Biometrika (2013), **xx**, x, *pp*. 1–9

Printed in Great Britain

Supplementary material for "Convergence of Sample Eigenvalues, Eigenvectors, and Principal Component Scores for Ultra-High Dimensional Data"

BY SEUNGGEUN LEE

Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, Michigan 48109, U.S.A.

leeshawn@umich.edu

FEI ZOU

Department of Biostatistics, University of North Carolina, 135 Dauer Drive, Chapel Hill, North Carolina 27599, U.S.A.

fzou@bios.unc.edu

FRED A. WRIGHT

Bioinformatics Research Center, North Carolina State University, 1 Lampe Drive, Raleigh, North Carolina 27607, U.S.A.

fred_wright@ncsu.edu

1. SPHERICITY CONDITIONS OF NON-SPIKED EIGENVALUES

Since eigenvalues can be rescaled by a constant, we assume that $\bar{\lambda} = 1$. Condition 1 is closely related to the sphericity measure in John (1971, 1972) and the ϵ_m condition of Jung & Marron (2009). Under the high dimension low sample size regime, Jung & Marron (2009) defined

$$\epsilon_m = \left(\sum_{v=m+1}^{p} \lambda_v \right)^2 \left(p \sum_{v=m+1}^{p} \lambda_v^2 \right)^{-1},$$

and derived their asymptotic results under the condition where $\epsilon_m^{-1} = o(p)$. Since under Condition 1,

$$\epsilon_m^{-1} = \left(p \sum_{v=m+1}^{p} \lambda_v^2 \right) \left(\sum_{v=m+1}^{p} \lambda_v \right)^{-2} = (p-m)^{-2} \sum_{v=m+1}^{p} p(\lambda_v - 1)^2 + (p-m)^{-1} p$$
$$= o(p) \quad \text{for fixed } n,$$

Condition 1 implies the ϵ_m condition. The relative growth rate of n to p plays a critical role in both conditions. For example, when $(p-m)^{-1} \sum_{v=m+1}^{p} (\lambda_v - 1)^2 = O(1)$, p must grow faster than n^2 to satisfy Condition 1. In Condition 2, we relax the assumption on $(\lambda_v - 1)^2$ but add an additional assumption on $(\lambda_v - 1)^4$.

2. NOTATIONS

Since $E_p\Lambda^{1/2}ZZ^T\Lambda^{1/2}E_p^T$ has the same eigenvalues as $\Lambda^{1/2}ZZ^T\Lambda^{1/2}$, with E_p^TU being the corresponding eigenvector matrix, without loss of generality we assume that $\sigma^2\Lambda$ is the population covariance matrix, and e_v is the vth eigenvector. Let $\varphi_v(\cdot)$ be a function on a matrix that returns its vth largest eigenvalue. Suppose suffixes A and B represent the first m and the remaining coordinates of any given matrix, respectively. The sample covariance matrix S can then be partitioned as

$$S = \begin{pmatrix} S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{pmatrix},$$

and u_v^T and Z^T can be partitioned as $(u_{A,v}^T, u_{B,v}^T)$ and $\left(Z_A^T, Z_B^T\right)$, respectively. Let $\Lambda_A = \text{diag}(\lambda_1, \ldots, \lambda_m)$, $\Lambda_B = \text{diag}(\lambda_{m+1}, \ldots, \lambda_p)$, $R_v = ||u_{B,v}||$, and $a_v = (1 - R_v^2)^{-1/2}u_{A,v}$. By the singular value decomposition, $\Lambda_B^{1/2}Z_B = n^{-1/2}VM^{1/2}H^T$, where $M = \text{diag}(\mu_1, \ldots, \mu_{p-m})$ is a $(p-m)\times(p-m)$ diagonal matrix of the ordered eigenvalues of S_{BB}, V is a $(p-m)\times(p-m)$ orthogonal matrix, and H is an $n\times(p-m)$ matrix. When $n \geq p-m$, H is column-orthogonal. When $n < p-m$, the first n columns of H are orthogonal while the rest of the columns are zero. Let $f_v = \lambda_{v+m} - 1$, we then have

