{i=1}^{g} b_{i}^{2}(n_{i} / n) \Sigma _{i} \). Note that \( x_{j} \) has the SSE model in general if one of the \( \pi _{i} \)s has the SSE model. For example, if \( \text{tr} (\Sigma _{i}^{2}) \geq \text{tr} (\Sigma _{j}^{2}) \) for \( i \geq 2 \) and \( \liminf _{p \to \infty } \lambda _{\max }^{2}(\Sigma _{i}) / \text{tr} (\Sigma _{j}^{2}) > 0 \), it holds that

\[
\liminf _{p \to \infty } \frac{\lambda _{\max }^{2}(\Sigma )}{\text{tr} (\Sigma _{j}^{2})} \geq \liminf _{p \to \infty } \frac{b_{1}^{2} \lambda _{\max }^{2}(\Sigma _{1})}{(\sum _{i=1}^{g} b_{i}^{2})^{2} \text{tr} (\Sigma _{j}^{2})} > 0.
\]

One may apply the one-sample test procedure (9) with \( x_{j} \) to (15). We give a demonstration using a microarray data set in Sect. 5.3.

It should be noted that the Gaussian assumption is required for the multisample test. We investigate robustness of our test procedure against the Gaussian assumption by some simulation studies. We set \( g = 4 , b_{1} = 1 , \mu _{1} = 0 \) and \( \Sigma _{i} = c_{i} \text{diag}(p^{2/3} , p^{1/3} , 1 , \ldots , 1) \) for \( i = 1 , \ldots , 4 \), where \( c_{1} = 1 \) and \( c_{2} = c_{3} = c_{4} = 2 \). We considered three cases:

(a) \( x_{ij} , j = 1 , \ldots , n_{i} \) are i.i.d. as \( N_{p}(\mu _{i} , \Sigma _{i}) \), where \( n_{1} = [p^{1/2}] , p = 2^{i}(s = 6 , \ldots , 12) \) and \( n_{i} = 2n_{1} (i = 2 , 3 , 4) \);

(b) \( x_{ij} = \Sigma _{i}^{1/2}(z_{i,j(0)} , \ldots , z_{i,j(s)})^{T} + \mu _{i} , j = 1 , \ldots , n_{i} \), having \( z_{i,j(s)} = (v_{i,j(s)} - v) / \sqrt{2}v \), where \( v_{i,j(s)} \)s are i.i.d. as the chi-squared distribution with degrees of freedom \( v = 5s \) (\( s = 1 , \ldots , 7 \)), and \( n_{1} = 50 \) and \( n_{i} = 100 \) (\( i = 2 , 3 , 4 \));

(c) \( x_{ij} = \Sigma _{i}^{1/2}(z_{i,j(0)} , \ldots , z_{i,j(3)})^{T} + \mu _{i} , j = 1 , \ldots , n_{i} \), where \( (z_{i,j(0)} , \ldots , z_{i,j(3)})^{T} \)s are i.i.d. as \( p \)-variate \( t \)-distribution, \( t_{p}(I_{p} , v) \), with mean zero, covariance matrix \( I_{p} \) and degrees of freedom \( v = 5s \) (\( s = 1 , \ldots , 7 \)), and \( n_{1} = 50 \) and \( n_{i} = 100 \) (\( i = 2 , 3 , 4 \)).

Case (a) corresponds to the cases (b) and (c) when \( v \to \infty \).

In each case of (a) to (c), we checked the performance by 2000 replications. We defined \( P_{r} = 1 \) (or 0) when \( H_{0} \) in (15) was falsely rejected (or not) for \( r = 1 , \ldots , 2000 \), and defined \( \overline{a} = \sum _{r=1}^{2000} P_{r} / 2000 \) to estimate the size. In Fig. 4, we plotted \( \overline{a} \) for (a) in the left panel, for (b) in the right upper panel and for (c) in the right lower panel. As for (a), our test procedure gave a good performance when \( p \) is large as expected in theory. One can observe that our test procedure is robust in (b). On the other hand, our test procedure seems to be a little affected by a heavy tail distribution in (c).

