数理解析研究所講究録
第2047巻 2017年 19-28

高次元固有ベクトルの一致性

筑波大学・数理物質系　　矢田　和善 (Kazuyoshi Yata)
Institute of Mathematics
University of Tsukuba
筑波大学・数理物質系　　青嶋　誠 (Makoto Aoshima)
Institute of Mathematics
University of Tsukuba

1　はじめに

情報化の進展に伴い，高次元データの統計解析が，ますます重要になってきている. 2000 年以降，確率論と理論物理の方面から，ランダム行列の理論に基づく幾つかの重要な結果がもたらされた. Johnstone [7] や Paul [8] 等は，標本固有値の漸近分布を導出した. しかしながら，そこでは，データの次元数 d と標本数 n が $n/d \to c > 0$ を満たす場合を考え，高次元において標本数は次元数と同程度を仮定した. 例えば，次元数は優に 10,000 を超えるが標本数は高々 100 程度といった高次元小標本においては，標本数を次元数と同程度には仮定できない. それゆえ，n が d に依存しないような設定で，もしくは，$n = n(d)$ であっても $n/d \to 0$ となる設定で，高次元漸近理論を展開する必要がある. Yata and Aoshima [10] は，高次元小標本における PCA の性質を研究し，PCA が一致性をもつための標本数 n の d に関するオーダー条件を導き，高次元小標本において PCA が不適解を起こすことを示した. この問題を解決する策として，Yata and Aoshima [12] は，高次元小標本データ空間の幾何学的表現を研究し，それに基づいて “ノイズ掃き出し法” とよばれる方法論を考案した. 一方で，Yata and Aoshima [13] は，高次元大標本も含む一般的な高次元データに対して，power spiked モデルと呼ばれる固有値モデルを考案し，高次元データに対する新しい PCA を構築した. 最近，Aoshima and Yata [5] は，ノイズ掃き出し法による固有ベクトルの推定量を用いることで，新たな高次元二標本検定法を考案した.

本稿では，高次元固有ベクトルの一致性について論じる. ノイズ掃き出し法による固有ベクトルの推定量を補正することで，緩い仮定のもとその一致性を与える新たな

方法論を提案する.

2 高次元固有値の一致性

平均に d 次のゼロベクトル，共分散行列に d 次の半正定値行列 $\boldsymbol{\Sigma}$ をもつ母集団を考える．母集団から $n\ (\geq 2)$ 個の d 次データベクトル $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ を無作為に抽出して，$d \times n$ データ行列 $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]$ を定義する．ただし，$d > n$ と仮定する．$\boldsymbol{\Sigma}$ の固有値を $\lambda_1 \geq \cdots \geq \lambda_d (\geq 0)$ とし，適当な直交行列 $\boldsymbol{H} = [\boldsymbol{h}_1, ..., \boldsymbol{h}_d]$ で $\boldsymbol{\Sigma}$ を

$$\boldsymbol{\Sigma} = \boldsymbol{H\Lambda H}^T, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_d)$$

と分解する．そのとき $\boldsymbol{Z} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{H}^T\boldsymbol{X}$ とおき，$\boldsymbol{Z} = [\boldsymbol{z}_1, ..., \boldsymbol{z}_d]^T$，$\boldsymbol{z}_s = (z_{s1}, ..., z_{sn})^T$ と表記する．ただし，$\boldsymbol{Z}$ の成分は，4 次モーメントに一様有界性を仮定する．標本共分散行列 $\boldsymbol{S} = \boldsymbol{XX}^T/n$ の固有値を $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_d\ (\geq 0)$，$\hat{\lambda}_j$ に対する固有ベクトルを $\hat{\boldsymbol{h}}_j$ として，スペクトル分解を

$$\boldsymbol{S} = \sum_{s=1}^{d} \hat{\lambda}_s \hat{\boldsymbol{h}}_s \hat{\boldsymbol{h}}_s^T$$