$$\begin{aligned}
&\sum_{i\neq j\neq k\neq l} f_if_jf_kf_l = 3\sum_{i\neq j} f_i^2f_j^2 - 3\sum_i f_i^4, \quad \sum_{i\neq j\neq k} f_i^2f_jf_k = -\sum_{i\neq j} f_i^2f_j^2 + \sum_i f_i^4,\\
&\sum_{i\neq j} f_i^3f_j = -\sum_i f_i^4, \quad \sum_{i\neq j\neq k} f_if_jf_k = 2\sum_i f_i^3, \quad \sum_{i\neq j} f_i^2f_j = -\sum_i f_i^3,\\
&\sum_{i\neq j} f_if_j = -\sum_i f_i^2. \qquad (1)
\end{aligned}$$

3. PROOF OF THEOREM 1

3·1. *Lemma 1*

We introduce the following lemma for future use.

LEMMA 1. *Suppose $\gamma_p \to \infty$ and $n \to \infty$ as $p \to \infty$. With either Condition 1 or 2,*

$$p^{-1}||Z_B^TZ_B - Z_B^T\Lambda_BZ_B|| \to 0 \textit{ in probability.}$$

Proof. *Step* 1. We first truncate and centralize z_{ij}, as described in Yin et al. (1988) and Bai & Yin (1993). Define a $(p-m)\times n$ matrix $\tilde{Z}$ with the (i,j)th element

$$\tilde{z}_{ij} = z_{ij}^BI(|z_{ij}^B| \leq \sigma\delta_pn^{1/4}p^{1/4}) - E\left\{z_{ij}^BI(|z_{ij}^B| \leq \sigma\delta_pn^{1/4}p^{1/4})\right\},$$

where z_{ij}^B is the (i,j)th element of Z_B. By the truncation and centralization lemma (Yin et al., 1988), there exists a sequence δ_p that converges to zero and satisfies

$$\begin{aligned}
&E(\tilde{z}_{ij}) = 0, \quad E(\tilde{z}_{ij}^2) \to \sigma^2, \quad E(\tilde{z}_{ij}^4) = O(1), \quad E(\tilde{z}_{ij}^k) = (\delta_pn^{1/4}p^{1/4})^{k-4}O(1)\ (k > 4),\\
&\varphi_1(p^{-1}\tilde{Z}^T\tilde{Z}) - \varphi_1(p^{-1}Z_B^TZ_B) = o(1), \quad \varphi_n(p^{-1}\tilde{Z}^T\tilde{Z}) - \varphi_n(p^{-1}Z_B^TZ_B) = o(1). \qquad (2)
\end{aligned}$$

Since both $\varphi_1(p^{-1}Z_B^TZ_B)$ and $\varphi_n(p^{-1}Z_B^TZ_B)$ converge to unity, $p^{-1}||\tilde{Z}^T\tilde{Z} - Z_B^TZ_B|| = o(1)$.

Step 2. We largely follow the combinatorial argument described in Bai & Silverstein (2010). Suppose there are two sets of integers, $\xi = (\xi_1, \ldots, \xi_{k+1})$ with $\xi_1 = \xi_{k+1}$ and $\zeta = (\zeta_1, \ldots, \zeta_k)$, where the ξ_us range from 1 to n and the ζ_us range from 1 to $p-m$. A graph $G(\xi,\zeta)$ is constructed by connecting k edges from ξ_u to ζ_u and another k edges from ζ_u to ξ_{u+1}. See Figure 3.1 of Bai & Silverstein (2010). An edge is single if there are no other edges to share the same two vertices. For the graph $G(\xi,\zeta)$, let $\tilde{Z}_{G(\xi,\zeta)} = \prod_{u=1}^{k} \tilde{z}_{\xi_u,\zeta_u} \tilde{z}_{\xi_{u+1},\zeta_u}$. Two graphs are isomorphic if there exist a permutation of ξ and a permutation of ζ such that the graphs are identical. The canonical graph for each isomorphic group is defined as $\xi_1 = \zeta_1 = 1, \xi_u \le \max(\xi_1,\ldots,\xi_{u-1})+1$, and $\zeta_u \le \max(\zeta_1,\cdots,\zeta_{u-1})+1$. Let $r+1$ and s be the number of distinct elements in ξ and ζ, respectively. Let $A_n = p^{-1}\tilde{Z}^T(\Lambda_B - I)\tilde{Z}$ and $\Delta(k,r,s)$ be a set of canonical graphs with no single edge. It can be shown that

