Intrinsic Dimensionality Estimation of High Dimension, Low Sample Size Data with D-Asymptotics

Kazuyoshi Yata

Graduate School of Pure and Applied Sciences
University of Tsukuba
Ibaraki 305-8571, Japan

Makoto Aoshima

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

Abstract High dimension, low sample size (HDLSS) data are becoming common in various fields such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. Such data have surprising and often counter-intuitive geometric structures because of the high-dimensional noise that dominates and corrupts the local neighborhoods. In this paper, we estimate the intrinsic dimension (ID) which allows one to distinguish between deterministic chaos and random noise of HDLSS data. A new ID estimating methodology is given and its properties are studied by using a d-asymptotic approach.

Key Words: Dual covariance matrix; Effective dimension; HDLSS; Large p small n; Maximum eigenvalue.

1. INTRODUCTION

Recently, a variety of methods have been developed to deal with nonlinear dimensionality reduction such as Isometic Feature Mapping (ISOMAP) (Tenenbaum et al. 2000), Local Linear Embedding (LLE) (Roweis and Saul 2000) and Hessian-based Locally Linear Embedding (HLLE) (Donoho and Grimes 2003), and others. Those methods focus on finding a low-dimensional curved manifold embedding of high-dimensional data. The dimensionality of the embedding is a key parameter in those algorisms. However, there is no consensus on how such dimensionality is determined and the dimensionality has often been chosen heuristically from the curve of residual variance as a function of dimension. Constructing a reliable

estimator of intrinsic dimension (ID) and understanding its statistical properties will clearly improve the performance of manifold learning methods.

The existing approaches to estimating ID can roughly be divided into two groups: the eigenvalue methods and the geometric methods. Eigenvalue methods are based on principal component analysis (PCA). Details can be found in Fukunaga and Olsen (1971), Verveer and Duin (1995), Bruske and Sommer (1998), and others. The geometric methods are mostly based on fractal dimensions or nearest neighbor distances. Details can be found in Grassberger and Procaccia (1983), Camastra and Vinciarelli (2002), Costa and Hero (2004), Wang and Marron (2008), and others. The statistical properties of a maximum likelihood estimator of ID were studied by Levina and Bickel (2005).

A currently very active area of data analysis is microarrays for measuring gene expression. A single measurement yields simultaneous expression levels for thousands to tens of thousands of genes. Because the measurements tend to be very expensive, the sizes of most datasets are in the tens, or maybe low hundreds, and so the dimension d of the data vectors is much larger than the sample size n. The current ID estimating methods may be very difficult to apply to such high dimension, low sample size (HDLSS) data since those methods naturally require very large samples in a high-dimensional space.

Related asymptotic studies assume that the dimension d increases, whereas the sample size n can be fixed or increases along with d. Bai and Silverstein (1998), Johnstone (2001), Baik et al. (2005), and Baik and Silverstein (2006) studied asymptotics where the ratio d/n goes to a constant. On the other hand, Hall et al. (2005) and Ahn et al. (2007) studied asymptotics specialized in the HDLSS case of $d \to \infty$ with a fixed n, which is called the *d-asymptotics*. They took a d-asymptotic approach and showed that, under some regularity conditions, the geometrical structure of HDLSS data becomes deterministic as d increases while n is fixed.

In this paper, we narrow down a target to the HDLSS case with Euclidean dimension and present a new ID estimating methodology with a d-asymptotic approach. Suppose we have a $d \times n$ data matrix $\boldsymbol{X}_{(d)} = [\boldsymbol{x}_{1(d)}, ..., \boldsymbol{x}_{n(d)}]$ with $d > n$, where $\boldsymbol{x}_{j(d)} =$

2

$(x_{1j(d)}, ..., x_{dj(d)})^T,\ j = 1, ..., n$, are independent and identically distributed as a d-dimensional multivariate distribution with mean zero and nonnegative definite covariance matrix $\boldsymbol{\Sigma}_d$. The eigenvalue decomposition of $\boldsymbol{\Sigma}_d$ is $\boldsymbol{\Sigma}_d = \boldsymbol{V}_d \boldsymbol{\Lambda}_d \boldsymbol{V}_d^T$, where $\boldsymbol{\Lambda}_d$ is a diagonal matrix of eigenvalues $\lambda_{1(d)} \geq \cdots \geq \lambda_{d(d)} \geq 0$ and $\boldsymbol{V}_d$ is an orthogonal matrix of corresponding eigenvectors. Then, $\boldsymbol{Z}_{(d)} = \boldsymbol{\Lambda}_d^{-1/2} \boldsymbol{V}_d^T \boldsymbol{X}_{(d)}$ is considered as a $d \times n$ data matrix from a distribution with the identity covariance matrix. Here, we write $\boldsymbol{Z}_{(d)}^T = [\boldsymbol{z}_{1(d)}, ..., \boldsymbol{z}_{d(d)}]$ and $\boldsymbol{z}_{i(d)}^T = (z_{i1(d)}, ..., z_{in(d)}),\ i = 1, ..., d$. Hereafter, the subscript d will be omitted for the sake of simplicity when it does not cause any confusion. We assume that the fourth moments of each variable are uniformly bounded and $||\boldsymbol{z}_i|| \neq 0$ for $i = 1, ..., d$, where $|| \cdot ||$ denotes the Euclidean norm. We consider a general setting as follows:

$$\lambda_i = a_i d^{\alpha_i}\ (i = 1, ..., m) \quad \text{and} \quad \lambda_j = c_j\ (j = m + 1, ..., d). \tag{1}$$

Here, $a_i(> 0),\ c_j(\geq 0)$ and $\alpha_i(\alpha_1 \geq \cdots \geq \alpha_m > 0)$ are unknown constants preserving the ordering that $\lambda_1 \geq \cdots \geq \lambda_d$, and m is an unknown positive integer. The experimenter determines the threshold level d^γ with a fixed $\gamma\ (> 0)$. Let k be the maximum integer $i\ (\leq m)$ such that $\alpha_i > \gamma$. We assume that $\gamma \neq \alpha_i\ (i = 1, ..., m)$, so that $\alpha_k > \gamma > \alpha_{k+1}$ and k is the number of the eigenvalues beyond the threshold level. In this paper, we consider k as ID that is the target to estimate.

In Section 2, a new ID estimating methodology is given and its properties are studied by using a d-asymptotic approach. In Section 3, we summarize the findings about the efficiency of the proposed methodology with the help of computer simulations. In Section 4, we demonstrate how the new methodology estimates ID of HDLSS data by using a gene expression dataset. We lay down lengthy proofs in the appendix.