5 Data analysis

5.1 Two-sample test with balanced samples

We analyzed gene expression data by using the proposed test procedure (9). We used microarray data sets of prostate cancer with 12625(=p) genes consisting of
two classes: \( \pi_1 \) : non-tumor (50(= \( n_1 \)) samples) and \( \pi_2 \) : prostate cancer (52(= \( n_2 \)) samples). See Singh et al. (2002) for the details. The data set is available at Jeffrey’s web page (URL: http://www.bioinf.ucd.ie/people/ian/). First, we checked (A-ii) by applying the method given in Sect. 5 of Aoshima and Yata (2018a) and confirmed that each class fits the SSE model. According to the test procedure by (9) with \( \hat{k} \) given in Appendix 1, we tested (13) at significant level 0.05. We used the first 50(= \( n_2 \)) samples of prostate cancer and constructed new data as 
\[
x_{ij} = x_{1j} - x_{2j}, \quad j = 1, \ldots, n_1,
\]
where \( x_{ij} \) denotes the \( j \)th sample in \( \pi_i \).

In addition, we considered testing (13) when \( H_0 \) was true. For each \( \pi_i \), we divided the samples into two parts, \( \mathbf{x}_{1i}, \ldots, \mathbf{x}_{in_i/2} \) and \( \mathbf{x}_{in_i/2+1}, \ldots, \mathbf{x}_{in_i} \) for \( i = 1, 2 \). We constructed new data as 
\[
x'_{ij} = x_{ij} - \mathbf{x}_{in_i/2+j}, \quad j = 1, \ldots, n_i/2.
\]
Then, we applied the test procedure (9) in the same way.

The results are summarized in Table 1. One can observe that the test procedure (9) worked well when \( H_0 \) in (13) was true. On the other hand, the preceding test procedures \( T_{BS} \) in (2) and \( T_{SD} \) in (10) did not control the size.

### 5.2 Two-sample test for matched samples

We considered applying the test procedure (9) to a microarray data set of matched samples. We used the colon adenocarcinomas data with 7457(= \( p \)) genes given by

---

**Fig. 4** The value of \( \hat{\pi} \) for (a) to (c). We considered three cases: (a) \( x_{ij}, j = 1, \ldots, n_1 \) are i.i.d. as \( \mathcal{N}_p(\mathbf{0}, \Sigma_i) \) in the left panel; (b) \( x_{ij} = \Sigma_i^{1/2}(z_{1j(i)}, \ldots, z_{pj(i)})^T + \mu_j, j = 1, \ldots, n_1 \), having \( z_{pj(i)} = (v_{pj(i)} - v)/\sqrt{2}v \), where \( v_{pj(i)} \)s are i.i.d. as the chi-squared distribution with degrees of freedom \( v \) in the right upper panel; and (c) \( x_{ij} = \Sigma_i^{1/2}(z_{1j(i)}, \ldots, z_{pj(i)})^T + \mu_j, j = 1, \ldots, n_1 \), where \( (z_{1j(i)}, \ldots, z_{pj(i)})^T \)s are i.i.d. as \( f_p(T_p, \nu) \) in the right lower panel.
1

Table 1 The values of three test statistics, $\hat{T}_{DT}/\hat{K}_{1}^{1/2}$, $T_{BS}/\hat{K}_{1}^{1/2}$ and $T_{SD}$, where $K_{1}$ is a consistent estimator of $K_{1}$ given in Bai and Saranadasa (1996). We also gave the value of $\hat{k}$ used in $\hat{T}_{DT}$, where $\hat{k}$ is given in Appendix 1. We used the data sets from Singh et al. (2002) having $p = 12625$