$$E\left\{\mathrm{tr}(A_n^k)\right\} = p^{-k}\sum_{r=0}^{k-1}\prod_{n_1=n-r}^{n} n_1 \sum_{s=1}^{k-r}\sum_{j_1\neq\cdots\neq j_s}\sum_{m_{s,1},\ldots,m_{s,s}}\prod_{i=1}^{s} f_{j_i}^{m_{s,i}} \sum_{G\in\Delta(k,r,s)} E(\tilde{Z}_G),$$

where $m_{s,1}\ge\cdots\ge m_{s,s}\ge 1$ and $\sum_{i=1}^{s} m_{s,i} = k$.

Step 3. Since $E(\tilde{Z}_G) = O(1)$ for all $G\in\Delta(2,r,s)$,

$$\begin{aligned} E\left\{\mathrm{tr}(A_n^2)\right\} &= p^{-2}n\sum_{j_1\neq j_2} f_{j_1}f_{j_2}O(1) + p^{-2}n\sum_{j_1} f_{j_1}^2 O(1) + p^{-2}n(n-1)\sum_{j_1} f_{j_1}^2 O(1) \\ &= p^{-2}n(n-1)\sum_{j_1} f_{j_1}^2 O(1) = o(1), \end{aligned} \tag{3}$$

under Condition 1. Applying Markov's inequality, for any $\epsilon > 0$,

$$\mathrm{pr}\left\{\max_{v=1,\ldots,n}\varphi_v(A_n^2) > \epsilon\right\} \le \mathrm{pr}\left\{\mathrm{tr}(A_n^2) > \epsilon\right\} \le \epsilon^{-1}E\left\{\mathrm{tr}(A_n^2)\right\} = o(1),$$

indicating that $||A_n|| = o_p(1)$. Thus, Lemma 1 holds under Condition 1. Using the same approach in (3) with (1)–(2), it can be shown that

$$E\left\{\mathrm{tr}(A_n^4)\right\} = (n^3 + n^{3/2}p^{1/2})p^{-4}\sum_{j_1\neq j_1} f_{j_1}^2 f_{j_2}^2 O(1) + (n^4 + n^2 p)p^{-4}\sum_{j_1} f_{j_1}^4 O(1),$$

which converges to zero under Condition 2. Therefore, $||A_n|| = o_p(1)$, which concludes the proof.

3·2. *Convergence of sample eigenvalues*

We first assume that c_v $(v\le m)$ is bounded away from zero. The vth eigenvalue of S is $d_v = \varphi_v(n^{-1}Z_A^T\Lambda_A Z_A + n^{-1}Z_B^T\Lambda_B Z_B)$. Since $||\lambda_v^{-1}\Lambda_A|| = O(1)$ $(v\le m)$, and $n^{-1}Z_A Z_A^T - \sigma^2 I = o(1)$, we can obtain

$$\left\|(n\lambda_v)^{-1}\Lambda_A^{1/2}Z_A Z_A^T\Lambda_A^{1/2} - \lambda_v^{-1}\sigma^2\Lambda_A\right\| = o(1).$$

By the continuity of eigenvalues and Lemma 1, $\lambda_v^{-1}\varphi_v(n^{-1}Z_A^T\Lambda_A Z_A) - \sigma^2 = o(1)$ $(v\le m)$, and $\gamma_p^{-1}\varphi_v(n^{-1}Z_B^T\Lambda_B Z_B) - \sigma^2 = o_p(1)$ $(v > m)$. For $v > m$, $\varphi_v(n^{-1}Z_A^T\Lambda_A Z_A) = 0$, since $Z_A^T\Lambda_A Z_A$ has rank m. Further, by Weyl's inequality (Bhatia, 1997), $\varphi_v(S) \le \varphi_v(n^{-1}Z_A^T\Lambda_A Z_A) + \varphi_1(n^{-1}Z_B^T Z_B)$ and $\varphi_v(S) \ge \varphi_v(n^{-1}Z_A^T\Lambda_A Z_A) + \varphi_n(n^{-1}Z_B^T Z_B)$.