2. ESTIMATION OF ID

The sample covariance matrix is $\boldsymbol{S} = n^{-1} \boldsymbol{X} \boldsymbol{X}^T$, and the $n \times n$ dual sample covariance matrix is defined by $\boldsymbol{S}_D = n^{-1} \boldsymbol{X}^T \boldsymbol{X}$. Note that $\boldsymbol{S}_D$ has the same eigenvalues as $\boldsymbol{S}$. Let us

write that

$$n\boldsymbol{S}_D = \boldsymbol{Z}^T \boldsymbol{\Lambda} \boldsymbol{Z} = \sum_{i=1}^{d} \lambda_i \boldsymbol{W}_i, \tag{2}$$

where $\boldsymbol{W}_i = \boldsymbol{z}_i \boldsymbol{z}_i^T$, $i = 1, ..., d$. Note that $E\{(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D\} = \boldsymbol{I}_n$. Ahn et al. (2007) claim that when the eigenvalues of $\boldsymbol{\Sigma}$ are sufficiently diffused in the sense that

$$\sum_{i=1}^{d} \tilde{\lambda}_i^2 \to 0 \quad \text{as} \quad d \to \infty, \quad \text{where} \quad \tilde{\lambda}_i = \lambda_i/(\textstyle\sum_{i=1}^{d} \lambda_i), \tag{3}$$

the sample eigenvalues behave as if they are from an identity covariance matrix. If $\boldsymbol{X}$ is Gaussian, the elements of $\boldsymbol{Z}$ are independent and standard univariate normal variables. Hence, as they claimed, it follows that $(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D \to \boldsymbol{I}_n$ w.p.1 as $d \to \infty$ with a fixed n under (3). If $\boldsymbol{X}$ is non-Gaussian, by Chebyshev's inequality, for any $\tau > 0$ and the uniform bound M for the fourth moments condition, one has for each off-diagonal element $(i' \neq j')$ of $(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D$ as $d \to \infty$ with a fixed n that $P(|\sum_{i=1}^{d} \tilde{\lambda}_i z_{ii'} z_{ij'}| > \tau) \leq \tau^{-2}\mathrm{var}(\sum_{i=1}^{d} \tilde{\lambda}_i z_{ii'} z_{ij'}) \leq \tau^{-2} \sum_{i=1}^{d} \tilde{\lambda}_i^2 \to 0$. Thus each off-diagonal element of $(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D$ converges to 0 in probability as $d \to \infty$ with a fixed n under (3). However, one has for each diagonal element (i') of $(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D$ as $d \to \infty$ that $P(|\sum_{i=1}^{d} \tilde{\lambda}_i z_{ii'}^2 - 1| > \tau) \leq \tau^{-2}\mathrm{var}(\sum_{i=1}^{d} \tilde{\lambda}_i z_{ii'}^2) = \tau^{-2}\{\sum_{i=1}^{d} \tilde{\lambda}_i^2 \mathrm{var}(z_{ii'}^2) + \sum_{i \neq j} \tilde{\lambda}_i \tilde{\lambda}_j \mathrm{cov}(z_{ii'}^2, z_{ji'}^2)\} \leq \tau^{-2} M \neq 0$, so that any diagonal element of $(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D$ has $O_p(1)$ and may not converge to 1 under (3). Hence, when $\boldsymbol{X}$ is non-Gaussian, we may claim that the matrix $(n/\sum_{i=1}^{d} \lambda_i)\boldsymbol{S}_D$ converges to a diagonal matrix with any diagonal element having $O_p(1)$ as $d \to \infty$ with a fixed n under (3). Therefore, no matter whether $\boldsymbol{X}$ is Gaussian or non-Gaussian, it is difficult to find a difference among the eigenvalues under (3) with a fixed n. We emphasize that the setting in (1), provided that $\alpha_1 < 1$ and $c_d > 0$, includes the case satisfying (3). Our new methodology attempts estimating ID of HDLSS data in such a situation as well by detecting differences among the eigenvalues clearly. Only when n is fixed, we suppose that the assumptions (A1) and (A2) hold:

(A1) There exists a constant ε_j (> 0) such that $||n^{-1/2}\boldsymbol{z}_j|| > \varepsilon_j$ w.p.1 as $d \to \infty$ for each j $(= 1, ..., k)$;

(A2) When $k \geq 2$, there exists a constant η_j (> 0) such that

$$|\text{Angle}(\boldsymbol{z}_j,\ \text{span}\{\boldsymbol{z}_1, ..., \boldsymbol{z}_{j-1}, \boldsymbol{z}_{j+1}, ..., \boldsymbol{z}_k\})| > \eta_j$$

w.p.1 as $d \to \infty$ for each j $(= 1, ..., k)$.

We suppose that the properties of $\boldsymbol{Z}$ still remain under (A1) and (A2). We first obtain the following theorem.

Theorem 2.1. *Assume that $n \geq k + 1$. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n \geq 0$ be the eigenvalues of $\boldsymbol{S}_D$. For $\gamma > 1/2$, consider the maximum integer i $(= \hat{k}_1)$ such that $d^{-\gamma} \hat{\lambda}_i \geq 1$ as an estimate of ID, k. Then, we have that $\hat{k}_1 \to k$ in probability as*

(I) $d \to \infty$ and $d^{2-2\gamma}/n \to 0$ when $\gamma \in (1/2, 1]$;

(II) $d \to \infty$ and either $n \to \infty$ or n is a fixed number satisfying (A1)–(A2) when $\gamma > 1$.

Remark 1. Let us demonstrate how to specify the threshold level d^γ. Let $\beta = \beta(d)$ be the noise level specified by the experimenter such that $0 < \beta(d) < 1$ and $\beta(d) \to 0$ as $d \to \infty$. We consider that each noise effect is less than $100\beta\%$ of the sum of all eigenvalues of $\boldsymbol{\Sigma}$. Then, we choose γ so as to satisfy the equation that $d^\gamma = \beta\text{tr}(\boldsymbol{\Sigma})$. Since $\boldsymbol{\Sigma}$ is unknown, we estimate γ by solving the equation that $d^{\hat{\gamma}} = \beta\text{tr}(\boldsymbol{S})$ instead. We first consider the case when n is fixed. We assume that there exists a constant ε_j (> 0) such that $||n^{-1/2} \boldsymbol{z}_j|| > \varepsilon_j$ w.p.1 as $d \to \infty$ for $j = 1, ..., d$. From the assumption, one has that $\text{tr}(\boldsymbol{S}) = n^{-1} \sum_{j=1}^{d} \lambda_j ||\boldsymbol{z}_j||^2 > \sum_{j=1}^{d} \lambda_j \varepsilon_j^2$. Hence, there exists a constant ε (> 0) such that $\text{tr}(\boldsymbol{S}) > \varepsilon\text{tr}(\boldsymbol{\Sigma})$ w.p.1. From the fourth moments condition, we have that $\text{tr}(\boldsymbol{S}) = O_p(\text{tr}(\boldsymbol{\Sigma}))$. Hence, there exists a random variable, $c_s \in (0, \infty)$, such that $\text{tr}(\boldsymbol{S}) = c_s\text{tr}(\boldsymbol{\Sigma})$. Then, it holds as $d \to \infty$ with a fixed n that $\hat{\gamma} = \log_d(\beta\text{tr}(\boldsymbol{S})) = \gamma + \log_d(c_s) = \gamma + o_p(1)$. For the case that $n \to \infty$, since we have as $d \to \infty$ and $n \to \infty$ that $\text{tr}(\boldsymbol{S})/\text{tr}(\boldsymbol{\Sigma}) = 1 + o_p(1)$, it holds that $\hat{\gamma} = \gamma + o_p(1)$.