|                  | $\hat{k}$ | $\hat{T}_{DT}/\hat{K}_{1}^{1/2}$ | $T_{BS}/\hat{K}_{1}^{1/2}$ | $T_{SD}$ |
|------------------|-----------|----------------------------------|-----------------------------|----------|
| Prostate cancer vs. Non-tumor | Value 2   | 44.1                             | 14.7                        | 11.0     |
|                  | $H_{0}$ in (13) | Rejected                       | Rejected                    | Rejected |
|                  | When $H_{0}$ is true. (Prostate cancer) |                                    |                             |          |
|                  | Value 3   | $-2.1$                           | 14.1                        | 19.0     |
|                  | $H_{0}$ in (13) | Accepted                       | Rejected                    | Rejected |
|                  | When $H_{0}$ is true. (Non-tumor) |                                    |                             |          |
|                  | Value 5   | $-72.2$                          | 14.3                        | 16.0     |
|                  | $H_{0}$ in (13) | Accepted                       | Rejected                    | Rejected |

Fig. 5 The first 10 eigenvalues of the data set given by Notterman et al. (2001). We calculated the eigenvalues for $x_{ij} - x_{2j}$, $j = 1, \ldots, 18$

Notterman et al. (2001). The data set includes the microarray data of $\pi_{1}$: adenocarcinomas cell and $\pi_{2}$: normal cell from 18(= n) persons. Note that one cannot use a two-sample test procedure because the data sets do not satisfy the independency. The data set is available at Princeton university’s web page (URL: http://genomics-pubs.princeton.edu/oncology/). First, we confirmed that $x_{ij} - x_{2j}$, $j = 1, \ldots, 18$, fit the SSE model. We plotted the first 10 eigenvalues in Fig. 5. We used the test procedure (9) similarly to Sect. 5.1. Then, $H_{0}$ in (13) was rejected.

Next, we estimated $\Delta$ using the two estimators given in Sect. 3.3. We had $T_{BS} = 3132.66$ and $\hat{T}_{DT} = 2125.49$, so that the ratio was $T_{BS}/\hat{T}_{DT} = 1.47$. Note that our estimation $\hat{T}_{DT}$ has the consistency property. See Sect. 3.3 for the details. It seems that $T_{BS}$ overestimated $\Delta$ because of huge noise.

5.3 Multi-sample problem using confidence region

We considered applying the test procedure given in Sect. 4.2 to a microarray data set. We used malignant gliomas and oligodendrogliomas data sets having 12625(= $p$) genes. See Nutt et al. (2003) for the details. We analyzed microarray data consisting
of four classes: \( \pi_1 \): classic oligodendrogliomas (CO) (7(=\( n_1 \)) samples), \( \pi_2 \): nonclassic oligodendrogliomas (NO) (15(=\( n_2 \)) samples), \( \pi_3 \): classic glioblastoma (CG) (14(=\( n_3 \)) samples) and \( \pi_4 \): nonclassic glioblastoma (NG) (14(=\( n_4 \)) samples). We transformed the data by (16) and we had \( \hat{k} = 1 \) for the transformed data. We constructed the confidence region by (11) for \( \mu = (\mu_1 - \mu_3) - (\mu_2 - \mu_4) \). When \( \mu = 0 \), we had

\[
\hat{R} = \{ \mu \in \mathcal{R}^p | 1.83 \times 10^7 \leq \| \bar{x} \|^2 - \eta(0) \leq 3.79 \times 10^7 \}.
\]

We calculated \( \| \bar{x} \|^2 - \eta(0) = 3.29 \times 10^7 \), so that \( \mu = 0 \in \hat{R} \). Hence, we concluded that \( \mu = 0 \) is a candidate of \( \mu = (\mu_1 - \mu_3) - (\mu_2 - \mu_4) \).

**Acknowledgements** We would like to thank an associate editor and two anonymous referees for their constructive comments.

### Appendix 1

In this section, we give estimators for the parameters in our new test statistic, \( \hat{T}_{DT} \), and discuss their asymptotic properties.

