The Convergence Of The Empirical Distribution Of Canonical Correlation Coefficients

Yanrong Yang, Guangming Pan *

Abstract

Suppose that \{X_{jk}; j = 1, \cdots, p_1; k = 1, \cdots, n\} are independent and identically distributed (i.i.d) real random variables with \(EX_{11} = 0\) and \(EX_{11}^2 = 1\), and that \{Y_{jk}; j = 1, \cdots, p_2; k = 1, \cdots, n\} are i.i.d real random variables with \(EY_{11} = 0\) and \(EY_{11}^2 = 1\), and that \{X_{jk}, j = 1, \cdots, p_1; k = 1, \cdots, n\} are independent of \{Y_{jk}, j = 1, \cdots, p_2; k = 1, \cdots, n\}. This paper investigates the canonical correlation coefficients \(r_1 \geq r_2 \geq \cdots \geq r_{p_1}\), whose squares \(\lambda_1 = r_1^2, \lambda_2 = r_2^2, \cdots, \lambda_{p_1} = r_{p_1}^2\) are the eigenvalues of the matrix

\[
S_{xy} = A_x^{-1}A_{xy}A_y^{-1}A_{xy}^T,
\]

where

\[
A_x = \frac{1}{n} \sum_{k=1}^{n} x_k x_k^T, \quad A_y = \frac{1}{n} \sum_{k=1}^{n} y_k y_k^T, \quad A_{xy} = \frac{1}{n} \sum_{k=1}^{n} x_k y_k^T,
\]

and

\[
x_k = (X_{1k}, \cdots, X_{p_1k})^T, \quad y_k = (Y_{1k}, \cdots, Y_{p_2k})^T, \quad k = 1, \cdots, n.
\]

When \(p_1 \to \infty, p_2 \to \infty\) and \(n \to \infty\) with \(\frac{p_1}{n} \to c_1, \frac{p_2}{n} \to c_2, c_1, c_2 \in (0, 1)\), it is proved that the empirical distribution of \(r_1, r_2, \cdots, r_{p_1}\) converges, with probability one, to a fixed distribution under the finite second moment condition.

Keywords: Canonical correlation coefficients; Empirical spectral distribution; Random matrix; Stieltjes transform; Lindeberg’s method.

---

*Y. R. Yang and G. M. Pan are with Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371.


\section{Introduction}

Canonical correlation analysis (CCA) deals with the relationship between two random variable sets. Suppose that there are two random variable sets: \( x = \{x_1, \ldots, x_{p_1}\} \) and \( y = \{y_1, \ldots, y_{p_2}\} \), where \( p_1 \leq p_2 \). Assume that there are \( n \) observations for each of the \( p_1 + p_2 \) variables and they are grouped into \( p_1 \times n \) random matrix \( X = (X_{ij})_{p_1 \times n} \) and \( p_2 \times n \) random matrix \( Y = (Y_{ij})_{p_2 \times n} \) respectively. CCA seeks the linear combinations \( a^T x \) and \( c^T y \) that are most highly correlated, that is to maximize

\[
 r = \text{Corr}(a^T x, c^T y) = \frac{a^T \Sigma_{xy} c}{\sqrt{a^T \Sigma_{xx} a} \sqrt{c^T \Sigma_{yy} c}},
\]

where \( \Sigma_{xx}, \Sigma_{yy} \) are population covariance matrices for \( x, y \) respectively; \( \Sigma_{xy} \) is the population covariance matrix between \( x \) and \( y \).

After finding the maximal correlation \( r_1 \) and associated combination vectors \( a_1, c_1 \), CCA considers seeking a second linear combination \( a_2^T x, c_2^T y \) that has the maximal correlation among all linear combinations uncorrelated with \( a_1^T x, c_1^T y \). This procedure can be iterated and successive canonical correlation coefficients \( r_1, \ldots, r_{p_1} \) can be found. Substituting population covariance matrices with sample covariance matrices, \( r_1, \ldots, r_{p_1} \) can be recast as the roots of the determinant equation

\[
 \det(A_{xy}^{-1} A_{xy}^T - r^2 A_x) = 0,
\]

where

\[
 A_x = \frac{1}{n} XX^T, \quad A_y = \frac{1}{n} YY^T, \quad A_{xy} = \frac{1}{n} XY^T.
\]

About this point, one may refer to page 284 of Mardia, Kent and Bibby (1979). The roots of the determinant equation above go under many names, because they figure equally in discriminant analysis, canonical correlation analysis, and invariant tests of linear hypotheses in the multivariate analysis of variance. These are standard techniques in multivariate statistical analysis. Section 4 of Wachter (1980) described how to transform these statistical settings to the determinant equation form. Johnstone (2008) also gave its applications in these aspects in multivariate statistical analysis.

The empirical distribution of the canonical correlation coefficients \( r_1, r_2, \ldots, r_{p_1} \) is defined as

\[
 F(x) = \frac{1}{p_1} \# \{i : r_i \leq x\},
\]

where \( \#\{\cdots\} \) denotes the cardinality of the set \( \{\cdots\} \). When the two variable sets \( x \) and \( y \) are independent and each set consists of i.i.d Gaussian random variables, Wachter (1980) proved that the empirical distribution of \( r_1, r_2, \ldots, r_{p_1} \) converges in probability and obtained an explicit expression for the limit of the empirical distribution when \( p_1, p_2 \) and \( n \) are all approaching infinity. From the determinant equation (1.2), it can be seen that \( \lambda_1 = r_1^2, \lambda_2 = r_2^2, \ldots, \lambda_{p_1} = r_{p_1}^2 \) are eigenvalues of the matrix \( S_{xy} = A_x^{-1} A_{xy} A_y^{-1} A_{xy}^T \). Hence the analysis of the empirical distribution of \( r_1, r_2, \ldots, r_{p_1} \) is equivalent to analyzing
the ESD of the matrix $S_{xy}$. Here for any $p \times p$ matrix $A$ with real eigenvalues $x_1 \leq x_2 \leq \ldots \leq x_p$, its ESD is defined as

$$F^A(x) = \frac{1}{p} \# \{i : x_i \leq x\}. \quad (1.4)$$

The aim of this paper is to prove that the result in Wachter (1980) remains true when the entries of $X$ and $Y$ have finite second moments but not necessarily Gaussian distribution.