Corollary 2.1. *Assume that $\alpha_1 > \alpha_2$ or $m = 1$ in (1). Recall that $\hat{\lambda}_1$ is the maximum eigenvalue of $\boldsymbol{S}_D$. Then, we have that $\hat{\lambda}_1/\lambda_1 = 1 + o_p(1)$ either as $d \to \infty$ and $d^{2-2\alpha_1}/n \to 0$ for $\alpha_1 \in (1/2, 1]$ or as $d \to \infty$ and $n \to \infty$ for $\alpha_1 > 1$.*

For $\gamma \in (1/2, 1]$ with a fixed n, one can not apply Theorem 2.1 to estimation of ID. In order to overcome this difficulty, we consider a new dual approach to attempt relaxing the convergence condition with respect to n. Suppose we have two $d \times n$ data matrices $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in}]$, $i = 1, 2$, where $\boldsymbol{x}_{ij} = (x_{i1j}, ..., x_{idj})^T$, $i = 1, 2$; $j = 1, ..., n$, are independent and identically distributed as a d-dimensional multivariate distribution stated before. We systematically write $\boldsymbol{S}^2 = n^{-2}\boldsymbol{X}_1\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{X}_2^T$ and define the $n \times n$ dual sample square matrix as $\boldsymbol{S}_D^2 = n^{-2}\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{X}_2^T\boldsymbol{X}_1$. Note that $\boldsymbol{S}_D^2$ has the same eigenvalues as $\boldsymbol{S}^2$. Let $\boldsymbol{Z}_i = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}^T\boldsymbol{X}_i$, $i = 1, 2$, and let us write that

$$n^2\boldsymbol{S}_D^2 = \boldsymbol{Z}_1^T\boldsymbol{\Lambda}\boldsymbol{Z}_2\boldsymbol{Z}_2^T\boldsymbol{\Lambda}\boldsymbol{Z}_1 = \Big(\sum_{i=1}^d \lambda_i\boldsymbol{Y}_i^T\Big)\Big(\sum_{i=1}^d \lambda_i\boldsymbol{Y}_i\Big), \tag{4}$$

where $\boldsymbol{Z}_1^T = [\boldsymbol{z}_{11}, ..., \boldsymbol{z}_{1d}]$, $\boldsymbol{z}_{1i}^T = (z_{1i1}, ..., z_{1in})$, $\boldsymbol{Z}_2^T = [\boldsymbol{z}_{21}, ..., \boldsymbol{z}_{2d}]$, $\boldsymbol{z}_{2i}^T = (z_{2i1}, ..., z_{2in})$ and $\boldsymbol{Y}_i = \boldsymbol{z}_{2i}\boldsymbol{z}_{1i}^T$, $i = 1, ..., d$. Note that $E\{(n/\sum_{i=1}^d \lambda_i^2)\boldsymbol{S}_D^2\} = \boldsymbol{I}_n$. Let $\tilde{\tilde{\lambda}}_i = \lambda_i^2/(\sum_{i=1}^d \lambda_i^2)$. By using Chebyshev's inequality, for any $\tau > 0$ and the uniform bound M for the fourth moments condition, one has for each off-diagonal element $(i' \neq j')$ of $(n/\sum_{i=1}^d \lambda_i^2)\boldsymbol{S}_D^2$ that $P(|\sum_{i,j}\sqrt{\tilde{\tilde{\lambda}}_i\tilde{\tilde{\lambda}}_j}(\boldsymbol{z}_{2i}/\sqrt{n})^T(\boldsymbol{z}_{2j}/\sqrt{n})z_{1ii'}z_{1jj'}| > \tau) \leq \tau^{-2}\text{var}(\sum_{i,j}\sqrt{\tilde{\tilde{\lambda}}_i\tilde{\tilde{\lambda}}_j}(\boldsymbol{z}_{2i}/\sqrt{n})^T(\boldsymbol{z}_{2j}/\sqrt{n})z_{1ii'}z_{1jj'}) \leq \tau^{-2}M\sum_{i,j}\tilde{\tilde{\lambda}}_i\tilde{\tilde{\lambda}}_j = \tau^{-2}M \neq o(1)$. Hence, we may not claim under (3) that any off-diagonal element of $(n/\sum_{i=1}^d \lambda_i^2)\boldsymbol{S}_D^2$ converges to 0, so that the matrix $(n/\sum_{i=1}^d \lambda_i^2)\boldsymbol{S}_D^2$ may not converge to even a diagonal matrix as $d \to \infty$ with a fixed n under (3). It gives us a hint of another ID estimating methodology to detect differences among the eigenvalues of $\boldsymbol{\Sigma}$ by using a d-asymptotic approach. Only when n is fixed, we suppose that the assumptions (A1') and (A2') hold:

(A1') There exists a constant ε_{ij} (> 0) such that $||n^{-1/2}\boldsymbol{z}_{ij}|| > \varepsilon_{ij}$ w.p.1 as $d \to \infty$ for each i, j $(i = 1, 2;\ j = 1, ..., k)$;

(A2') When $k \geq 2$, there exists a constant η_{ij} (> 0) such that

$$|\text{Angle}(\boldsymbol{z}_{ij},\ \text{span}\{\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{ij-1}, \boldsymbol{z}_{ij+1}, ..., \boldsymbol{z}_{ik}\})| > \eta_{ij}$$

w.p.1 as $d \to \infty$ for each i, j $(i = 1, 2;\ j = 1, ..., k)$.

We suppose that the properties of $\boldsymbol{Z}_i$, $i = 1, 2$, still remain under (A1') and (A2'). Then, we obtain the following theorem.

Theorem 2.2. *Assume that $n \geq k+1$. Let $\hat{\lambda}_1^2 \geq \cdots \geq \hat{\lambda}_n^2 \geq 0$ be the eigenvalues of S_D^2. Consider the maximum integer i $(= \hat{k}_2)$ such that $d^{-2\gamma}\hat{\lambda}_i^2 \geq 1$ as an estimate of ID, k. Then, we have for $\gamma > 1/2$ that $\hat{k}_2 \to k$ in probability either as $d \to \infty$ and $n \to \infty$ or as $d \to \infty$ while n is a fixed number satisfying (A1')–(A2').*

Remark 2. Assume that X_i, $i = 1, 2$, are Gaussian. Then, we can extend the range of allowable γ thresholds to $\gamma \in (1/4, 1/2]$ to claim the assertion in Theorem 2.2 as $d \to \infty$ and $d^{2-4\gamma}/n \to 0$.