Hence,

$$d_v/\lambda_v - \sigma^2 c_v^{-1}(c_v+1) = o_p(1) \quad (v \le m),$$
$$d_v/\gamma_p - \sigma^2 = o_p(1) \quad (v > m).$$

When $c_v = o(1)$, $\gamma_p^{-1}\lambda_1(n^{-1}Z_A^T\Lambda_A Z_A) \to 0$. Hence, $d_v\gamma_p^{-1} = \sigma^2 + o_p(1)$, which concludes the proof.

3·3. *Convergence of sample eigenvectors*

The proof largely follows the arguments in Paul (2007) and Lee et al. (2010). We assume without loss of generality that σ^2 is unity. From the definitions of eigenvalues and eigenvectors,

$$\begin{aligned}&\left\{S_{AA} + S_{AB}(d_vI - S_{BB})^{-1}S_{BA}\right\} a_v \\ &= \left\{S_{AA} + n^{-1}\Lambda_A^{1/2}Z_AHM(d_vI-M)^{-1}H^TZ_A^T\Lambda_A^{1/2}\right\} a_v = d_v a_v,\end{aligned} \tag{4}$$

and

$$a_v^T\left\{I + n^{-1}\Lambda_A^{1/2}Z_AHM(d_vI-M)^{-2}H^TZ_A^T\Lambda_A^{1/2}\right\} a_v = (1-R_v^2)^{-1}. \tag{5}$$

We first assume that c_v $(v \le m)$ is bounded away from zero. Let

$$\eta_v = c_v^{-1}(c_v+1), \quad \mathcal{R}_v = \sum_{k\neq v}\left\{\eta_v(\lambda_k - \lambda_v)\right\}^{-1}\lambda_v e_{A,k}e_{A,k}^T,$$

$$\mathcal{D}_v = \lambda_v^{-1}\left\{S_{AA} + S_{AB}(d_vI - S_{BB})^{-1}S_{BA} - \eta_v\Lambda_A\right\},$$

where $e_{A,v}$ contains the first m elements of e_v. From (4), it can be shown that

$$\left(\lambda_v^{-1}\eta_v\Lambda_A - \eta_v I\right)a_v = -\mathcal{D}_v a_v + \left(\lambda_v^{-1}d_v - \eta_v\right)a_v.$$

Since $\mathcal{R}_v\left(\lambda_v^{-1}\eta_v\Lambda_A - \eta_vI\right) = I - e_{A,v}e_{A,v}^T$,

$$(I - e_{A,v}e_{A,v}^T)a_v = -\mathcal{R}_v\mathcal{D}_va_v + \left(\lambda_v^{-1}d_v - \eta_v\right)\mathcal{R}_va_v,$$

which indicates that $a_v - e_v = o_p(1)$ if both $||\mathcal{R}_v\mathcal{D}_v||$ and $|\lambda_v^{-1}d_v - \eta_v|||\mathcal{R}_v||$ are $o_p(1)$. For $k = 1,\ldots,m$ and $l = 1,\ldots,m$, we show that

$$\begin{aligned}e_{A,k}^T\mathcal{D}_ve_{A,l} &= \lambda_v^{-1}e_{A,k}^TS_{AA}e_{A,l} \\ &\quad + \lambda_v^{-1}e_{A,k}^TS_{AB}(d_vI - S_{BB})^{-1}S_{BA}e_{A,l} - \lambda_v^{-1}\eta_ve_{A,k}^T\Lambda_Ae_{A,l}.\end{aligned} \tag{6}$$