#### Estimation of \( x_{jl} \)

Let \( X = [x_1, \ldots, x_n] \), \( \bar{X} = [\bar{x}, \ldots, \bar{x}] \) and \( P_n = I_n - 1_n 1_n^T/n \), where \( 1_n = (1, \ldots, 1)^T \). Recall \( S \) is the sample covariance matrix. One can write that \( S = (X - \bar{X})(X - \bar{X})^T/(n - 1) = XP_nX^T/(n - 1) \). Let us write the eigen-decomposition of \( S \) as \( S = \sum_{j=1}^p \hat{\lambda}_j \hat{h}_j \hat{h}_j^T \) having eigenvalues \( \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0 \) and the corresponding \( p \)-dimensional unit eigenvectors \( \hat{h}_1, \ldots, \hat{h}_p \). We assume \( P(h_j^T \hat{h}_j \geq 0) = 1 \) for all \( j \) without loss of generality. We also define the following \( n \times n \) dual sample covariance matrix:

\[
S_D = (n - 1)^{-1}P_nXX^TP_n = (n - 1)^{-1}(X - \bar{X})^T(X - \bar{X}).
\]

Note that \( S \) and \( S_D \) share non-zero eigenvalues. Let us 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 \( n \)-dimensional unit eigenvector corresponding to \( \hat{\lambda}_j \). In high-dimensional settings, we calculate \( \hat{h}_j \) by using \( \hat{u}_j \) as follows:

\[
\hat{h}_j = ((n - 1)\hat{\lambda}_j)^{-1/2}(X - \bar{X})\hat{u}_j.
\]

Note that \( 1_n^T S_D 1_n = 0 \), so that \( 1_n^T \hat{u}_j = \sum_{l=1}^n \hat{u}_{jl} = 0 \) when \( \hat{\lambda}_j > 0 \).

For high-dimensional data, the sample eigenvalues and eigenvectors get huge noise. See Jung and Marron (2009) and Shen et al. (2016) for the details. In order to remove the huge noise, Yata and Aoshima (2012) focused on a geometric representation of \( S_D \) and proposed the NR method. If one applies the NR method, the \( \hat{\lambda}_j \)s and \( \hat{h}_j \)s are estimated by

\[
\hat{\lambda}_j = \hat{\lambda}_j - \frac{\text{tr}(S_D) - \sum_{l=1}^{j-1} \hat{\lambda}_l}{n - j - 1} \quad (j = 1, \ldots, n - 2)
\]  

(17)
\[
\hat{h}_j = \{(n-1)\hat{x}_j\}^{-1/2}(x - \bar{x})\hat{u}_j \quad (j = 1, \ldots, n-2).
\] (18)

Note that \(P(\hat{x}_j \geq 0) = 1\) for \(j = 1, \ldots, n-2\). We emphasize that \(\hat{x}_j\)'s and \(\hat{h}_j\)'s have consistency properties under much milder conditions than \(\hat{x}_j\)'s and \(\hat{h}_j\)'s. However, for the estimation of \(x_{jl} = x_l^T h_j\), Aoshima and Yata (2018a) showed that \(x_l^T \hat{h}_j\) involves a huge bias and gave a modification for all \(j, l\) by

\[
\hat{h}_{jl} = \left(\frac{n-1}{n-2}\right) \frac{(X - \bar{X})\hat{u}_{jl}}{(n-1)\hat{x}_j} = \left(\frac{n-1}{n-2}\right)^{1/2}(X - \bar{X})\hat{u}_{jl} \quad \text{for all } j, l.
\]

From Lemma B.1 in Aoshima and Yata (2018a), we have the following result.

**Proposition 5** Assume (A-i) and (A-ii). It holds for \(j = 1, \ldots, k\) that as \(m \to \infty\)

\[
\sum_{l=1}^{n} \frac{(\hat{x}_{jl} - x_{jl})^2}{n\lambda_j} = O_P(\frac{\lambda_j^2}{n} + \frac{\mu^T \Sigma \mu}{\lambda_j^2}).
\]

Note that \(\text{Var}(x_{jl}) = \lambda_j\). If \(\lambda_j^2/(n\lambda_j^2) = o(1)\), the normalized mean squared error in Proposition 5 tends to 0 under \(H_0\) in (1). See Sect. 5.1 in Aoshima and Yata (2018a) for the details.