Remark 3. Suppose that we have a $d \times n$ data matrix, $X = [x_1, ..., x_n] = [x_{11}, ..., x_{1n_1}, x_{21}, ..., x_{2n_2}]$, where $n_1 + n_2 = n$. One may define X_1 and X_2 as $X_i = [x_{i1}, ..., x_{in_i}]$, $i = 1, 2$. We suggest that one may take $n_1 = n_2$ $(= n')$ when $n = 2n'$ or $n_1 = n'+1$ and $n_2 = n'$ when $n = 2n'+1$. Then, one may generally define $S_D^2 = (n_1 n_2)^{-1} X_1^T X_2 X_2^T X_1$.

Corollary 2.2. *When the population mean may not be zero, let $\overline{X}_i = [\bar{x}_{i1}, ..., \bar{x}_{id}]^T$ $(i = 1, 2)$ having n-vector $\bar{x}_{ij} = (\bar{x}_{ij}, ..., \bar{x}_{ij})^T$ with $\bar{x}_{ij} = \sum_{s=1}^{n} x_{ijs}/n$ for each j $(= 1, ..., d)$. Let us write that $\Lambda^{-1/2} V^T (X_i - \overline{X}_i) = [\acute{z}_{i1}, ..., \acute{z}_{id}]^T$ $(i = 1, 2)$. Assume $n \geq k+2$ and define S_D^2 after replacing X_i with $X_i - \overline{X}_i$. Then, the assertion in Theorem 2.2 is still justified under the convergence condition given by replacing z_{ij} with $\acute{z}_{ij}$ in (A1')–(A2').*

Corollary 2.3. *Assume that $\alpha_1 > \alpha_2$ or $m = 1$ in (1). Recall that $\hat{\lambda}_1^2$ is the maximum eigenvalue of S_D^2. Then, we have that $\lambda_1^{-1}\sqrt{\hat{\lambda}_1^2} = 1 + o_p(1)$ as $d \to \infty$ and $n \to \infty$ for $\alpha_1 > 1/2$.*

Remark 4. Earlier literature may not handle the case that $\alpha_1 \leq 1$ when $n/d \to 0$ for estimating the maximum eigenvalue. One can use Corollaries 2.1 and 2.3 for the case. We conducted computer simulations with the following setup: $d = 1000$; $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (d^{2/3}, d^{1/2}, d^{1/3}, d^{1/6})$ and $\lambda_i = 1$, $i = 5, ..., d$, so that $\lambda_1 = 1000^{2/3} = 100$. We considered that (i) $n = 20$ in Corollary 2.1 ($n = 10$ in Corollary 2.3) and (ii) $n = 40$ in Corollary 2.1 ($n = 20$ in Corollary 2.3). By averaging the outcomes from 1000 replications, we obtained from Corollary 2.1 that $\hat{\lambda}_1 = 152.1$ for (i) and $\hat{\lambda}_1 = 126.7$ for (ii). On the other hand, we

obtained from Corollary 2.3 that $\sqrt{\hat{\lambda}_1^2} = 96.7$ for (i) and $\sqrt{\hat{\lambda}_1^2} = 98.7$ for (ii). We observed superiority of $\sqrt{\hat{\lambda}_1^2}$ in average to $\hat{\lambda}_1$ for other parameter configurations as well. We emphasize that Corollaries 2.1 and 2.3 are applicable for the case that $\alpha_1 > 1$ as well.

3. SIMULATION

In order to study the performance of the ID estimating methodologies, we resort to computer simulations. We fixed ID at $k = 4$ and the sample size at $n = 30 \; (= 15 + 15)$. We set $\gamma = 3/5$, namely the threshold level is $d^{3/5}$. We conducted numerous simulation studies. However, we omit the details and present a case for brevity. We considered that $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (d, d^{9/10}, d^{4/5}, d^{7/10})$, $(\lambda_5, \lambda_6, \lambda_7, \lambda_8) = (d^{1/2}, d^{2/5}, d^{3/10}, d^{1/5})$ and $\lambda_i = 1$, $i = 9, ..., d$. In Figs.1-3, we evaluated the performance of the ID estimating methodologies given by Theorems 2.1 and 2.2 in terms of the average ID number and the probability of correct decision. We used the whole sample of size $n = 30$ to define the data matrix $\boldsymbol{X} : d \times 30$ for the calculation of $\boldsymbol{S}_D$ in Theorem 2.1, whereas we divided the whole sample into $\boldsymbol{X}_1 : d \times 15$ and $\boldsymbol{X}_2 : d \times 15$ for the calculation of $\boldsymbol{S}_D^2$ in Theorem 2.2. The findings were obtained by averaging the outcomes from $1000 \; (= R$, say) replications. Under a fixed scenario, suppose that the rth replication ends with estimate $k_r \; (r = 1, ..., R)$, for the ID estimating methodology, $\hat{k}_1$ (or $\hat{k}_2$), given by Theorem 2.1 (or Theorem 2.2). Let us simply write $\hat{k} = R^{-1} \sum_{r=1}^{R} k_r$ for each ID estimating methodology. Fig.1 shows that $\hat{k}_2$ estimates ID ($k = 4$) better than $\hat{k}_1$ for a long span of $d \in [500, 1500]$. We also consider the Monte Carlo variability. Let us write $V(\hat{k}) = R^{-1} \sum_{r=1}^{R} k_r^2 - (R^{-1} \sum_{r=1}^{R} k_r)^2$ for the sample variance of each ID estimation. Fig.2 shows that $\hat{k}_2$ keeps variance $V(\hat{k}_2)$ lower than $\hat{k}_1$. At the end of the rth replication, we also checked whether it holds that $k_r = k \; (= 4)$, and defined $p_r = 1$ (or 0) according as $k_r = k$ (or $k_r \neq k$), $r = 1, ..., R$. Then, $\overline{p} = R^{-1} \sum_{r=1}^{R} p_r$ estimates the probability of correct decision, $P(\hat{k}_1 = k)$ (or $P(\hat{k}_2 = k)$), for each ID estimating methodology. Fig.3 shows that $\overline{p}_1$ estimating $P(\hat{k}_1 = k)$ decreases as d increases, while $\overline{p}_2$ estimating $P(\hat{k}_2 = k)$ increases as d increases. As stated in Theorem 2.1, the experimenter needs to take samples depending on d in the ID estimating methodology, $\hat{k}_1$. The sample size fixed at $n = 30$ is not large enough