**Theorem 1.** Assume that

(a) $X = (X_{ij})_{1 \leq i \leq p_1, 1 \leq j \leq n}$ where $X_{ij}, 1 \leq i \leq p_1, 1 \leq j \leq n$, are i.i.d real random variables with $EX_{11} = 0$ and $E|X_{11}|^2 = 1$.

(b) $Y = (Y_{ij})_{1 \leq i \leq p_2, 1 \leq j \leq n}$ where $Y_{ij}, 1 \leq i \leq p_2, 1 \leq j \leq n$ are i.i.d real random variables with $EY_{11} = 0$ and $E|Y_{11}|^2 = 1$.

(c) $p_1 = p_1(n)$ and $p_2 = p_2(n)$ with $\frac{p_1}{n} \rightarrow c_1$ and $\frac{p_2}{n} \rightarrow c_2$, $c_1, c_2 \in (0, 1)$, as $n \rightarrow \infty$.

(d) $S_{xy} = A_x^{-1}A_yA_y^{-1}A_y^T$ where $A_x = \frac{1}{n}XX^T$, $A_y = \frac{1}{n}YY^T$ and $A_{xy} = \frac{1}{n}XY^T$.

(e) $X$ and $Y$ are independent.

Then as $n \rightarrow \infty$ the empirical distribution of the matrix $r_1, r_2, \ldots, r_{p_1}$ converges almost surely to a fixed distribution function whose density is

$$\rho(r) = ((r - L)(r + L)(H - r)(H + r))^{\frac{1}{2}}/[\pi c_1 r(1 - r)(1 + r)], \quad r \in [L, H], \quad (1.5)$$

where $L = |(c_2 - c_1)^2 - (c_1 - c_1c_2)^2|$ and $H = |(c_2 - c_1)(c_1 - c_1c_2) + (c_1 - c_1c_2)^2|$; and atoms of size $\max(0, 1 - c_2/c_1)$ at zero and size $\max(0, 1 - (1 - c_2)/c_1)$ at unity.

**Remark 1.** The inverse of a matrix, such as $A_x^{-1}$ and $A_y^{-1}$, is the Moore-Penrose pseudoinverse, i.e. in the spectral decomposition of the initial matrix, replace each nonzero eigenvalue by its reciprocal and leave the zero eigenvalues alone. This is because under the finite second moment condition, the matrices $A_x$ and $A_y$ may be not invertible under the classical inverse matrix definition. However, with the additional assumption that $EX_{11}^4 < \infty$ and $EY_{11}^4 < \infty$, we have the conclusion that the smallest eigenvalues of the sample matrices $A_x$ and $A_y$ converge to $(1 - \sqrt{c_1})^2$ and $(1 - \sqrt{c_2})^2$ respectively [Theorem 5.11 of Bai and Silverstein (2009)], which are not zero since $c_1, c_2 \in (0, 1)$. So $A_x$ and $A_y$ are invertible with probability one under the finite fourth moment condition.

As stated previously, it is sufficient to analyze the limiting spectral distribution (LSD) of the matrix $S_{xy}$, where LSD denotes the limit of the empirical spectral distribution as $n \rightarrow \infty$.

The strategy of the proof of Theorem 1 is as follows. Since the matrix $S_{xy}$ is not symmetric, it is difficult to work on it directly. Instead we consider the $n \times n$ symmetric matrix

$$P_yP_xP_y \quad (1.6)$$

where

$$P_x = X^T(XX^T)^{-1}X, \quad P_y = Y^T(YY^T)^{-1}Y.$$
Note that $P_x$ and $P_y$ are projection matrices. It is easy to see that the eigenvalues of the matrix $P_y P_x P_y$ are the same as those of the matrix $S_{xy}$ other than $n - p_1$ zero eigenvalues, i.e.

$$F^{P_y P_x P_y}(x) = \frac{p_1}{n} F^{S_{xy}}(x) + \frac{n-p_1}{n} I_{[0, +\infty)}(x).$$

(1.7)

By (1.7) and the result in Wachter (1980), one can easily obtain the limit of $F^{P_y P_x P_y}(x)$ when the entries of $X$ and $Y$ are Gaussian distributed. To move from the Gaussian case to non-Gaussian case, we mainly use Lindeberg’s method (see Lindeberg (1922) and Chatterjee (2006)) and the Stieltjes transform. The Stieltjes transform for any probability distribution function $G(x)$ is defined as

$$m_G(z) = \int \frac{1}{x - z} dG(x), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C}, \ v = \Im z > 0\}.$$  

(1.8)

An additional key technique is to introduce a perturbation matrix in order to deal with the random matrix $(XX^T)^{-1}$ under the finite second moment condition.

## 2 Proof of Theorem 1

We divide the proof of Theorem 1 into 4 parts:

### 2.1 Step 1: Introducing a perturbation matrix

Let

$$A = P_y P_x P_y.$$  

In view of (1.7) it is enough to investigate $F^A$ to prove Theorem 1. In order to deal with the matrix $(XX^T)^{-1}$, we make a perturbation of the matrix $A$ and obtain a new matrix

$$B = P_y P_{tx} P_y,$$

where $P_{tx} = \frac{1}{n} XX^T (\frac{1}{n} XX^T + tI_{p_1})^{-1} X$, $t > 0$ is a small constant number and $I_{p_1}$ is the identity matrix of the size $p_1$.