**Estimation of \(K_{1n}\)**

We use the CDM method given by Yata and Aoshima (2010) to estimate \(K_{1n}\). Let \(n_1 = \lfloor n/2 \rfloor\) and \(n_2 = n - n_1\). Let \(X_1 = [x_1, \ldots, x_{n_1}]\) and \(X_2 = [x_{n_1+1}, \ldots, x_n]\). We define

\[
S_D^{(1)} = \{(n_1 - 1)(n_2 - 1)\}^{-1/2}(X_1 - \bar{X}_1)^T(X_2 - \bar{X}_2),
\]

where \(\bar{X}_i = [\bar{x}_{n(i)}, \ldots, \bar{x}_{n(i)}]\) with \(\bar{x}_{n_1} = \sum_{i=1}^{n_1} x_i/n_1\) and \(\bar{x}_{n_2} = \sum_{i=n_1+1}^{n} x_i/n_2\). We estimate \(\lambda_j\) by the \(j\)-th singular value, \(\hat{\lambda}_j\), of \(S_D^{(1)}\), where

\[
\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{n_2-1} \geq 0.
\]
Yata and Aoshima (2010) showed that \( \hat{\lambda}_j \) has several consistency properties for high-dimensional non-Gaussian data. Note that \( E\{ \text{tr} (S_{D(1)}^T S_{D(1)}^c) \} = \text{tr} (\Sigma^2) \). We estimate \( \Psi_r \) by \( \hat{\Psi}_1 = \text{tr} (S_{D(1)}^T S_{D(1)}^c) \) and

\[
\hat{\Psi}_r = \text{tr} (S_{D(1)}^T S_{D(1)}^c) - \sum_{s=1}^{r-1} \hat{\lambda}_s^2 \quad \text{for } r = 2, \ldots, n(2) - 1. \tag{20}
\]

Note that \( P(\hat{\Psi}_r \geq 0) = 1 \) for \( r = 1, \ldots, n(2) - 1 \). Then, Aoshima and Yata (2018a) gave the following result.

**Lemma 1** (Aoshima and Yata 2018a) Assume (A-i) and (A-ii). Then, it holds that \( \hat{\Psi}_r / \Psi_r = 1 + o_P(1) \) as \( m \to \infty \) for \( r = 1, \ldots, k + 1 \).

Thus, we estimate \( \text{tr} (\Sigma^2) \) by \( \hat{\Psi}_{k+1} \). Let

\[
\hat{K}_{1*} = 2 \hat{\Psi}_{k+1} / \{ n(n - 1) \}.
\]

Then, from Lemma 1, under (A-i) and (A-ii), it holds that

\[
\hat{K}_{1*} / K_{1*} = 1 + o_P(1) \quad \text{as } m \to \infty.
\]

**Estimation of \( k \)**

Recently, Jung et al. (2018) proposed a test of the number of spiked components for high-dimensional data. On the other hand, Aoshima and Yata (2018a) gave estimation of \( k \) in (A-ii) by using the CDM method. Let \( \hat{\tau}_r = \hat{\Psi}_{r+1} / \hat{\Psi}_r = (1 - \hat{\lambda}_r^2 / \hat{\psi}_r) \) for all \( r \), where \( \hat{\Psi}_r \) given by (20). Note that \( \hat{\tau}_r \in [0, 1) \) for \( \hat{\lambda}_r > 0 \). Then, Aoshima and Yata (2018a) gave the following results.