8

to use $\hat{k}_1$ efficiently. On the other hand, $\hat{k}_2$ estimates ID surprisingly well in such HDLSS cases.
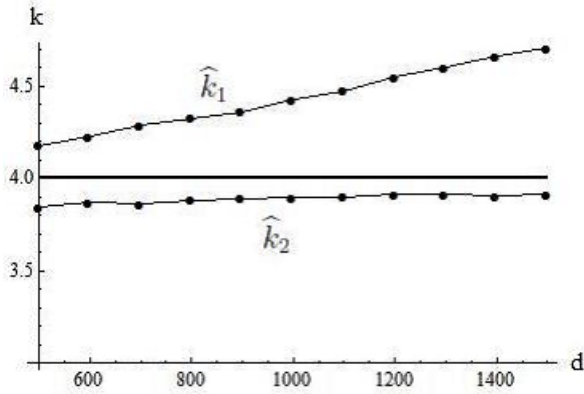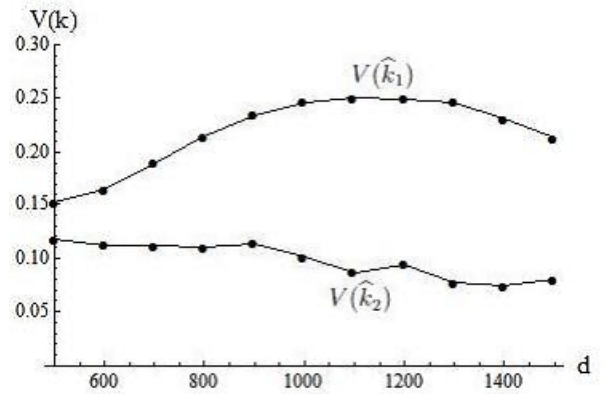

Fig.1. Average ID number
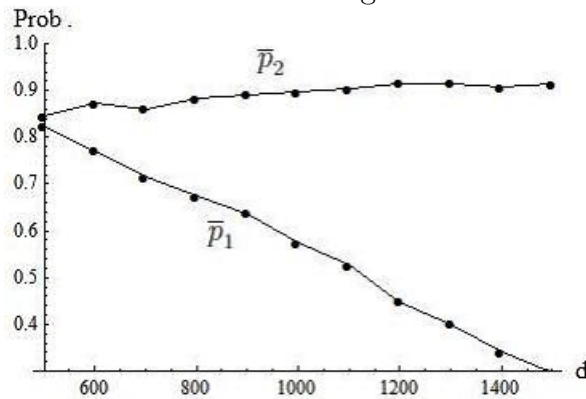

Fig.2. Variance of ID estimation

Fig.3. Probability of correct decision

4. EXAMPLE

We analyzed gene expression data given by Chiaretti et al. (2004) in which dataset consisted of $12625 (= d)$ genes and $128 (= 2n)$ microarrays from different patients. Note that the expression measures have been obtained using the three-step robust multichip average (RMA) preprocessing method. Refer to Pollard et al. (2005) as well for the details. Here, we had data matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of each size 12625×64. Let $\boldsymbol{X} : 12625 \times 128 = [\boldsymbol{X}_1, \boldsymbol{X}_2]$. We first specified the threshold level d^γ as follows. We considered that each noise effect is less than 5% (or $\beta = 0.05$) of the sum of all eigenvalues of $\boldsymbol{\Sigma}$. Then, with the help of Remark 1, we chose $\gamma = 0.524$ so as to satisfy the equation that $d^\gamma = \beta \mathrm{tr}(\boldsymbol{S})$, where

9

$\boldsymbol{S} = (\boldsymbol{X} - \overline{\boldsymbol{X}})(\boldsymbol{X} - \overline{\boldsymbol{X}})^T/128$. Since $\gamma > 1/2$, we used the ID estimating methodology, $\hat{k}_2$, given by Theorem 2.2. Here, $\boldsymbol{S}_D^2$ was defined in view of Corollary 2.2. We calculated the eigenvalues of $d^{-2\gamma}\boldsymbol{S}_D^2$ as $(d^{-2\gamma}\hat{\lambda}_1^2, d^{-2\gamma}\hat{\lambda}_2^2, d^{-2\gamma}\hat{\lambda}_3^2, d^{-2\gamma}\hat{\lambda}_4^2, ...) = (7.17, 1.53, 1.34, 0.48, ...)$. Hence, we obtained $\hat{k}_2 = 3$. So, we claimed that the ID of this HDLSS dataset is 3. In addition, we observed that $\hat{k}_2 = 2$ for $\beta = 0.06$ and $\hat{k}_2 = 3$ for $\beta = 0.04$.

APPENDIX

Throughout this section, let us write $\boldsymbol{R}_n = \{\boldsymbol{e}_n \in \boldsymbol{R}^n : ||\boldsymbol{e}_n|| = 1\}$. Let $\boldsymbol{U}_1 = n^{-1}\sum_{i=1}^k \lambda_i \boldsymbol{W}_i$ and $\boldsymbol{U}_2 = n^{-1}\sum_{i=k+1}^d \lambda_i \boldsymbol{W}_i$, where $\boldsymbol{W}_i$'s are defined in (2). Let $\boldsymbol{V}_1 = n^{-1}\sum_{i=1}^k \lambda_i \boldsymbol{Y}_i^T$ and $\boldsymbol{V}_2 = n^{-1}\sum_{i=k+1}^d \lambda_i \boldsymbol{Y}_i^T$, where $\boldsymbol{Y}_i$'s are defined in (4).

Lemma 1. *Assume that $n \geq k + 1$. Let $\hat{\delta}_1 \geq \cdots \geq \hat{\delta}_n \geq 0$ be the eigenvalues of $\boldsymbol{U}_1$. Then, it holds that $\liminf d^{-\gamma}\hat{\delta}_k > 1$ w.p.1 as $d \to \infty$ either when $n \to \infty$ or n is a fixed number satisfying assumptions (A1)–(A2).*

Proof. Let $\hat{\boldsymbol{z}}_j = (||n^{-1/2}\boldsymbol{z}_j||)^{-1}n^{-1/2}\boldsymbol{z}_j$ for $j = 1, ..., k$. Then, let us write that $d^{-\alpha_k}\boldsymbol{U}_1 = \sum_{j=1}^k a_j d^{\alpha_j - \alpha_k}||n^{-1/2}\boldsymbol{z}_j||^2 \hat{\boldsymbol{z}}_j \hat{\boldsymbol{z}}_j^T$. We first consider the case when n is fixed. From (A1), there exists a constant $\omega_j \ (> 0)$ such that $a_j d^{\alpha_j - \alpha_k}||n^{-1/2}\boldsymbol{z}_j||^2 > \omega_j$ w.p.1 as $d \to \infty$ for $j = 1, ..., k$. When $k = 1$, we can claim as $d \to \infty$ that there exists a constant $\zeta_1 \ (> 0)$ such that $d^{-\alpha_1}\hat{\delta}_1 > \zeta_1$ w.p.1. We consider the case when $k \geq 2$. Let us write $\boldsymbol{R}_{n,j} = \{\boldsymbol{e}_n \in \boldsymbol{R}_n : \boldsymbol{e}_n = \sum_{i=1(\backslash j)}^k b_i \hat{\boldsymbol{z}}_i, \ b_i \in R\}$, where $(\backslash j)$ excludes number j. From (A2), it holds as $d \to \infty$ that $\hat{\boldsymbol{z}}_j, \ j = 1, ..., k$, are linearly independent and there exists a constant $\xi_j \ (> 0)$ such that $|\hat{\boldsymbol{z}}_j^T \boldsymbol{e}_{nj}| > \xi_j$ w.p.1 as $d \to \infty$ for $j = 1, ..., k$, where $\boldsymbol{e}_{nj}$ is an arbitrary element of $\boldsymbol{R}_{n,j}$. Thus we can claim as $d \to \infty$ that there exists a constant $\zeta_j \ (> 0)$ such that $d^{-\alpha_k}\hat{\delta}_j > \zeta_j$ w.p.1 for $j = 1, ..., k$. Noting that $d^{\alpha_k - \gamma}\zeta_i > 1$ w.p.1 as $d \to \infty$, it holds that $\liminf d^{-\gamma}\hat{\delta}_k > 1$ w.p.1 as $d \to \infty$. Next, we consider the case when $n \to \infty$. From the facts that $||n^{-1/2}\boldsymbol{z}_i|| = 1 + o_p(1)$ and $n^{-1}\boldsymbol{z}_i^T \boldsymbol{z}_j = o_p(1)$ for $i \neq j$ as $n \to \infty$, we can claim (A1)–(A2) in the case. Thus, in a way similar to above, it concludes the results. $\square$