We claim that, with probability one,

$$\lim_{t \to 0} \lim_{n \to \infty} L(F^A, F^B) = 0.$$  

(2.1)

where $L(F^A, F^B)$ is the Levy distance between two distribution functions $F^A(\lambda)$ and $F^B(\lambda)$. By Lemma 6 in the Appendix,

$$L^3(F^A, F^B) \leq \frac{1}{n} tr(A - B)^2 \leq \frac{1}{n} tr(P_x - P_{tx})^2$$

$$= \frac{1}{n} tr(\frac{1}{n} XX^T [\frac{1}{n} XX^T]^{-1} - \frac{1}{n} XX^T + tI_{p_1})^{-1})^2$$

$$\leq \frac{t^2}{n} tr(\frac{1}{n} XX^T + tI_{p_1})^{-2},$$

(2.2)
where the second inequality uses the fact that $\|P_y\| = 1$ with the norm being the spectral norm and the last inequality uses the definition of the Moore-Penrose pseudoinverse so that we may write

$$\frac{1}{n}XX^T[(\frac{1}{n}XX^T)^{-1} - (\frac{1}{n}XX^T + tI_{p_1})^{-1}]$$

$$= U^T \begin{pmatrix} \mu_1 & \cdots & \mu_m \\ 0 & \ddots & 0 \\ \vdots & \ddots & 0 \end{pmatrix} \frac{t}{\mu_1(\mu_1 + t)} \begin{pmatrix} \mu_1 & \cdots & \mu_m \\ 0 & \ddots & 0 \\ \vdots & \ddots & 0 \end{pmatrix} \frac{t}{\mu_m(\mu_m + t)} - \frac{1}{t} \begin{pmatrix} \mu_1 & \cdots & \mu_m \\ 0 & \ddots & 0 \\ \vdots & \ddots & 0 \end{pmatrix}$$

$$= U^T \begin{pmatrix} \frac{t}{\mu_1 + t} & \cdots & \frac{t}{\mu_m + t} \\ 0 & \ddots & 0 \\ \vdots & \ddots & 0 \end{pmatrix} U.$$

Here $\mu_1, \ldots, \mu_m$ are the nonzero eigenvalues of the matrix $\frac{1}{n}XX^T$ and $U^T$ is the eigenvectors matrix of $\frac{1}{n}XX^T$.

Given $t > 0$, by Theorem 3.6 in Bai and Silverstein (2009) (or see Jonsson (1982) and Marčenko and Pastur (1967)) and the Helly-Bray theorem, we have with probability one

$$\frac{1}{n} tr(\frac{1}{n}XX^T + tI_{p_1})^{-2} = \frac{p_1}{n} \int \frac{1}{(\lambda + t)^2} dF_{p_1}(\lambda) \to c_1 \int_a^b \frac{1}{(\lambda + t)^2} dF_{c_1}(\lambda)$$

$$= \int_a^b \frac{\sqrt{(b - \lambda)(\lambda - a)}}{(\lambda + t)^2} d\lambda \leq \int_a^b \frac{\sqrt{(b - \lambda)(\lambda - a)}}{\lambda^3} d\lambda \leq M,$$

where $F_{p_1}$ is the ESD of the sample matrix $\frac{1}{n}XX^T$, $F_{c_1}$ is the Marcenko-Pastur Law, $b = (1 + \sqrt{c_1})^2$ and $a = (1 - \sqrt{c_1})^2$. Here and in what follows $M$ stands for a positive constant number and it may be different from line to line. This, together with (2.2), implies (2.1), as claimed.

Let $B$ and $A$, respectively, denote analogues of the matrices $B$ and $A$ with the elements of $X$ replaced by i.i.d. Gaussian distributed random variables, independent of the entries of $Y$. By (2.1) and the fact that, for any $\lambda \in \mathbb{R}$,

$$|F^A(\lambda) - F^A(\lambda)| \leq |F^A(\lambda) - F^B(\lambda)| + |F^B(\lambda) - F^B(\lambda)| + |F^B(\lambda) - F^A(\lambda)|,$$

in order to prove that, for any fixed $t > 0$, with probability one,

$$\lim_{n \to \infty} |F^A(\lambda) - F^A(\lambda)| = 0,$$  (2.3)
it suffices to prove with probability one,

$$\lim_{n\to\infty} |F^B(\lambda) - F^B(\lambda)| = 0.$$  \hspace{1cm} (2.4)

If we have (2.3), then for any $\lambda \in \mathbb{R}$, with probability one,

$$\lim_{n\to\infty} |F^{P^*_yP^*_y}(\lambda) - F^{P^*_yP^*_y}(\lambda)| = 0.$$  \hspace{1cm} (2.5)

Since $P_y$ and $P_x$ stand symmetric positions in the matrix $P_xP_y$, as in (2.3) and (2.5), one can similarly prove that for any $\lambda \in \mathbb{R}$, with probability one,

$$\lim_{n\to\infty} |F^{P^*_xP^*_x}(\lambda) - F^{P^*_xP^*_x}(\lambda)| = 0,$$  \hspace{1cm} (2.6)

where $P^*_y$ is obtained from the matrix $P_y$ with all the entries of $Y$ replaced by i.i.d Gaussian distributed random variables, independent of $P^*_x$. Then (2.5) and (2.6) imply that for any $\lambda \in \mathbb{R}$, with probability one,

$$\lim_{n\to\infty} |F^{P^*_xP^*_y}(\lambda) - F^{P^*_xP^*_y}(\lambda)| = 0.$$  \hspace{1cm} (2.7)

With the theorem obtained in Wachter (1980) and (2.7), our theorem is easily derived. Hence the subsequent parts are devoted to proving (2.4).

### 2.2 Step 2: Truncation, Centralization, Rescaling and Tightness of $F^B$

With (1.8) of Bai and Silverstein (2004) and the arguments above and below, we can choose $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$, $n^{1/2}\varepsilon_n \to \infty$ as $n \to \infty$, and $P(\{|X_{ij}| \geq n^{1/2}\varepsilon_n\}) \leq \frac{\varepsilon_n}{n}$. Define

$$\tilde{X}_{ij} = X_{ij}I(\{|X_{ij}| < n^{1/2}\varepsilon_n\}), \quad \hat{X}_{ij} = \tilde{X}_{ij} - E\tilde{X}_{11},$$

$$P_{tx} = \frac{1}{n}X^T(1 - XX^T + I_{p_1})^{-1}X, \quad \hat{P}_{tx} = \frac{1}{n}X^T(1 - XX^T + tI_{p_1})^{-1}X,$$

$$\hat{P}_{tx} = \frac{1}{n}X^T(1 - XX^T + tI_{p_1})^{-1}X, \quad \hat{B} = P_y\hat{P}_{tx}P_y, \quad \hat{B} = P_y\hat{P}_{tx}P_y,$$

where $\hat{X} = (\hat{X}_{ij})_{1 \leq i \leq p_1; 1 \leq j \leq n}$ and $\bar{X} = (\bar{X}_{ij})_{1 \leq i \leq p_1; 1 \leq j \leq n}$.