とおく．最近，Yata and Aoshima [13] は，power spiked モデルとよばれる固有値モデルを考案し，高次元データに対する新しい PCA を研究した．いま，$\boldsymbol{\Sigma}_{(1)} = \sum_{s=1}^{m} \lambda_s \boldsymbol{h}_s \boldsymbol{h}_s^T$，$\boldsymbol{\Sigma}_{(2)} = \sum_{s=m+1}^{d} \lambda_s \boldsymbol{h}_s \boldsymbol{h}_s^T$ とおき，$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{(1)} + \boldsymbol{\Sigma}_{(2)}$ という分解を考える．ただし，$m < \infty$．そのとき，次の条件を満たすような $\lambda_1 \geq \cdots \geq \lambda_d$ を power spiked モデルと定義する．

λ_m に対して，$\lim_{d\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^{k_m})/\lambda_m^{k_m} = 0$ なる (有界な) ある自然数 k_m が存在する．

本稿では簡単のため，$k_m = 2$ の場合を考える．すなわち，

$$\lim_{d\to\infty} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)}^2)}{\lambda_j^2} = 0, \quad j = 1, ..., m \tag{1}$$

なる spiked モデルを仮定する．いま，

$$\delta_j = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{(2)})}{n\lambda_j}, \quad j = 1, ..., m$$

とおく．(1) のもと，次の定理を得る．

定理 1 ([13]). 各 $j=1,...,m$ について，条件

(C-i) $$\frac{\sum_{s,t=m+1}^{d}\lambda_s\lambda_t E\{(z_{sk}^2-1)(z_{tk}^2-1)\}}{n\lambda_j^2}=o(1)$$

のもと，$d,n\to\infty$ のとき次が成り立つ.

$$\frac{\hat{\lambda}_j}{\lambda_j}=1+\delta_j+o_p(1).$$

注意. もし，$z_{1k},...,z_{dk}$ が互いに独立ならば，(1) のもと (C-i) を満たす.

定理 1 より，$\delta_j\to\infty$，$d,n\to\infty$ のとき，$\hat{\lambda}_j$ は "$\lambda_j/\hat{\lambda}_j=o_p(1)$" なる強不一致性をもつ. 一方で，Yata and Aoshima [12] は，高次元小標本データ空間の幾何学的表現を研究し，それに基づいて "ノイズ掃き出し法" とよばれる方法論を考案し，次のような固有値の推定量を提案した.

$$\tilde{\lambda}_j=\hat{\lambda}_j-\frac{\mathrm{tr}(\boldsymbol{S})-\sum_{i=1}^{j}\hat{\lambda}_i}{n-j}\quad(j=1,...,n-1). \tag{2}$$

そのとき，(1) のもと，次の定理を得る.

定理 2 ([13]). 各 $j=1,...,m$ について，(C-i) のもと，$d,n\to\infty$ のとき次が成り立つ.

$$\frac{\tilde{\lambda}_j}{\lambda_j}=1+o_p(1).$$

定理 2 より，$\delta_j\to\infty$，$d,n\to\infty$ の場合においても，$\tilde{\lambda}_j$ は一致性をもつ.

一方で，(C-i) が仮定できない場合，Yata and Aoshima [11] は，母集団分布の仮定を必要としないクロスデータ行列法という方法論を考案した. 詳細は Aoshima and Yata [1] や青嶋・矢田 [2, 3] を参照されたい.

3 高次元固有ベクトルの一致性

$\boldsymbol{\Sigma}$ の固有ベクトルについて，ノイズ掃き出し法による推定を考える．推定量 (2) に基づいて，$\boldsymbol{\Sigma}$ の固有ベクトル $\boldsymbol{h}_j$ を

$$\tilde{\boldsymbol{h}}_j = \sqrt{\frac{\hat{\lambda}_j}{\tilde{\lambda}_j}}\hat{\boldsymbol{h}}_j$$

で推定する．ただし，$\boldsymbol{h}_j$ には符号の自由度があるため，各 j で $\tilde{\boldsymbol{h}}_j^T\boldsymbol{h}_j \geq 0$ を仮定する．ここで，$||\tilde{\boldsymbol{h}}_j||^2 = \hat{\lambda}_j/\tilde{\lambda}_j > 1$ であることに注意する．ただし，$||\cdot||$ はユークリッドノルムを表す．(1) のもと次の定理を得る．