Proof of Theorem 2.1. Let us write $\boldsymbol{S}_D = \boldsymbol{U}_1 + \boldsymbol{U}_2$. We first consider the latter part, $\boldsymbol{U}_2$. When $k < m$, one has for all diagonal elements of $d^{-\gamma}\boldsymbol{U}_2$ as (I) or (II) that

$$\sum_{i'=1}^{n} P\Big((d^\gamma n)^{-1}\Big| \sum_{i=k+1}^{d} \lambda_i z_{ii'}^2 \Big| > \tau\Big)$$

$$\leq \tau^{-2} M d^{-2\gamma} n^{-1}\Big(\Big(\sum_{i=k+1}^{m} \lambda_i \Big)^2 + \Big(\sum_{i=m+1}^{d} \lambda_i \Big)^2 + 2\Big(\sum_{i=k+1}^{m} \lambda_i \Big)\Big(\sum_{i=m+1}^{d} \lambda_i \Big)\Big)$$

$$= O(d^{2\alpha_{k+1}-2\gamma}/n) + O(d^{2-2\gamma}/n) + O(d^{1+\alpha_{k+1}-2\gamma}/n) = o(1)$$

by using Chebyshev's inequality, for any $\tau > 0$ and the uniform bound M for the fourth moments condition. Thus all diagonal elements of $d^{-\gamma}\boldsymbol{U}_2$ have $o_p(1)$. Let us write $u_{i'j'}(= (d^\gamma n)^{-1} \sum_{i=k+1}^{d} \lambda_i z_{ii'} z_{ij'})$ for $i' \neq j'$ as an off-diagonal element of $d^{-\gamma}\boldsymbol{U}_2$. Then, by using Markov's inequality, we claim as $d \to \infty$ and either $n \to \infty$ or n is fixed that

$$P\Big(\sum_{i'\neq j'} u_{i'j'}^2 > \tau\Big) \leq \tau^{-1} d^{-2\gamma}\Big(\sum_{i=m+1}^{d} \lambda_i^2 \Big) + O(d^{2\alpha_{k+1}-2\gamma}) = O(d^{1-2\gamma}) + o(1) = o(1) \quad (5)$$

by noting that $\gamma > 1/2$. Thus we have $\sum_{i'\neq j'} u_{i'j'}^2 = o_p(1)$. Let $\boldsymbol{e}_n = (e_1, ..., e_n)^T$ be an arbitrary element of $\boldsymbol{R}_n$. Since it holds that $\sum_{i'\neq j'} e_{i'}^2 e_{j'}^2 \leq 1$, we obtain that

$$\sum_{i'\neq j'} e_{i'} e_{j'} u_{i'j'} = o_p(1). \quad (6)$$

Hence, we can claim as (I) or (II) that $d^{-\gamma}\boldsymbol{e}_n^T\boldsymbol{S}_D\boldsymbol{e}_n = d^{-\gamma}\boldsymbol{e}_n^T\boldsymbol{U}_1\boldsymbol{e}_n + o_p(1)$. When $k = m$, we can claim that $d^{-\gamma}\boldsymbol{e}_n^T\boldsymbol{S}_D\boldsymbol{e}_n = d^{-\gamma}\boldsymbol{e}_n^T\boldsymbol{U}_1\boldsymbol{e}_n + o_p(1)$ in a similar way. By applying Lemma 1 to the former part, $\boldsymbol{U}_1$, we obtain the result. $\square$

Proof of Corollary 2.1. Let us write that $\lambda_1^{-1}\boldsymbol{S}_D = \boldsymbol{B}_1 + \boldsymbol{B}_2$, where $\boldsymbol{B}_1 = n^{-1}\boldsymbol{W}_1 = n^{-1}\boldsymbol{z}_1\boldsymbol{z}_1^T$ and $\boldsymbol{B}_2 = (n\lambda_1)^{-1} \sum_{i=2}^{d} \lambda_i \boldsymbol{W}_i$. Noting that $\alpha_1 > \alpha_2$ or $m = 1$, similarly to the proof of Theorem 2.1, we have that $\boldsymbol{e}_n^T\boldsymbol{B}_2\boldsymbol{e}_n = o_p(1)$ for any $\boldsymbol{e}_n \in \boldsymbol{R}_n$ either as $d \to \infty$ and $d^{2-2\alpha_1}/n \to 0$ when $\alpha_1 \in (1/2, 1]$ or as $d \to \infty$ and $n \to \infty$ when $\alpha_1 > 1$. From the fact that $||n^{-1/2}\boldsymbol{z}_1|| = 1 + o_p(1)$ as $n \to \infty$, we claim either as $d \to \infty$ and $d^{2-2\alpha_1}/n \to 0$ for $\alpha_1 \in (1/2, 1]$ or as $d \to \infty$ and $n \to \infty$ for $\alpha_1 > 1$ that $\max(\boldsymbol{e}_n^T\lambda_1^{-1}\boldsymbol{S}_D\boldsymbol{e}_n) = \max(\boldsymbol{e}_n^T n^{-1}\boldsymbol{z}_1\boldsymbol{z}_1^T\boldsymbol{e}_n + o_p(1)) = 1 + o_p(1)$

11

with respect to any $\boldsymbol{e}_n \in \boldsymbol{R}_n$. It concludes the result. $\square$