Let $\eta_{ij} = 1 - I(\{|X_{ij}| < n^{1/2}\varepsilon_n\})$. We then get by Lemma 4 in the appendix

$$\sup_{\lambda} |F^B(\lambda) - \hat{F}^B(\lambda)| \leq \frac{1}{n}\text{rank}(P_yP_{tx}P_y - P_y\hat{P}_{tx}P_y) \leq \frac{1}{n}\text{rank}(P_{tx} - \hat{P}_{tx})$$

$$\leq \frac{1}{n}[\text{rank}(X^T - \bar{X}) + \text{rank}(XX^T - \tilde{X}\tilde{X}^T) + \text{rank}(X - \tilde{X}^T)] \leq \frac{4}{n}\sum_{i=1}^{p_1}\sum_{j=1}^{n}\eta_{ij}.$$

Denote $q = P(\eta_{ij} = 1) = P(\{|X_{ij}| \geq n^{1/2}\varepsilon_n\})$. We conclude from Lemma 5 that for any $\delta > 0$,

$$P(\sup_{\lambda} |F^B(\lambda) - \hat{F}^B(\lambda)| \geq \delta) \leq P\left(\frac{1}{n}\sum_{i=1}^{p_1}\sum_{j=1}^{n}\eta_{ij} \geq \delta\right)$$
\[ P\left( \sum_{i=1}^{p_1} \sum_{j=1}^{n} \eta_{ij} - np_1q \geq np_1\left( \frac{\delta}{p_1} - q \right) \right) \]

\[ \leq 2\exp\left( - \frac{n^2 p_1^2 \left( \frac{\delta}{p_1} - q \right)^2}{2np_1q + np_1\left( \frac{\delta}{p_1} - q \right)} \right) \leq 2\exp(-nh), \]

for some positive \( h \). It follows from Borel-Cantelli’s lemma that

\[ \sup_{\lambda} |F^B(\lambda) - F^{\tilde{B}}(\lambda)| \to 0, \quad a.s. \quad as \ n \to \infty. \]

Next, we prove that

\[ \sup_{\lambda} |F^{\tilde{B}}(\lambda) - F^{\tilde{B}}(\lambda)| \to 0, \quad a.s. \quad as \ n \to \infty. \]  \hspace{1cm} (2.8)

Again by Lemma 4 we have

\[ \sup_{\lambda} |F^{\tilde{B}}(\lambda) - F^{\tilde{B}}(\lambda)| \leq \frac{1}{n} \text{rank}(\hat{B} - \tilde{B}) \leq \frac{1}{n} \text{rank}\left[ \hat{P}_{tx} - \tilde{P}_{tx} \right] \]

\[ \leq \frac{1}{n} \text{rank}\left[ \frac{1}{n} \hat{X}^T \left( \left( \frac{1}{n} \hat{X}\hat{X}^T + tI_{p_1} \right)^{-1} - \left( \frac{1}{n} \tilde{X}\tilde{X}^T + tI_{p_1} \right)^{-1} \right) \tilde{X} \right] \]

\[ + \frac{1}{n} \text{rank}\left[ \frac{1}{n} \hat{X}^T \left( \frac{1}{n} \hat{X}\hat{X}^T + tI_{p_1} \right)^{-1} E\tilde{X} \right] + \frac{1}{n} \text{rank}\left[ \frac{1}{n} (E\tilde{X})^T \left( \frac{1}{n} \hat{X}\hat{X}^T + tI_{p_1} \right)^{-1} \tilde{X} \right] \]

\[ + \frac{1}{n} \text{rank}\left[ \frac{1}{n} (E\tilde{X})^T \left( \frac{1}{n} \hat{X}\hat{X}^T + tI_{p_1} \right)^{-1} E\tilde{X} \right]. \]

Since all elements of \( E\tilde{X} \) are identical, \( \text{rank}\left( E\tilde{X} \right) = 1 \). Moreover, from (2.10)

\[ \left( \frac{1}{n} \tilde{X}\tilde{X}^T + tI_{p_1} \right)^{-1} - \left( \frac{1}{n} \hat{X}\hat{X}^T + tI_{p_1} \right)^{-1} \]

\[ = \left( \frac{1}{n} \tilde{X}\tilde{X}^T + tI_{p_1} \right)^{-1} \left( \frac{1}{n} \hat{X}\hat{X}^T - \frac{1}{n} \tilde{X}\tilde{X}^T \right) \left( \frac{1}{n} \tilde{X}\tilde{X}^T + tI_{p_1} \right)^{-1} \]

\[ = \frac{1}{n} \left( \tilde{X}\tilde{X}^T + tI_{p_1} \right)^{-1} \left( -E\tilde{X}E\tilde{X}^T + \tilde{X}E\tilde{X}^T + (E\tilde{X})\tilde{X}^T \right) \left( \frac{1}{n} \hat{X}\hat{X}^T + tI_{p_1} \right)^{-1}. \]