定理 3 ([13])．各 $j = 1, ..., m$ について，(C-i) と条件

(C-ii) $\displaystyle \liminf_{d\to\infty} \frac{\lambda_j}{\lambda_{j'}} > 1 \text{ for } j < j' \leq m$

のもと，$d, n \to \infty$ のとき次が成り立つ．

$$\boldsymbol{h}_j^T\hat{\boldsymbol{h}}_j = (1+\delta_j)^{-1/2} + o_p(1) \text{ and } \boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_j = 1 + o_p(1).$$

それゆえ，$\tilde{\boldsymbol{h}}_j$ は（$||\tilde{\boldsymbol{h}}_j||^2 > 1$ であるが）$\boldsymbol{h}_j$ の内積に関する一致性をもつ．ノイズ掃き出し法による推定量 $\tilde{\boldsymbol{h}}_j$ は，例えば，Aoshima and Yata [4, 5] では高次元二標本検定と高次元判別分析に応用され，Ishii et al. [6] では高次元共分散行列の同等性検定に応用されている．

ここで，定理 1 から 3 より，(C-i) と (C-ii) のもと，$d, n \to \infty$ のとき次を得る．

$$||\hat{\boldsymbol{h}}_j - \boldsymbol{h}_j||^2 = 2\{1-(1+\delta_j)^{-1/2}\} + o_p(1) \text{ and } ||\tilde{\boldsymbol{h}}_j - \boldsymbol{h}_j||^2 = \delta_j + o_p(1).$$

すなわち，$\liminf_{d,n\to\infty}\delta_j > 0$ のとき，$\hat{\boldsymbol{h}}_j$ と $\tilde{\boldsymbol{h}}_j$ はノルムに関する一致性をもたない．

4 高次元固有ベクトルのスパース推定

本節では，ノルムに関する一致性をもつように，$\tilde{\boldsymbol{h}}_1$ を補正する．いま，$\boldsymbol{h}_1 = (h_1, ..., h_d)^T$ とおく．さらに，

$$D = \{s \mid h_s \neq 0,\ s = 1, ..., d\}$$

とおく．そのとき，次のモデルを仮定する．

$$\liminf_{d,n\to\infty} n^{1/2}|h_s| > 0 \quad \text{for all } s \in D. \tag{3}$$

ここで，$(\sum_{s=2}^{d} \boldsymbol{h}_s \boldsymbol{h}_s^T)\boldsymbol{x}_j = (y_{j1},...,y_{jd})^T$ とおく．そのとき，各 y_{js} の 4 次モーメントが有界であることを仮定する．いま，$\tilde{\boldsymbol{h}}_1 = (\tilde{h}_1,...,\tilde{h}_d)^T$ とおき，$\tilde{h}_1,...,\tilde{h}_d$ を絶対値の大きい順に並べ替えたものを $\tilde{h}_{(1)},...,\tilde{h}_{(d)}$ とおく．すなわち，$|\tilde{h}_{(1)}| \ge \cdots \ge |\tilde{h}_{(d)}|$ となる．$||\tilde{\boldsymbol{h}}_1||^2 > 1$ であることに注意すれば，

$$\sum_{s=1}^{k-1} \tilde{h}_{(s)}^2 < 1 \quad \text{and} \quad \sum_{s=1}^{k} \tilde{h}_{(s)}^2 \ge 1$$

となる k が一意に定まる．そのとき，$\tilde{\boldsymbol{h}}_1$ を次のようにスパース化する．

$$\acute{\boldsymbol{h}}_1 = (\acute{h}_1,...,\acute{h}_d)^T.$$

ただし，

$$\acute{h}_s = \begin{cases} \tilde{h}_s & (|\tilde{h}_s| \ge |\tilde{h}_{(k)}|) \\ 0 & (|\tilde{h}_s| < |\tilde{h}_{(k)}|) \end{cases}, \quad (s = 1,...,d)$$