Lemma 2. *Assume that $n \geq k + 1$. Let $\tilde{\delta}_1 \geq \cdots \geq \tilde{\delta}_k \geq 0$ be singular values of $\boldsymbol{V}_1$. Let $\tilde{\boldsymbol{u}}_{j(1)} \in \boldsymbol{R}_n$ be a left-singular vector and $\tilde{\boldsymbol{u}}_{j(2)} \in \boldsymbol{R}_n$ be a right-singular vector corresponding to $\tilde{\delta}_j$ ($j = 1, ..., k$). Let us write the singular value decomposition as $\boldsymbol{V}_1 = \sum_{j=1}^k \tilde{\delta}_j \tilde{\boldsymbol{u}}_{j(1)} \tilde{\boldsymbol{u}}_{j(2)}^T$. Then, it holds that $\liminf d^{-\gamma} \tilde{\delta}_k > 1$ w.p.1 as $d \to \infty$ and either $n \to \infty$ or n is a fixed number satisfying (A1')-(A2').*

Proof. Let $\hat{\boldsymbol{z}}_{ij} = (\|n^{-1/2} \boldsymbol{z}_{ij}\|)^{-1} n^{-1/2} \boldsymbol{z}_{ij}$ for $j = 1, ..., k$ ($i = 1, 2$). Then, we have that $d^{-\alpha_k} \boldsymbol{V}_1 = \sum_{j=1}^k a_j d^{\alpha_j - \alpha_k} (\|n^{-1/2} \boldsymbol{z}_{1j}\| \hat{\boldsymbol{z}}_{1j})(\|n^{-1/2} \boldsymbol{z}_{2j}\| \hat{\boldsymbol{z}}_{2j})^T$. The result is obtained in similar fashion to the proof of Lemma 1. $\square$

Lemma 3. *We have for $\gamma > 1/2$ that $d^{-\gamma} \boldsymbol{e}_{1n}^T \boldsymbol{V}_2 \boldsymbol{e}_{2n} = o_p(1)$ as $d \to \infty$ either when $n \to \infty$ or n is a fixed number, where $\boldsymbol{e}_{1n}$ and $\boldsymbol{e}_{2n}$ are arbitrary elements of $\boldsymbol{R}_n$.*

Proof. Let us write $v_{i'j'} = n^{-1} \sum_{i=k+1}^d \lambda_i z_{1ii'} z_{2ij'}$ as an (i', j') element of $\boldsymbol{V}_2$. We first consider off-diagonal elements of $\boldsymbol{V}_2$. We have that $E\{n^2 (d^{-\gamma} v_{i'j'})^2\} = O(d^{1-2\gamma}) + o(1) = o(1)$ for $i' \neq j'$. Thus in a way similar to (5)-(6), we claim that $d^{-\gamma} \boldsymbol{e}_{1n}^T (\boldsymbol{V}_2 - \text{diag}(v_{11}, ..., v_{nn})) \boldsymbol{e}_{2n} = o_p(1)$. Next, we consider diagonal elements of $\boldsymbol{V}_2$. One has for all diagonal elements of $d^{-\gamma} \boldsymbol{V}_2$ as $d \to \infty$ that

$$\sum_{i'=1}^n P\left(d^{-\gamma} |v_{i'i'}| > \tau\right) = \sum_{i'=1}^n P\left((d^\gamma n)^{-1} \left| \sum_{i=k+1}^d \lambda_i z_{1ii'} z_{2ii'} \right| > \tau\right)$$
$$\leq \tau^{-2} d^{-2\gamma} n^{-1} \left(\sum_{i=k+1}^d \lambda_i^2 \right) = O(d^{1-2\gamma}/n) + o(1) = o(1)$$

by using Chebyshev's inequality, for any $\tau > 0$. Thus all diagonal elements of $d^{-\gamma} \boldsymbol{V}_2$ have $o_p(1)$. It concludes the result. $\square$

Proof of Theorem 2.2. Let us write that $\boldsymbol{S}_{D(1)} = n^{-1} \sum_{i=1}^d \lambda_i \boldsymbol{Y}_i^T = \boldsymbol{V}_1 + \boldsymbol{V}_2$. Let $\sqrt{\hat{\lambda}_1^2} \geq \cdots \geq \sqrt{\hat{\lambda}_n^2} \geq 0$ be singular values of $\boldsymbol{S}_{D(1)}$. Let $\hat{\boldsymbol{u}}_{j(1)} \in \boldsymbol{R}_n$ be a left-singular vector and $\hat{\boldsymbol{u}}_{j(2)} \in \boldsymbol{R}_n$ be a right-singular vector corresponding to $\sqrt{\hat{\lambda}_j^2}$ ($j = 1, ..., n$). Then, we have the singular value decomposition as $\boldsymbol{S}_{D(1)} = \sum_{j=1}^n \sqrt{\hat{\lambda}_j^2} \hat{\boldsymbol{u}}_{j(1)} \hat{\boldsymbol{u}}_{j(2)}^T$. From Lemmas 2 and 3, it

12

holds for $\gamma > 1/2$ that $\liminf d^{-\gamma}\sqrt{\hat{\lambda}_k^2} > 1$ w.p.1 and $d^{-\gamma}\sqrt{\hat{\lambda}_j^2} = o_p(1)$ for $j = k+1, ..., n$, as $d \to \infty$ either when $n \to \infty$ or n is a fixed number satisfying (A1')–(A2'). Noting that $\boldsymbol{S}_D^2 = \boldsymbol{S}_{D(1)}\boldsymbol{S}_{D(1)}^T = \sum_{j=1}^n \hat{\lambda}_j^2 \hat{\boldsymbol{u}}_{j(1)}\hat{\boldsymbol{u}}_{j(1)}^T$, it concludes the result. $\qquad\square$