Hence

\[ \sup_{\lambda} |F^{\tilde{B}}(\lambda) - F^{\tilde{B}}(\lambda)| \leq \frac{M}{n} \to 0. \]
Let \( \hat{\sigma}^2 = E(|\hat{X}_{ij}|^2) \) and \( \mathbf{B} = \frac{1}{\hat{\sigma}^2} \hat{X}^T(\frac{1}{\hat{\sigma}^2} \hat{X} \hat{X}^T + t \mathbf{I}_{p_1})^{-1} \hat{X} \). Then by Lemma 6, we have

\[
L^2(\mathcal{F}^\mathbf{B}, F^\mathbf{B}) \leq \frac{1}{n} tr(\hat{\mathbf{B}} - \mathbf{B})^2
\]

\[
= \frac{(\hat{\sigma}^2 - 1)^2 t^2}{n} tr\left( \frac{1}{n} \hat{X} \hat{X}^T (\frac{1}{n} \hat{X} \hat{X}^T + \hat{\sigma}^2 t \mathbf{I}_{p_1})^{-1} (\frac{1}{n} \hat{X} \hat{X}^T + t \mathbf{I}_{p_1})^{-1} \right)^2
\]

\[
= \frac{(\hat{\sigma}^2 - 1)^2 t^2}{n} tr\left( \frac{1}{n} \hat{X} \hat{X}^T + \hat{\sigma}^2 t \mathbf{I}_{p_1} \right)^{-1} - \hat{\sigma}^2 t(\frac{1}{n} \hat{X} \hat{X}^T + \hat{\sigma}^2 t \mathbf{I}_{p_1})^{-1} (\frac{1}{n} \hat{X} \hat{X}^T + t \mathbf{I}_{p_1})^{-1} \right)^2
\]

\[
\leq \frac{(\hat{\sigma}^2 - 1)^2 t^2}{n} p_1 \left( \frac{1}{\sigma^2} \right)^2 \left( \frac{1}{\hat{\sigma}^2} \right)^2 \leq \frac{4}{t^2} \rightarrow 0,
\]

because \( \hat{\sigma}^2 \rightarrow 1 \) and \( p_1/n \rightarrow c_1 \) as \( n \rightarrow \infty \); where the first equality uses the formula (2.10); the second inequality uses the matrix inequality that

\[
tr(C) \leq p_1 ||C||,
\]

holding for any \( p_1 \times p_1 \) normal matrix \( C \); and the last inequality uses the fact that

\[
||\left( \frac{1}{\sigma^2} X \right)^{-1} || \leq \frac{1}{\sigma^2}, \quad ||\left( \frac{1}{\hat{\sigma}^2} \hat{X} \hat{X}^T + t \mathbf{I}_{p_1} \right)^{-1} || \leq \frac{1}{t}.
\]

In view of the truncation, centralization and rescaling steps above, in the sequel, we shall assume that the underlying variables satisfy

\[
|X_{ij}| \leq n^{1/2} \varepsilon_n, \quad E X_{ij} = 0, \quad E X_{ij}^2 = 1, \quad (2.9)
\]

and for simplicity we shall still use notation \( X_{ij} \) instead of \( \hat{X}_{ij} \).

We now turn to investigating the tightness of \( F^\mathbf{B} \). For any constant number \( K > 0 \),

\[
\int_{\lambda > K} dF^\mathbf{B} \leq \frac{1}{K} \int \lambda dF^\mathbf{B} = \frac{1}{K} \frac{1}{n} tr[\mathbf{P}_y \mathbf{P}_{tx} \mathbf{P}_y]
\]

Since the largest eigenvalue of \( \mathbf{P}_y \) is 1 and \( \mathbf{P}_{tx} \) is a nonnegative matrix we obtain

\[
tr[\mathbf{P}_y \mathbf{P}_{tx} \mathbf{P}_y] = tr[\mathbf{P}_y \mathbf{P}_{tx}]
\]

\[
\leq tr[\mathbf{P}_{tx}] = tr[\frac{1}{n} XX^T(\frac{1}{n} XX^T + t \mathbf{I}_{p_1})^{-1}] \leq n.
\]

The last inequality has used the facts that \( t > 0 \) and that all the eigenvalues of \( \frac{1}{n} XX^T(\frac{1}{n} XX^T + t \mathbf{I}_{p_1})^{-1} \) are less than 1.

It follows that \( F^\mathbf{B} \) is tight.
2.3 Step 3: Convergence of the random part

The aim in this section is to prove that

\[
\frac{1}{n} tr B^{-1}(z) - E \frac{1}{n} tr B^{-1}(z) \to 0 \quad \text{a.s. as } n \to \infty.
\]

To this end we introduce some notation. Let \( x_k \) denote the \( k \)th column of \( X \) and \( e_k \) the column vector of the size of \( p_1 \) with the \( k \)th element being 1 and otherwise 0. Moreover, define \( X_k \) to be the matrix obtained from \( X \) by replacing the elements of the \( k \)th column of \( X \) with 0.

Fix \( v = \zeta z > 0 \). Define \( \mathcal{F}_k \) to be the \( \sigma \)-field generated by \( x_1, \ldots, x_k \). Let \( E_k(\cdot) \) denote the conditional expectation with respect to \( \mathcal{F}_k \) and \( E_0(\cdot) \) denote expectation. That is, \( E_k(\cdot) = E(\cdot | \mathcal{F}_k) \) and \( E_0(\cdot) = E(\cdot) \). Let

\[
B^{-1}(z) = (P_y P_{tz} P_y - zI)^{-1}, \quad B_k = P_y P_{tk} P_y, \quad B_k^{-1}(z) = (P_y P_{tk} P_y - zI)^{-1},
\]

where \( P_{tx} = \frac{1}{n} X^T(\frac{1}{n} X X^T + \delta \mathbf{I}_{p_1})^{-1} X \), \( P_{tx} = \frac{1}{n} X^T(\frac{1}{n} X X^T + \delta \mathbf{I}_{p_1})^{-1} X \).

Define \( H_k^{-1} = (\frac{1}{n} X_k X_k^T + \delta \mathbf{I}_{p_1})^{-1} \) and \( H^{-1} = (\frac{1}{n} X X^T + \delta \mathbf{I}_{p_1})^{-1} \).