The first term of (6) is

$$\lambda_v^{-1}e_{A,k}^TS_{AA}e_{A,l} = (n\lambda_v)^{-1}(\lambda_k\lambda_l)^{1/2}Z_{A,k}^TZ_{A,l} = \begin{cases}\lambda_v^{-1}(\lambda_k\lambda_l)^{1/2}o_p(1), & k \neq l, \\ \lambda_v^{-1}\lambda_k\left\{1+o_p(1)\right\}, & k = l,\end{cases} \tag{7}$$

and the third term of (6) equals

$$\begin{cases}0, & k \neq l, \\ \lambda_v^{-1}\eta_v\lambda_k, & k = l.\end{cases} \tag{8}$$

From Proposition 1 of Lee et al. (2010),

$$n^{-1}Z_{A,k}^THM(d_vI-M)^{-1}H^TZ_{A,l} = \begin{cases}o_p(1), & l \neq k, \\ n^{-1}\mathrm{tr}\left\{M(d_vI-M)^{-1}\right\} + o_p(1), & l = k.\end{cases}$$

Although Proposition 1 of Lee et al. (2010) requires that $||HM(d_vI - M)^{-1}H^T|| = O(1)$, it can easily be shown that the same result holds for $||HM(d_vI - M)^{-1}H^T|| = O(1) + o_p(1)$. Since $\mu_n(d_v - \mu_n)^{-1} < n^{-1}\text{tr}\left\{M(d_vI - M)^{-1}\right\} < \mu_1(d_v - \mu_1)^{-1}$, and both $\mu_1\gamma_p^{-1}$ and $\mu_n\gamma_p^{-1}$ are $1 + o_p(1)$,

$$n^{-1}\text{tr}\left\{M(d_vI - M)^{-1}\right\} = \{1 + o_p(1)\}\{c_v + c_v o_p(1)\}^{-1} + o_p(1). \tag{9}$$

Hence, the second term of (6) equals

$$\begin{cases} \lambda_v^{-1}(\lambda_k\lambda_l)^{1/2}o_p(1), & k \neq l, \\ \lambda_v^{-1}\lambda_k\left[\{1 + o_p(1)\}\{c_v + c_v o_p(1)\}^{-1} + o_p(1)\right], & k = l. \end{cases} \tag{10}$$

Summing up (7), (8), and (10), we get

$$e_{A,k}^T\mathcal{D}_v e_{A,l} = \lambda_v^{-1}(\lambda_k\lambda_l)^{1/2}o_p(1) \quad (k = 1, \ldots, m; l = 1, \ldots, m).$$

Decomposing $\mathcal{R}_v\mathcal{D}_v e_{A,l} = \sum_{k\neq v}\lambda_v\{\eta_v(\lambda_k - \lambda_v)\}^{-1}e_{A,k}e_{A,k}^T\mathcal{D}_v e_{A,l}$, we conclude that $||\mathcal{R}_v\mathcal{D}_v|| = o_p(1)$. Since $||\mathcal{R}_v|| = O(1)$ and $\left|d_v\lambda_v^{-1} - \eta_v\right|||\mathcal{R}_v|| = o_p(1)$,

$$a_v - e_v = o_p(1). \tag{11}$$

From the same argument of (9),

$$n^{-1}\lambda_v Z_{Av}^T HM(d_vI - M)^{-2}H^T Z_{Av} = \frac{1 + o_p(1)}{c_v + c_v o_p(1)} + o_p(1) = c_v^{-1} + o_p(1).$$

Using (11),

$$\begin{aligned}(5) &= \{e_v + o_p(1)\}^T\left\{I + n^{-1}\Lambda_A^{1/2}Z_A HM(d_vI - M)^{-2}H^T Z_A^T\Lambda_A^{1/2}\right\}\{e_v + o_p(1)\} \\ &= 1 + c_v^{-1} + o_p(1). \end{aligned} \tag{12}$$

By combining (11) and (12),

$$\langle e_v, u_v\rangle = \langle e_{A,v}, a_v\rangle(1 - R_v^2)^{1/2} = c_v^{1/2}(1 + c_v)^{-1/2} + o_p(1), \tag{13}$$

which completes the proof for the case where c_v is bounded away from zero.