とする．そのとき，(1) と (3) のもと次の定理を得る．

定理 4. $j = 1$ に対して (C-i) と (C-ii) を仮定する．条件

(C-iii) $d/\lambda_1^2 \to 0,\ d \to \infty$

のもと，$d, n \to \infty$ のとき次が成り立つ．

$$||\acute{\boldsymbol{h}}_1 - \boldsymbol{h}_1||^2 = o_p(1). \tag{4}$$

注意. $\boldsymbol{h}_j$ $(j \ge 2)$ についてもノルムに関する一致性を証明できるが，本稿では割愛する．Shen et al. [9] はあるチューニングパラメータを用いた $\boldsymbol{h}_1$ のスパースな推定量を与え，その高次元一致性を議論した．しかしながら，その推定量がチューニングパラメータに大きく依存することに注意する．一方で，$\acute{\boldsymbol{h}}_1$ はそのようなパラメータに依存せず，自動的にスパースな推定量を与えることができる．

5 シミュレーション

本節では，高次元小標本のもとで，$\acute{\boldsymbol{h}}_1$ と $\hat{\boldsymbol{h}}_1$ の精度を数値的に検証する．母集団分布には，d 次元正規分布 $N_d(\boldsymbol{0},\boldsymbol{\Sigma})$ を考え，次の 2 つの設定を考える．

(a) $d=2^s$，$s=6,...,11$，$n=\lceil d^{1/3}\rceil$ とおく．ただし，$\lceil x\rceil$ は x 以上の最小の整数を表す．$\boldsymbol{\Sigma}=\mathrm{diag}(d^{2/3},1,...,1)$ とおく．すなわち，$\boldsymbol{h}_1=(1,0,...,0)$ である．

(b) $d=500, n=2^s$，$s=3,...,8$ とおく．

$$\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Gamma}_{\lceil d^{2/3}\rceil} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{I}_{d-\lceil d^{2/3}\rceil}\end{pmatrix}\quad(=\boldsymbol{\Sigma}_b)$$

とおく．ただし，$\boldsymbol{\Gamma}_t=\boldsymbol{I}_t+\boldsymbol{1}_t\boldsymbol{1}_t^T$ であり，$\boldsymbol{I}_t$ は t 次の単位行列である．このとき，$\lambda_1\approx d^{2/3}$ であり，最初の $\lceil d^{2/3}\rceil$ 個の成分が非ゼロである $\boldsymbol{h}_1=(\lceil d^{2/3}\rceil^{-1/2},...,\lceil d^{2/3}\rceil^{-1/2},0,...,0)^T$ となる．

設定 (a) と (b) において，A:$||\acute{\boldsymbol{h}}_1-\boldsymbol{h}_1||^2$ と B:$||\hat{\boldsymbol{h}}_1-\boldsymbol{h}_1||^2$ をそれぞれ 1000 回発生させ，その平均を図 1 と図 2 にプロットした．

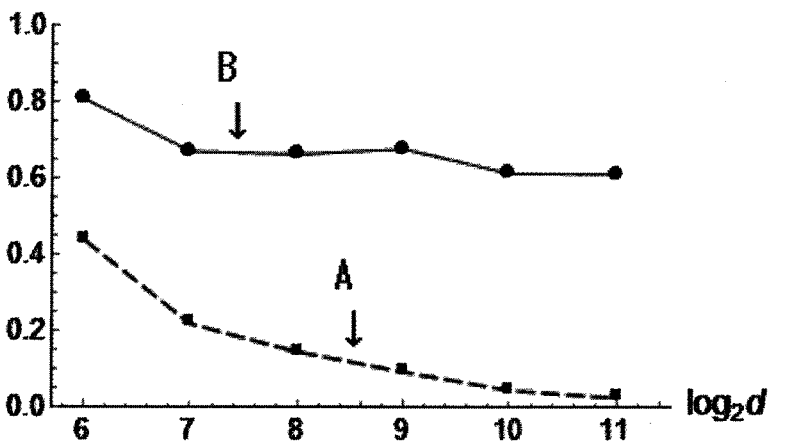

図 1　(a) $d=2^s$，$s=6,...,11$，$n=\lceil d^{1/3}\rceil$，$\boldsymbol{\Sigma}=\mathrm{diag}(d^{2/3},1,...,1)$.