**Proposition 6** (Aoshima and Yata 2018a) Assume (A-i) and (A-ii). It holds that as \( m \to \infty \)

\[
P(\hat{\tau}_r < 1 - c_r) \to 1 \quad \text{with some fixed constant } c_r \in (0, 1) \quad \text{for } r = 1, \ldots, k;
\]

\[
\hat{\tau}_{k+1} = 1 + o_P(1).
\]

**Proposition 7** (Aoshima and Yata 2018a) Assume (A-i), (A-ii) and (A-v). Assume also \( \hat{\lambda}_{k+1}^2 / \Psi_{k+1} = O(n^{-c}) \) as \( m \to \infty \) with some fixed constant \( c > 1/2 \). It holds that as \( m \to \infty \)

\[
P\left( \hat{\tau}_{k+1} > \{ 1 + (k + 1)\gamma(n) \}^{-1} \right) \to 1,
\]

where \( \gamma(n) \) is a function such that \( \gamma(n) \to 0 \) and \( n^{1/2} \gamma(n) \to \infty \) as \( n \to \infty \).

From Propositions 6 and 7, if one can assume the conditions in Proposition 7, one may consider \( k \) as the first integer \( r \) (\( \approx \hat{k}_o \), say) such that

\[
\hat{\tau}_{r+1} \{ 1 + (r + 1)\gamma(n) \} > 1 \quad (r \geq 0). \tag{21}
\]
Then, it holds that $P(\hat{k}_o = k) \to 1$ as $m \to \infty$. Note that $\hat{P}_{n_o} = 0$ from the fact that rank$(S_{D(1)}) \leq n(2) - 1$. Finally, one may choose $k$ as

$$\hat{k} = \min\{\hat{k}_o, n(2) - 2\}$$

in actual data analysis. According to Aoshima and Yata (2018a), we use $\gamma(n) = (n^{-1} \log n)^{1/2}$ in (21). If $\hat{k} = 0$, that is, the data is the NSSE model.

### Appendix 2

Let $\psi_r = \lambda_r^2/(n^2 \lambda_r) + \mu^T \Sigma \mu/(n \lambda_r)$ for $r = 1, \ldots, k$.

### Proofs of Propositions 1 and 2

From the fact that $\lambda_{k+1}^2 \leq \text{tr} (\Sigma^2)$, we note that

$$K_{2*} \leq 4\Delta_{*}^2 \lambda_{k+1}/n = O(\Delta_{*} K_{1*}^{1/2}).$$

(22)

Then, under (A-iv), it holds that as $m \to \infty$

$$\text{Var} \left( T_{D*} \right)/\Delta_{*}^2 = K_{*}/\Delta_{*}^2 \to 0.$$  

(23)

Thus, we can conclude the result of Proposition 2. On the other hand, from (22), under (A-ii) and (A-iii), it holds that $K_{2*} = o(K_{1*})$. Then, from Theorem 5 in Aoshima and Yata (2015), we can conclude the result of Proposition 1.

### Proof of Theorem 1

From (S6.28) in Appendix B of Aoshima and Yata (2018a), we claim that for $r = 1, \ldots, k$,

$$\sum_{j \leq l} \frac{n \hat{x}_{il} \hat{x}_{il'} - n \hat{x}_{il} \hat{x}_{il'}}{n(n-1)} = O_P \left\{ \psi_r^{1/2} (\psi_r^{1/2} + \lambda_r^{1/2}/n^{1/2} + h_r^T \mu) \right\}$$

(24)

as $m \to \infty$ under (A-i) and (A-ii). Here, by noting that tr$(\Sigma^2)/\lambda_k^2 = O(1)$ and $\mu^T \Sigma \mu \leq \Delta \lambda_1$ we have that for $r = 1, \ldots, k$

$$\lambda_1^2/(n^2 \lambda_r) = o(K_{1*}^{1/2}); \quad \mu^T \Sigma \mu/(n \lambda_r) = O(\Delta_{*} \lambda_1/(n \lambda_r)) = o(\Delta_{*});$$

$$\lambda_1^2/n^2 = o(K_{1*}); \quad \text{and} \quad \mu^T \Sigma \mu/n^2 = O(\Delta_{*} \lambda_1/n^2) = o(\Delta_{*} K_{1*}^{1/2})$$

under (A-ii), (A-v) and (A-vi), so that for $r = 1, \ldots, k$

$$\psi_r = o(\max\{K_{1*}^{1/2}, \Delta_{*}\}) \quad \text{and} \quad \psi_r^{1/2} \lambda_r^{1/2}/n^{1/2} = o(\max\{K_{1*}^{1/2}, \Delta_{*}\})$$