Note that \( X = X_k + x_k e_k^T \), that the elements of \( x_k e_k \) are all zero and hence that

\[
XX^T - X_k X_k^T = x_k e_k^T.
\]

This implies that

\[
H_k^{-1} - H^{-1} = \frac{1}{n} H^{-1} x_k e_k^T H_k^{-1} = \frac{1}{1 + \frac{1}{n} x_k^T H_k^{-1} x_k} \frac{1}{n} H^{-1} x_k e_k^T H_k^{-1},
\]

where we make use of the formula

\[
A_1^{-1} - A_2^{-1} = A_2^{-1}(A_2 - A_1)A_1^{-1}, \quad (2.10)
\]

holding for any two invertible matrices \( A_1 \) and \( A_2 \); and

\[
(U + uv^T)^{-1} u = \frac{U^{-1} u}{1 + v^T U^{-1} u}, \quad (2.11)
\]

holding for any invertible matrices \( U \) and \( (U + uv^T) \), vectors \( u \) and \( v \). We then write

\[
B_k - B = P_y (P_{tk} - P_{tz}) P_y = P_y (C_1 + C_2 + C_3 + C_4) P_y, \quad (2.12)
\]

where

\[
C_1 = \frac{1}{n} \frac{X_k^T H_k^{-1} x_k e_k^T H_k^{-1} X_k}{1 + \frac{1}{n} x_k^T H_k^{-1} x_k}, \quad C_2 = \frac{1}{n} \frac{X_k^T H_k^{-1} x_k e_k^T X_k}{1 + \frac{1}{n} x_k^T H_k^{-1} x_k},
\]

\[
C_3 = \frac{1}{n} \frac{e_k x_k^T H_k^{-1} X_k}{1 + \frac{1}{n} x_k^T H_k^{-1} x_k}, \quad C_4 = \frac{1}{n} \frac{e_k x_k^T H_k^{-1} e_k}{1 + \frac{1}{n} x_k^T H_k^{-1} x_k}. \quad (2.13)
\]
Now write
\[
\frac{1}{n} tr B^{-1}(z) - E \frac{1}{n} tr B^{-1}(z) = \frac{1}{n} \sum_{k=1}^{n} [E_k tr B^{-1}(z) - E_{k-1} tr B^{-1}(z)]
\]
\[
= \frac{1}{n} \sum_{k=1}^{n} (E_k - E_{k-1}) (tr B^{-1}(z) - tr B_k^{-1}(z))
\]
\[
= \frac{1}{n} \sum_{k=1}^{n} (E_k - E_{k-1}) \left[ \sum_{i=1}^{4} \text{tr} (B_k^{-1}(z) P_y C_i P_y B^{-1}(z)) \right],
\]
where the last step uses (2.10) and (2.12). Let \( \| \cdot \| \) denote the spectral norm of matrices or the Euclidean norm of vectors. It is observed that
\[
\| B^{-1}(z) \| \leq \frac{1}{v}, \quad \| B_k^{-1}(z) \| \leq \frac{1}{v}, \quad \| P_y \| \leq 1, \quad \frac{1}{p_1} tr H_k^{-1} \leq \frac{1}{t}.
\]
and since \( x^T_k H_k^{-1} x_k \geq 0 \) we have
\[
\frac{1}{1 + \frac{1}{n} x^T_k H_k^{-1} x_k} \leq 1.
\]
(2.15)

It follows that
\[
| tr B_k^{-1}(z) P_y C_i P_y B^{-1}(z) | = \frac{1}{n^2} \left| x^T_k H_k^{-1} x_k P_y B^{-1}(z) B_k^{-1}(z) P_y x^T_k H_k^{-1} x_k \right|
\]
\[
\leq \frac{1}{v^2 n^2} \| x^T_k H_k^{-1} x_k \|^2 \leq \frac{1}{v^2 n} | x^T_k H_k^{-1} x_k | + \frac{t}{v^2 n} | x^T_k H_k^{-1} x_k |,
\]
(2.16)

where the last inequality uses the facts that \( \| x^T_k H_k^{-1} x_k \|^2 = x^T_k H_k^{-1} x_k x^T_k H_k^{-1} x_k \) and
\[
H_k^{-1} x_k x^T_k H_k^{-1} = n H_k^{-1} - (1/n) x_k x^T_k + t(I_{p_1} - (I_{p_1}) H_k^{-1} = n H_k^{-1} - n t H_k^{-2}.
\]

We then conclude from Lemma 2, (2.14)-(2.16) that
\[
E \left| \frac{1}{n} \sum_{k=1}^{n} (E_k - E_{k-1}) tr B_k^{-1}(z) P_y C_i P_y B_k^{-1}(z) \right|^4 \]
\[
\leq \frac{M}{n^3} \sum_{k=1}^{n} E \left| tr B_k^{-1}(z) P_y C_i P_y B_k^{-1}(z) \right|^4
\]
\[
\leq \frac{M}{n^4} \sum_{k=1}^{n} E \left| x^T_k H_k^{-1} x_k \right|^4 + \frac{M}{n^2} \sum_{k=1}^{n} E \left| x^T_k H_k^{-2} x_k \right|^4
\]
\[
= O\left(\frac{1}{n^4}\right),
\]
where the last step uses the facts that via Lemma 3 and (2.9)
\[
\frac{1}{n^3} E \left| x^T_k H_k^{-1} x_k \right|^4 \leq \frac{1}{n^4} ME \left| x^T_k H_k^{-1} x_k - tr H_k^{-1} \right|^4 + \frac{1}{n^4} ME \left| tr H_k^{-1} \right|^4 \leq M
\]
(2.17)
and that
\[ \frac{1}{n^4}E|x_k^T H_k x_k|^4 \leq M. \] (2.18)

Similarly, we can also obtain for \( i = 2, 3, 4 \),
\[ E\left| \frac{1}{n} \sum_{k=1}^{n} (E_k - E_{k-1}) tr B^{-1}(z) P_y C_i P_y B_k^{-1}(z) \right|^4 \leq \frac{M}{n^2}. \] (2.19)

It follows from Borel-Cantelli’s lemma that
\[ \frac{1}{n} tr B^{-1}(z) = E \frac{1}{n} tr B^{-1}(z) \quad \text{a.s. } n \to \infty. \] (2.20)

2.4 Step 4: From Gaussian distribution to general distributions

This section is to prove that
\[ E\left[ \frac{1}{n} tr B^{-1}(z) \right] - E\left[ \frac{1}{n} tr D^{-1}(z) \right] \to 0 \quad \text{as } n \to \infty, \] (2.21)
where \( D^{-1}(z) = (P_y P_y^T P_y - z I)^{-1}, P_y^T = \frac{1}{n} G^T (\frac{1}{n} G G^T + t I_{p_1})^{-1} G \) and \( G = (G_{ij})_{p_1 \times n} \) consists of i.i.d. Gaussian random variables. We would point out that (2.4) follows immediately from (2.20), (2.21), tightness of \( F^B \) and the well-known inversion formula for Stieltjes transform[Theorem B.8 of Bai and Silverstein (2009)]. We use Lindeberg’s method in Chatterjee (2006) to prove this result.