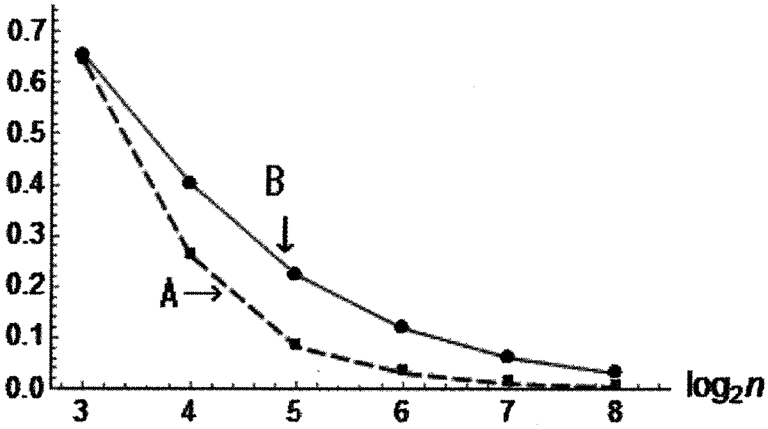

図 2　(b) $d = 500, n = 2^s,\ s = 3, ..., 8,\ \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_b$.

これらの図からも分かるように，スパースな推定量 $\acute{\boldsymbol{h}}_1$ が従来の推定量 $\hat{\boldsymbol{h}}_1$ に比べ，高次元小標本のもと非常に良い推定量となっている．ここでは割愛するが，設定を変えて実験をしたときにも，$\acute{\boldsymbol{h}}_1$ が $\hat{\boldsymbol{h}}_1$ に比べ，高次元小標本のもと優れた推定量となっていることが確認されている．

6　定理 4 の証明

まず，$j = 1$ に対して条件 (C-i), (C-ii) と (C-iii) を仮定する．いま，$\boldsymbol{S}_D = n^{-1}\boldsymbol{X}^T\boldsymbol{X}$ とおく．$\boldsymbol{S}_D$ は $\boldsymbol{S}$ と正の固有値を共有する双対標本共分散行列という．$\boldsymbol{S}_D$ の固有値を $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$ とし，$\hat{\lambda}_j$ に対する固有ベクトルを $\hat{\boldsymbol{u}}_j$ として，スペクトル分解を

$$\boldsymbol{S}_D = \sum_{s=1}^{n} \hat{\lambda}_s \hat{\boldsymbol{u}}_s \hat{\boldsymbol{u}}_s^T$$

とおく．そのとき，

$$\tilde{\boldsymbol{h}}_j = (n\tilde{\lambda}_j)^{-1/2}\boldsymbol{X}\hat{\boldsymbol{u}}_j$$

と書ける．ここで，Yata and Aoshima [13] の補題 9 より，$d, n \to \infty$ のとき次が成り立つ．

$$\hat{\boldsymbol{u}}_1^T \frac{\boldsymbol{z}_1/n^{1/2}}{||\boldsymbol{z}_1/n^{1/2}||} = 1 + o_p(n^{-1/2}).$$

すなわち,

$$\hat{\boldsymbol{u}}_1 = \{1+o_p(1)\}\boldsymbol{z}_1/n^{1/2} + \boldsymbol{\omega}$$

と書ける．ただし，$\boldsymbol{z}_1^T\boldsymbol{\omega}=0$, $||\boldsymbol{\omega}||^2 = o_p(n^{-1/2})$ である．いま，$\boldsymbol{y}_{(s)} = (y_{1s},...,y_{ns})^T$, $s=1,...,d$ とおく．そのとき，定理 2 より次が成り立つ．

$$\begin{aligned}\tilde{\boldsymbol{h}}_1 =& \{1+o_p(1)\}\boldsymbol{h}_1 + \frac{\{1+o_p(1)\}}{n\lambda_1^{1/2}}(\boldsymbol{y}_{(1)}^T\boldsymbol{z}_1,...,\boldsymbol{y}_{(d)}^T\boldsymbol{z}_1)^T \\ &+ \frac{\{1+o_p(1)\}}{n^{1/2}\lambda_1^{1/2}}(\boldsymbol{y}_{(1)}^T\boldsymbol{\omega},...,\boldsymbol{y}_{(d)}^T\boldsymbol{\omega})^T. \end{aligned} \tag{5}$$