(25)
Also, note that $|h_r^T \mu| \leq \Delta^{1/2} = O(\Delta_*^{1/2})$ under (A-vi), so that for $r = 1, \ldots, k$

$$\psi_r^{1/2} h_r^T \mu = o(\max\{K_*^{1/2}, \Delta_*\}).$$

Thus, from (24) we can conclude the first result of Theorem 1. In addition, from the first result, it holds that

$$\widehat{T}_{DT} = T_{DT} + o_P(K_*^{1/2})$$

under (A-i) to (A-iii), (A-v) and (A-vi). Thus from Proposition 1, it concludes the second result of Theorem 1.

**Proofs of Corollaries 1 and 2**

From the first result of Theorem 1, it holds that

$$\frac{\widehat{T}_{DT}}{\Delta_*} = \frac{T_{DT}}{\Delta_*} + o_P(1)$$

under (A-i), (A-ii) and (A-iv) to (A-vi). Thus from Proposition 2, it concludes the result of Corollary 1. In addition, from Corollary 1 and Lemma 1, it holds

$$P\left(\frac{\widehat{T}_{DT}}{K_*^{1/2}} > z_{\alpha}\right) = P\left(\frac{\widehat{T}_{DT}/\Delta_*}{z_*^{1/2} \kappa_{1*}^{1/2}/\Delta_*} > P\{1 + o_P(1) > o_P(1)\} \to 1$$

under (A-i), (A-ii) and (A-iv) to (A-vi). It concludes the result of Corollary 2.

**Proof of Theorem 2**

First, we consider the case when (A-iii) is met. We note that $K_*^{1/2} / K_*^{1/2} \to 1$ as $m \to \infty$ under (A-ii) and (A-iii). Then, from Theorem 1 and Lemma 1, under (A-i) to (A-iii), (A-v) and (A-vi), we have that

$$P\left(\frac{\widehat{T}_{DT}}{K_*^{1/2}} > z_{\alpha}\right) = P\left(\frac{\widehat{T}_{DT} - \Delta_*}{K_*^{1/2}} > \left(z_{\alpha}K_*^{1/2} - \Delta_*\right)/K_*^{1/2} + o_P(1)\right)$$

$$= \Phi\left(\left(\Delta_* - z_{\alpha}K_*^{1/2}\right)/K_*^{1/2}\right) + o(1) = \Phi\left(\Delta_*/K_*^{1/2} - z_{\alpha}\right) + o(1).$$

(25)

It concludes the result about Size in Theorem 2. On the other hand, under (A-iv), from (23), it holds that

$$\Phi\left(\left(\Delta_* - z_{\alpha}K_*^{1/2}\right)/K_*^{1/2}\right) = 1 + o(1).$$

Hence, from (25) and Corollary 2, by considering a convergent subsequence of $\Delta_*^2/K_*^{1/2}$, we can conclude the result about Power in Theorem 2.
Proofs of Propositions 3 and 4

By noting that \( E(x_l - \mu) = 0 \) \((l = 1, \ldots, n)\) and \( \text{Var}(T_\mu) = K_{1s} \), the results are obtained straightforwardly from the results of Proposition 1 and Theorem 1.