To facilitate statements, denote
\[ X_{11}, \ldots, X_{in}, X_{21}, \ldots, X_{p_1n} \] respectively by \( \hat{X}_1, \ldots, \hat{X}_n, \hat{X}_{n+1}, \ldots, \hat{X}_{p_1n} \)
and
\[ G_{11}, \ldots, G_{in}, G_{21}, \ldots, G_{p_1n} \] respectively by \( \hat{G}_1, \ldots, \hat{G}_n, \hat{G}_{n+1}, \ldots, \hat{G}_{p_1n} \).

For each \( j, \ 0 \leq j \leq p_1n \), set
\[ Z_j = (\hat{X}_1, \ldots, \hat{X}_j, \hat{G}_{j+1}, \ldots, \hat{G}_{p_1n}) \text{ and } Z^0_j = (\hat{X}_1, \ldots, \hat{X}_{j-1}, 0, \hat{G}_{j+1}, \ldots, \hat{G}_{p_1n}). \] (2.22)

Note that \( X \) in \( B^{-1}(z) \) consists of the entries of \( Z_{p_1n} \). Hence we denote \( \frac{1}{n} tr B^{-1}(z) \) by \( \frac{1}{n} tr (B(Z_{p_1n}) - z I)^{-1} \). Define the mapping \( f \) from \( R^{p_1n} \) to \( C \) as
\[ f(Z_{p_1n}) = \frac{1}{n} tr (B(Z_{p_1n}) - z I)^{-1}. \] (2.23)

Furthermore we use the entries of \( Z_j, \ j = 0, 1, \ldots, p_1n - 1 \), respectively, to replace \( \hat{X}_1, \ldots, \hat{X}_{p_1n} \), the entries of \( X \) in \( B \), to constitute a series of new matrices. For these new matrices, we define \( f(Z_j), \ j = 0, 1, \ldots, p_1n - 1 \) as \( f(Z_{p_1n}) \) is defined for the matrix \( B \). For example, \( f(Z_0) = \frac{1}{n} tr D^{-1}(z) \). We then write
\[ E\left[ \frac{1}{n} tr B^{-1}(z) \right] - E\left[ \frac{1}{n} tr D^{-1}(z) \right] = \sum_{j=1}^{p_1n} E\left( f(Z_j) - f(Z_{j-1}) \right). \]
A third Taylor expansion yields

\[ f(Z_j) = f(Z_j^0) + \dot{X}_j \partial_j f(Z_j^0) + \frac{1}{2} \dot{X}_j^2 \partial^2_j f(Z_j^0) + \frac{1}{2} \dot{X}_j^3 \int_0^1 (1 - \tau)^2 \partial^3_j f(Z_j^{(1)}(\tau)) d\tau, \]

\[ f(Z_{j-1}) = f(Z_j^0) + \dot{G}_j \partial_j f(Z_j^0) + \frac{1}{2} \dot{G}_j^2 \partial^2_j f(Z_j^0) + \frac{1}{2} \dot{G}_j^3 \int_0^1 (1 - \tau)^2 \partial^3_j f(Z_j^{(2)}(\tau)) d\tau, \]

where \( \partial^r f(\cdot), \ r = 1, 2, 3, \) stand for the \( r \)-fold derivative of the function \( f \) in the \( j \)-th coordinate, and

\[
Z_j^{(1)}(t) = (\dot{X}_1, \ldots, \dot{X}_{j-1}, \tau \dot{X}_j, \dot{G}_{j+1}, \ldots, \dot{G}_m),
\]

\[
Z_j^{(2)}(t) = (\dot{X}_1, \ldots, \dot{X}_{j-1}, \tau \dot{G}_j, \dot{G}_{j+1}, \ldots, \dot{G}_m).
\]

Since \( \dot{X}_j \) and \( \dot{G}_j \) are both independent of \( Z_j^0 \), \( E[\dot{X}_j] = E[\dot{G}_j] = 0 \) and \( E[\dot{X}_j^2] = E[\dot{G}_j^2] = 1 \), we obtain

\[
E\left[ \frac{1}{n} \text{tr} B^{-1}(z) \right] - E\left[ \frac{1}{n} \text{tr} D^{-1}(z) \right] = \frac{1}{2} \sum_{j=1}^{m-1} E \left[ \dot{X}_j^3 \int_0^1 (1 - \tau)^2 \partial^3_j f(Z_j^{(1)}(\tau)) d\tau - \dot{G}_j^3 \int_0^1 (1 - \tau)^2 \partial^3_j f(Z_j^{(2)}(\tau)) d\tau \right].
\]

Next we evaluate \( \partial^3_j f(Z_{j1n}(\tau)) \). Note that

\[
\frac{\partial H^{-1}}{\partial X_{ij}} = -H^{-1} \frac{\partial H}{\partial X_{ij}} H^{-1}.
\]