Now we assume $c_v = o(1)$. Without loss of generality, we only consider the first eigenvector. Since $e_1^T u_1 < (1 - R_1^2)^{1/2}$, all we need to show is that $R_1 \to 1$. The first sample eigenvalue is

$$d_1 = u_1^T S u_1 = u_{A,1}^T S_{AA} u_{A,1} + 2u_{A,1}^T S_{AB} u_{B,1} + u_{B,1}^T S_{BB} u_{B,1}. \tag{14}$$

The first term of (14) is

$$\begin{aligned} u_{A,1}^T S_{AA} u_{A,1} &= n^{-1}u_{A,1}^T\Lambda_A^{1/2}Z_A Z_A^T\Lambda_A^{1/2}u_{A,1} \\ &\leq (1 - R_1^2)\varphi_1(n^{-1}\Lambda_A^{1/2}Z_A Z_A^T\Lambda_A^{1/2}) = (1 - R_1^2)\lambda_1\{1 + o(1)\}, \end{aligned}$$

and the second term of (14) is

$$\begin{aligned}
u_{A,1}^T S_{AB} u_{B,1} &\le \lambda_1^{1/2} ||u_{A,1}|| \max_{v=1,\ldots,m} \left(n^{-1} z_{A,v}^T Z_B^T \Lambda_B^{1/2} u_{B,1} \right) \\
&\le R_1 \lambda_1^{1/2} (1 - R_1^2)^{1/2} \max_{v=1,\ldots,m} \left(n^{-1} ||\Lambda_B^{1/2} Z_B z_{A,v}|| \right) \\
&\le R_1 \lambda_1^{1/2} (1 - R_1^2)^{1/2} n^{-1} \left(\sum_{v=1}^m z_{A,v}^T Z_B^T \Lambda_B Z_B z_{A,v} \right)^{1/2} \\
&= R_1 \gamma_p \left\{ c_1 (1 - R_1^2) \right\}^{1/2} O_p(1),
\end{aligned}$$

where $z_{A,v}$ is the vth row vector of Z_A, and the second inequality is obtained by the Cauchy-Schwarz inequality. The third term of (14) is $u_{B,1}^T S_{BB} u_{B,1} < R_1^2 \mu_1$. Combining all the three terms, we get $(d_1 - \mu_1)\gamma_p^{-1} \le R_1^2 - 1 + o_p(1)$. Since $d_1 - \mu_1 \ge 0$ by the interlacing inequality (Horn & Johnson, 1990), we conclude that $R_1 \to 1$, which completes the proof.

4. PROOF OF COROLLARY 1

Since $\lambda_v/\gamma_p^\alpha$ converges to $\widetilde{c}_v$ as $p \to \infty$, c_v converges to ∞, $\widetilde{c}_v$, and 0 for $\alpha > 1$, $\alpha = 1$, and $\alpha < 1$, respectively, as $p \to \infty$, the results in Corollary 1 can be easily obtained.

5. PROOF OF THEOREM 2

5·1. *Convergence of sample principal component scores*

Without loss of generality, we assume $\text{corr}(p_v, \widehat{p}_v) \ge 0$. Let $\bar{p}_v$ and $\bar{p}_v^*$ be the averages of elements in p_v and $\widehat{p}_v$, respectively. From $S_{AA} u_{A,v} + S_{AB} u_{B,v} = d_v u_{A,v}$, we obtain

$$\begin{aligned}
\frac{p_v^T \widehat{p}_v}{n(\sigma^2 \lambda_v d_v)^{1/2}} &= \frac{1}{(\sigma^2 \lambda_v d_v)^{1/2}} e_{A,v}^T S_{AA} u_{A,v} + \frac{1}{(\sigma^2 \lambda_v d_v)^{1/2}} e_{A,v}^T S_{AB} u_{B,v} \\
&= \frac{d_v^{1/2}}{\sigma \lambda_v^{1/2}} e_{A,v}^T u_{A,v} = \frac{d_v^{1/2}}{\sigma \lambda_v^{1/2}} e_v^T u_v. \qquad (15)
\end{aligned}$$