Proof of Corollary 2.2. Let us write that $\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}^T(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i) = [\acute{\boldsymbol{z}}_{i1}, ..., \acute{\boldsymbol{z}}_{id}]^T$ and $\acute{\boldsymbol{z}}_{ij} = (\acute{z}_{ij1}, ..., \acute{z}_{ijn})^T$ for $i = 1, 2$ and $j = 1, ..., d$. Then, we have that $\acute{z}_{ijl} = z_{ijl} - \bar{z}_{ij}$ for $l = 1, ..., n$, where $\bar{z}_{ij} = \sum_{l=1}^n z_{ijl}/n$. Let $E(z_{ijl}) = \mu_j$ for $j = 1, ..., d$. We write that $\acute{z}_{ijl} = \tilde{z}_{ijl} + z_{oij}$, where $\tilde{z}_{ijl} = z_{ijl} - \mu_j$ and $z_{oij} = \mu_j - \bar{z}_{ij}$ $(i = 1, 2; \ j = 1, ..., d; \ l = 1, ..., n)$. Now, let us write that n-vectors $\tilde{\boldsymbol{z}}_{ij} = (\tilde{z}_{ij1}, ..., \tilde{z}_{ijn})^T$ and $\boldsymbol{z}_{oij} = (z_{oij}, ..., z_{oij})^T$ for $i = 1, 2$ and $j = 1, ..., d$. Then, we can write that $(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2) = \sum_{j=1}^d \lambda_j(\tilde{\boldsymbol{z}}_{1j} + \boldsymbol{z}_{o1j})(\tilde{\boldsymbol{z}}_{2j} + \boldsymbol{z}_{o2j})^T$. Let $\boldsymbol{V}_o = n^{-1}\sum_{j=k+1}^d \lambda_j(\tilde{\boldsymbol{z}}_{1j} + \boldsymbol{z}_{o1j})(\tilde{\boldsymbol{z}}_{2j} + \boldsymbol{z}_{o2j})^T$. We first consider the term $n^{-1}\sum_{j=k+1}^d \lambda_j\tilde{\boldsymbol{z}}_{1j}\boldsymbol{z}_{o2j}^T$. Let us write that $v_{oi'j'} = n^{-1}\sum_{j=k+1}^d \lambda_j\tilde{z}_{1ji'}z_{o2j}$ as (i, j) element of $n^{-1}\sum_{j=k+1}^d \lambda_j\tilde{\boldsymbol{z}}_{1j}\boldsymbol{z}_{o2j}^T$. Then, we have for $\gamma > 1/2$ that $E\{n^2(d^{-\gamma}v_{oi'j'})^2\} = O(d^{1-2\gamma}) + o(1) = o(1)$. Thus, in a way similar to the proof of Lemma 3, we claim for $\gamma > 1/2$ that $n^{-1}d^{-\gamma}\boldsymbol{e}_{1n}^T\sum_{j=k+1}^d \lambda_j\tilde{\boldsymbol{z}}_{1j}\boldsymbol{z}_{o2j}^T\boldsymbol{e}_{2n} = o_p(1)$ as $d \to \infty$ for any $\boldsymbol{e}_{1n}, \boldsymbol{e}_{2n} \in \boldsymbol{R}_n$. Similarly, we claim for $\gamma > 1/2$ that $n^{-1}d^{-\gamma}\boldsymbol{e}_{1n}^T\sum_{j=k+1}^d \lambda_j\boldsymbol{z}_{o1j}\boldsymbol{z}_{o2j}^T\boldsymbol{e}_{2n} = o_p(1)$ as $d \to \infty$. Thus we have for $\gamma > 1/2$ that $n^{-1}d^{-\gamma}\boldsymbol{e}_{1n}^T\boldsymbol{V}_o\boldsymbol{e}_{2n} = n^{-1}d^{-\gamma}\boldsymbol{e}_{1n}^T\sum_{j=k+1}^d \lambda_j\tilde{\boldsymbol{z}}_{1j}\tilde{\boldsymbol{z}}_{2j}^T\boldsymbol{e}_{2n} + o_p(1)$ as $d \to \infty$. Here, let us write $\boldsymbol{V}_{o2} = n^{-1}\sum_{j=k+1}^d \lambda_j\tilde{\boldsymbol{z}}_{1j}\tilde{\boldsymbol{z}}_{2j}^T$. Then, note that $\boldsymbol{V}_{o2}$ is essentially equal to $\boldsymbol{V}_2$. Hence, we can claim the assertion in Lemma 3 by replacing z_{ijl} with $\tilde{z}_{ijl}$. Thus, similarly to the proof of Theorem 2.2, we obtain the result in Theorem 2.2 given by replacing both $n \geq k+1$ and $\boldsymbol{z}_{ij}$ in (A1')–(A2') with $n \geq k+2$ and $\acute{\boldsymbol{z}}_{ij}$ respectively. $\qquad\square$

Proof of Corollary 2.3. Let $\boldsymbol{S}_{D(1)} = n^{-1}\sum_{i=1}^d \lambda_i\boldsymbol{Y}_i^T$ as before. Let us write $\lambda_1^{-1}\boldsymbol{S}_{D(1)} = \boldsymbol{M}_1 + \boldsymbol{M}_2$, where $\boldsymbol{M}_1 = n^{-1}\boldsymbol{Y}_1^T$ and $\boldsymbol{M}_2 = (n\lambda_1)^{-1}\sum_{i=2}^d \lambda_i\boldsymbol{Y}_i^T$. Note that $\alpha_1 > \alpha_2$ or $m = 1$. In a way similar to the proof of Lemma 3, we have for $\alpha_1 > 1/2$ that $\boldsymbol{e}_{1n}^T\boldsymbol{M}_2\boldsymbol{e}_{2n} = o_p(1)$ for any $\boldsymbol{e}_{1n}, \boldsymbol{e}_{2n} \in \boldsymbol{R}_n$ as $d \to \infty$ and $n \to \infty$. The result can be obtained similarly to the proof of Corollary 2.1. $\qquad\square$

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

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

Bai, Z., Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Prob.* 26: 316–345.

Baik, J., Ben Arous, G., Peche, S. (2005). Phase transition of the largest eigenvalue for non-null complex covariance matrices. *Ann. Prob.* 33: 1643–1697.

Baik, J., Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Mult. Anal.* 97: 1382–1408.

Bruske, J., Sommer, G. (1998). Intrinsic dimensionality estimation with optimally topology preserving maps. *IEEE Trans. on PAMI* 20: 572–575.

Camastra, F., Vinciarelli, A. (2002). Estimating the intrinsic dimension of data with a fractal-based approach. *IEEE Trans. on PAMI* 24: 1404–1407.

Chiaretti, S., Li, X., Gentleman, R., Vitale, A., Vignetti, M., Mandelli, F., Ritz, J., Foa, R. (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood* 103: 2771–2778.

Costa, J., Hero, A. O. (2004). Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. on Signal Processing* 52: 2210–2221.

Donoho, D. L., Grimes, C. (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *PANS* 100: 5591–5596.

Fukunaga, K., Olsen, D. R. (1971). An algorithm for finding intrinsic dimensionality of data. *IEEE Trans. on Computers* C-20: 176–183.

Grassberger, P., Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica* D 9: 189–208.

Hall, P., Marron, J. S., Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Statist. Soc.* B 67: 427–444.

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

Levina, E., Bickel, P. J. (2005). Maximum likelihood estimation of intrinsic dimension. In *Advances in NIPS 17* (Eds. L. K. Saul, Y. Weiss, L. Bottou).

Pollard, K.S., Dudoit, S., van der Laan, M.J. (2005). Multiple Testing Procedures: R multtest Package and Applications to Genomics. *In Bioinformatics and Computational Biology Solutions Using R and Bioconductor, R. Gentleman, V. Carey, W. Huber, R. Irizarry, S. Dudoit (Editors). Springer (Statistics for Biology and Health Series)*: 251–272.

Roweis, S., Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* 290: 2323–2326.

Tenenbaum, J. B., de Silva, V., Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290: 2319–2322.

Verveer, P., Duin, R. (1995). An evaluation of intrinsic dimensionality estimators. *IEEE Trans. on PAMI* 17: 81–86.

Wang, X., Marron, J. S. (2008). A scale-based approach to finding effective dimensionality in manifold learning. *Electronic J. Statist.* 2: 127–148.