Proof of Theorem 3

By noting that \( \mu^T \Sigma \mu \leq \Delta \lambda_1 \) we have that for \( r = 1, \ldots, k \)

\[
\lambda_1^2 / (n^2 \lambda_r) = O(\lambda_1 / n^{3/2}) = o(\Delta); \quad \mu^T \Sigma \mu / (n \lambda_r) = O(\Delta \lambda_1 / (n \lambda_r)) = o(\Delta);
\]
\[
\lambda_2^2 / n^3 = o(\Delta^2); \quad \text{and} \quad \mu^T \Sigma \mu / n^2 = O(\Delta_0, \lambda_1 / n^2) = o(\Delta^2)
\]
as \( m \to \infty \) under (A-vii). Thus, from (24) it holds that

\[
\hat{T}_{\Delta T} / \Delta = T_{\Delta T} / \Delta + o_P(1)
\]
under (A-i), (A-ii) and (A-vii). Then, from Proposition 2, we can conclude the result of Theorem 3.

Proof of Proposition 5

From Lemma B.1 in Appendix B of Aoshima and Yata (2018a), under (A-i) and (A-ii), it holds for \( j = 1, \ldots, k \) that as \( m \to \infty \)

\[
\sum_{i=1}^{n} \frac{(\bar{x}_{ij} - x_{ij})^2}{n} = O_P(n \psi_j).
\]

Thus, we can conclude the result of Proposition 5.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, 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 changes were made. 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/4.0/.

References

Anderson, T. W. (2003). An Introduction to Multivariate Statistical Analysis (3rd ed.). New York: Wiley.
Aoshima, M., & Yata, K. (2011). Two-stage procedures for high-dimensional data. Sequential Analysis, 30, 356–399. (Editor’s special invited paper).
Aoshima, M., & Yata, K. (2014). A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. Annals of the Institute of Statistical Mathematics, 66, 983–1010.
Aoshima, M., & Yata, K. (2015). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. Methodology and Computing in Applied Probability, 17, 419–439.
Aoshima, M., & Yata, K. (2018a). Two-sample tests for high-dimension, strongly spiked eigenvalue models. Statistica Sinica, 28, 43–62.
Aoshima, M., & Yata, K. (2018b). Distance-based classifier by data transformation for high-dimension, strongly spiked eigenvalue models. *Annals of the Institute of Statistical Mathematics*, in press (https://doi.org/10.1007/s10463-018-0655-z).

Bai, Z., & Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica*, 6, 311–329.

Bennett, B. M. (1951). Note on a solution of the generalized Behrens–Fisher problem. *Annals of the Institute of Statistical Mathematics*, 2, 87–90.

Chen, S. X., & Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38, 808–835.

Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, 29, 995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16, 41–50.

Ishii, A., Yata, K., & Aoshima, M. (2016). Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context. *Journal of Statistical Planning and Inference*, 170, 186–199.

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

Jung, S., Lee, M. H., & Ahn, J. (2018). On the number of principal components in high dimensions. *Biometrika*, 105, 389–402.

Katayama, S., Kano, Y., & Srivastava, M. S. (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 116, 410–421.

Nishiyama, T., Hyodo, M., Seo, T., & Pavlenko, T. (2013). Testing linear hypotheses of mean vectors for high-dimension data with unequal covariance matrices. *Journal of Statistical Planning and Inference*, 143, 1898–1911.

Notterman, D. A., Alon, U., Sierk, A. J., & Levine, A. J. (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Research*, 61, 3124–3130.

Nutt, C. L., Mani, D. R., Betensky, R. A., Tamayo, P., Cairncross, J. G., Ladd, C., et al. (2003). Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Research*, 63, 1602–1607.

Shen, D., Shen, H., Zhu, H., & Marron, J. S. (2016). The statistics and mathematics of high dimension low sample size asymptotics. *Statistica Sinica*, 26, 1747–1770.

Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, 37, 53–86.

Srivastava, M. S., & Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99, 386–402.

Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1, 203–209.

Yata, K., & Aoshima, M. (2010). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *Journal of Multivariate Analysis*, 101, 2060–2077.

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

Yata, K., & Aoshima, M. (2015). Principal component analysis based clustering for high-dimension, low-sample-size data. *arXiv preprint*, arXiv:1503.04525.