Since $e_v^T u_v - \left\{ c_v (c_v + 1)^{-1} \right\}^{1/2} = o_p(1)$ and $\lambda_v^{-1} d_v - \sigma^2 c_v^{-1}(c_v + 1) = o_p(1)$, (15) converges to unity in probability, and $n^{-1/2} J^T \left\{ p_v (\sigma^2 \lambda_v)^{-1/2} - \widehat{p}_v d_v^{-1/2} \right\} = o_p(1)$ where $J = (1, \ldots, 1)^T$. Thus,

$$\begin{aligned}
\frac{\bar{p}_v J^T \widehat{p}_v}{n(\sigma^2 \lambda_v d_v)^{1/2}} &= \frac{\bar{p}_v J^T \left\{ p_v (\sigma^2 \lambda_v)^{-1/2} - p_v (\sigma^2 \lambda_v)^{-1/2} + \widehat{p}_v d_v^{-1/2} \right\}}{n(\sigma^2 \lambda_v)^{1/2}} \\
&= \frac{\bar{p}_v^2}{\sigma^2 \lambda_v} - \frac{\bar{p}_v J^T \left\{ p_v (\sigma^2 \lambda_v)^{-1/2} - \widehat{p}_v d_v^{-1/2} \right\}}{n(\sigma^2 \lambda_v)^{1/2}} = o_p(1). \qquad (16)
\end{aligned}$$

Since $(p_v - \bar{p}_v J)^T (p_v - \bar{p}_v J)/(n\sigma^2 \lambda_v) = 1 + o_p(1)$, we can easily show that $(\widehat{p}_v - \bar{p}_v^* J)^T (\widehat{p}_v - \bar{p}_v^* J)/(n d_v) = 1 + o_p(1)$. Combining (15) and (16), we conclude the proof.

5·2. *Convergence of predicted principal component scores*

Let $u_v^{\perp} = (I - e_v e_v^T) u_v \left\{1 - (u_v^T e_v)^2\right\}^{-1/2}$. Then $u_v = (u_v^T e_v) e_v + \left\{1 - (u_v^T e_v)^2\right\}^{1/2} u_v^{\perp}$. We partition $u_v^{\perp}$ into $\left(u_{A,v}^{\perp}, u_{B,v}^{\perp}\right)$. Following the same argument in Lee et al. (2010),

$$\lambda_v^{-1} E(\widehat{p}_{vj}^2) = \lambda_v^{-1} E(d_v) \to \sigma^2 c_v^{-1}(c_v + 1), \tag{17}$$

and

$$\begin{aligned}\lambda_v^{-1} E(\widehat{q}_v^2 \mid u_v) &= (u_v^T e_v)^2 + \lambda_v^{-1} \left\{1 - (u_v^T e_v)^2\right\} (u_{A,v}^{\perp T} \Lambda_A u_{A,v}^{\perp} + u_{B,v}^{\perp T} u_{B,v}^{\perp}) \\ &\quad + 2\lambda_v^{-1} u_v^T e_v \left\{1 - (u_v^T e_v)^2\right\}^{1/2} e_{A,v} \Lambda_A u_{A,v}^{\perp} \\ &\quad \to \sigma^2 c_v (1 + c_v)^{-1} \text{ in probability.}\end{aligned} \tag{18}$$

The proof follows from (17) and (18).