ここで，マルコフの不等式を用いると，任意の $\tau>0$ について次が成り立つ．

$$\begin{aligned}\sum_{s=1}^d P\Big(|\boldsymbol{y}_{(s)}^T\boldsymbol{z}_1|/(n\lambda_1^{1/2}) > \tau n^{-1/2}\Big) &= \sum_{s=1}^d P\Big(|\boldsymbol{y}_{(s)}^T\boldsymbol{z}_1|^4/(n\lambda_1)^2 > \tau^4\Big) \\ &= O(d/\lambda_1^2) \to 0. \end{aligned} \tag{6}$$

さらに，各 s で $\sigma_{(s)} = E(y_{js}^2)$ とおき，任意の $\tau>0$ について次が成り立つ．

$$\begin{aligned}&\sum_{s=1}^d P\Big(|(||\boldsymbol{y}_{(s)}||^2/n - \sigma_{(s)})|/\lambda_1) > \tau n^{-1/2}\Big) \\ &= \sum_{s=1}^d P\Big(|(||\boldsymbol{y}_{(s)}||^2/n - \sigma_{(s)})|^2/\lambda_1^2) > \tau^2 n^{-1}\Big) = O(d/\lambda_1^2) \to 0.\end{aligned}$$

それゆえ，$d>n$ に注意し，すべての s について次が成り立つ．

$$\frac{||\boldsymbol{y}_{(s)}||^2}{n\lambda_1} = \frac{\sigma_{(s)}}{\lambda_1} + o_p(n^{-1/2}) = o_p(n^{-1/2}). \tag{7}$$

それゆえ，(5) から (7) より，次が成り立つ．

$$\tilde{\boldsymbol{h}}_1 = \{1+o_p(1)\}\boldsymbol{h}_1 + \Big(o_p(n^{-1/2}),...,o_p(n^{-1/2})\Big)^T.$$

よって，仮定 (3) より題意を得る． □

謝辞　本研究は，科学研究費補助金 基盤研究 (A) 15H01678 研究代表者: 青嶋 誠「大規模複雑データの理論と方法論の総合的研究」，および，若手研究 (B) 26800078 研究代表者: 矢田 和善「高次元漸近理論の統一的研究」から研究助成を受けています．

参考文献

[1] Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data, *Sequential Analysis (Editor's special invited paper)*, **30**, 356-399.

[2] 青嶋　誠, 矢田和善 (2013a). 論説：高次元小標本における統計的推測, 数学, **65**, 225-247.

[3] 青嶋　誠, 矢田和善 (2013b). 日本統計学会研究業績賞受賞者特別寄稿論文：高次元データの統計的方法論, 日本統計学会誌, **43**, 123-150.

[4] Aoshima, M. and Yata, K. (2016). A distance-based classifier for high-dimensional data under the strongly spiked eigenvalue model, submitted.

[5] Aoshima, M. and Yata, K. (2017). Two-sample tests for high-dimension, strongly spiked eigenvalue models, *Statistica Sinica*, in press (arXiv:1602.02491).

[6] Ishii, A., Yata, K. and 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.

[7] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *The Annals of Statistics*, **29**, 295-327.

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

[9] Shen, D., Shen, H. and Marron, J.S. (2013). Consistency of sparse PCA in high dimension, low sample size contexts, *Journal of Multivariate Analysis*, **115** 317-333.

[10] Yata, K. and Aoshima, M. (2009). PCA consistency for non-Gaussian data in high dimension, low sample size context, *Communications in Statistics. Theory and Methods, Special Issue: Honoring Zacks, S. (ed. Mukhopadhyay, N.)*, **38**, 2634-2652.

[11] Yata, K. and 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.

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

[13] Yata, K. and Aoshima, M. (2013). PCA consistency for the power spiked model in high-dimensional settings, *Journal of Multivariate Analysis*, **122**, 334-354.