6. Examples

Example 1. Suppose there are two independent groups of variables, one with p_1 variables and the other with p_2 variables, and the covariance structure is block compound symmetric. That is, the population covariance matrix is

$$\Sigma = \begin{Bmatrix} (1-\rho_1) I_{p_1,p_1} & 0 \\ 0 & (1-\rho_2) I_{p_2,p_2} \end{Bmatrix} + \begin{pmatrix} \rho_1 J_{p_1} J_{p_1}^T & 0 \\ 0 & \rho_2 J_{p_2} J_{p_2}^T \end{pmatrix},$$

where $p = p_1 + p_2$, $I_{p,p}$ is a $p \times p$ identity matrix, and J_p is a $p \times 1$ vector with all elements equal to unity. Define $r_k = (p_k - 1)/(p - 2)$ $(k = 1, 2)$. Suppose the r_ks are bounded away from 0, $(p_1 - 1)\rho_1 \geq (p_2 - 1)\rho_2$, and $(p_1 - 1)\rho_1 \asymp (p_2 - 1)\rho_2$. Then the first two population eigenvalues, after rescaling, equal

$$\lambda_k = \frac{1 + (p_k - 1)\rho_k}{C} \approx \frac{r_k \rho_k}{C} p \quad (k \leq 2),$$

where $C = 1 - r_1\rho_1 - r_2\rho_2$. The non-spiked eigenvalues are

$$\frac{1}{C}(\underbrace{1-\rho_1, \ldots, 1-\rho_1}_{p_1 - 1}, \underbrace{1-\rho_2, \ldots, 1-\rho_2}_{p_2 - 1}).$$

When $\rho_1 \neq \rho_2$, the non-spiked eigenvalues are not identical, but Condition 1 holds if $p \gg n^2$. Now let us consider 3 scenarios for the ρ_ks: large ($\rho_k \gg 1/n$), small ($\rho_k \asymp 1/n$), and very small ($\rho_k \ll 1/n$). From Theorem 1, the first two sample eigenvalues are consistent for large ρ_ks, inconsistent but separable from the bulk for small ρ_ks, and indistinguishable for very small ρ_ks. Similarly, the first two sample eigenvectors are consistent for large ρ_ks, neither consistent nor asymptotically perpendicular to the corresponding population eigenvectors for small ρ_ks, and asymptotically perpendicular for very small ρ_ks.

Example 2. Suppose there is a group of variables with a compound symmetric correlation structure and another group of independent variables. Specifically, the population covariance matrix is

$$\Sigma = \begin{Bmatrix} (1-\rho) I_{p_1,p_1} & 0 \\ 0 & I_{p_2,p_2} \end{Bmatrix} + \begin{pmatrix} \rho J_{p_1} J_{p_1}^T & 0 \\ 0 & 0 \end{pmatrix}.$$

The first population eigenvalue, after rescaling, is

$$\lambda_1 = \frac{1 + (p_1 - 1)\rho}{C} \asymp \frac{\rho p_1}{C},$$

where $C = 1 - r_1\rho$ and $r_1 = (p_1 - 1)/(p - 1)$. The non-spiked eigenvalues are

$$\frac{1}{C}(\underbrace{1, \dots, 1}_{p_2}, \underbrace{1 - \rho, \dots, 1 - \rho}_{p_1 - 1}).$$

We consider three scenarios on the sizes of the groups: large ($p_1 \asymp p$), moderate ($p_1 \asymp p/n$) and small ($p_1 \ll p/n$). From Theorem 1, the first eigenvalue is consistent for large p_1, inconsistent but separable from the bulk for moderate p_1, and indistinguishable for small p_1. The behaviors of sample eigenvectors can also be inferred from Theorem 1.

REFERENCES

BAI, Z. & SILVERSTEIN, J. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer Verlag.

BAI, Z. D. & YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21**, 1275–1294.

BHATIA, R. (1997). *Matrix Analysis*. Springer Verlag.

HORN, R. & JOHNSON, C. (1990). *Matrix Analysis*. Cambridge Univ Pr.

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58**, 123–127.

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59**, 169–173.

JUNG, S. & MARRON, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37**, 4104–4130.

LEE, S., ZOU, F. & WRIGHT, F. (2010). Convergence and prediction of principal component scores in high-dimensional settings. *Ann. Statist.* **38**, 3605–3629.

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17**, 1617–1642.

YIN, Y. Q., BAI, Z. D. & KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78**, 509–521.