A simple but tedious calculation indicates that

\[
\frac{\partial B}{\partial X_{ij}} = \frac{1}{n} P_y e_j e_i^T H^{-1} X P_y + \frac{1}{n} P_y X^T H^{-1} e_i e_j^T P_y
\]

\[ -\frac{1}{n^2} P_y X^T H^{-1} (e_i e_j^T X^T + X e_j e_i^T) H^{-1} X P_y, \]

\[
\frac{\partial^2 B}{\partial X_{ij}^2} = \frac{2}{n} P_y e_j e_i^T H^{-1} e_i e_j^T P_y - \frac{2}{n^2} P_y e_j e_i^T H^{-1} (e_i e_j^T X^T + X e_j e_i^T) H^{-1} X P_y
\]

\[ -\frac{2}{n^2} P_y X^T H^{-1} (e_i e_j^T X^T + X e_j e_i^T) H^{-1} e_i e_j^T P_y - \frac{2}{n^2} P_y X^T H^{-1} e_i e_j^T H^{-1} X P_y
\]

\[ + \frac{2}{n^3} P_y X^T [H^{-1} (e_i e_j^T X^T + X e_j e_i^T)]^2 H^{-1} X P_y, \]

12
\[
\frac{\partial^3 B}{\partial X_{ij}^3} = -\frac{6}{n^2} P_y e_j e_i^T H^{-1}(e_i e_j^T X^T + X e_j e_i^T) H^{-1} e_i e_j^T P_y - \frac{6}{n^2} P_y e_i e_i^T H^{-1} e_i e_j^T H^{-1} X P_y
\]
\[
+ \frac{6}{n^3} P_y e_j e_i^T [H^{-1}(e_i e_j^T X^T + X e_j e_i^T)^2 H^{-1} X P_y - \frac{6}{n^2} P_y X^T H^{-1} e_i e_j^T H^{-1} X P_y + \frac{6}{n^3} P_y X^T H^{-1} e_i e_j^T H^{-1} X P_y
\]
\[
+ \frac{6}{n^3} P_y X^T e_i e_j^T X^T + X e_j e_i^T) H^{-1} e_i e_j^T H^{-1} X P_y + \frac{6}{n^3} P_y X^T H^{-1} e_i e_j^T H^{-1} (e_i e_j^T X^T + X e_j e_i^T) H^{-1} X P_y.
\]

Also, by the formula
\[
\frac{1}{n} \frac{\partial \text{tr} B^{-1}(z)}{\partial X_{ij}} = -\frac{1}{n} \text{tr}(\frac{\partial B}{\partial X_{ij}} B^{-2}(z)),
\]

it is easily seen that
\[
\frac{1}{n} \frac{\partial^3 \text{tr} B^{-1}(z)}{\partial X_{ij}^3} = -\frac{6}{n} \text{tr}(\frac{\partial B}{\partial X_{ij}} B^{-1}(z)) \frac{\partial B}{\partial X_{ij}} B^{-1}(z) \frac{\partial B}{\partial X_{ij}} B^{-2}(z))
\]
\[
- \frac{1}{n} \text{tr}(\frac{\partial B}{\partial X_{ij}} B^{-2}(z)) + \frac{3}{n} \text{tr}(\frac{\partial B}{\partial X_{ij}} B^{-2}(z) \frac{\partial B}{\partial X_{ij}} B^{-1}(z))
\]
\[
+ \frac{3}{n} \text{tr}(\frac{\partial^2 B}{\partial X_{ij}^2} B^{-1}(z) \frac{\partial B}{\partial X_{ij}} B^{-2}(z)).
\]

There are lots of terms in the expansion of \( \frac{1}{n} \frac{\partial^3 \text{tr} B^{-1}(z)}{\partial X_{ij}^3} \) and therefore we do not enumerate all the terms here. By using the formula that, for any matrices \( A, B \) and column vectors \( e_j \) and \( e_k \),
\[
\text{tr}(A e_j e_k^T B) = e_k^T B A e_j,
\] (2.25)

all the terms of \( \frac{1}{n} \frac{\partial^3 \text{tr} B^{-1}(z)}{\partial X_{ij}^3} \) can be dominated by a common expression. That is
\[
|| \frac{1}{n} \frac{\partial^3 \text{tr} B^{-1}(z)}{\partial X_{ij}^3} || \leq \frac{M}{n^3} ||H^{-1}|| \cdot ||X^T H^{-1}|| + \frac{M}{n^4} ||X^T H^{-1}||^3
\]
\[
+ \frac{M}{n^4} ||H^{-1}|| \cdot ||X^T H^{-1}||^2
\]
\[
+ \frac{M}{n^4} ||H^{-1}|| \cdot ||X^T H^{-1}|| \cdot ||X^T H^{-1} X||
\]
\[
+ \frac{M}{n^5} ||H^{-1}|| \cdot ||X^T H^{-1}|| \cdot ||X^T H^{-1} X||^2
\]
\[
+ \frac{M}{n^5} ||X^T H^{-1}||^3 \cdot ||X^T H^{-1} X||
\]
\[
+ \frac{M}{n^6} ||X^T H^{-1}||^3 \cdot ||X^T H^{-1} X||^2
\]
\[
+ \frac{M}{n^7} ||X^T H^{-1}||^3 \cdot ||X^T H X||^3.
\] (2.26)
Obviously

\[ ||H^{-1}|| \leq \frac{1}{t}. \]  \hspace{1cm} (2.27)

It is observed that

\[ \|X^T H^{-1} X\|^2 = \lambda_{\text{max}}(X^T H^{-1} X X^T H^{-1} X) = \lambda_{\text{max}}(H^{-1} X X^T H^{-1} X) \]
\[ \leq n^2 [1 + 2t ||H^{-1}|| + t^2 ||H^{-2}||] \leq M n^2, \]  \hspace{1cm} (2.28)

where \( \lambda_{\text{max}}(\cdot) \) denotes the maximum eigenvalue of the corresponding matrix; and the first inequality above utilizes the fact that

\[ H^{-1} X X^T = n H^{-1} (\frac{1}{n} X X^T + t I_{p_1} - t I_{p_1}) = n I_{p_1} - n t H^{-1}. \]

Similarly, we can obtain

\[ ||X^T H^{-1}|| \leq M \sqrt{n}. \]  \hspace{1cm} (2.29)

We conclude from (2.26)-(2.29) that

\[ \frac{1}{n} \frac{\partial^3 \text{tr} B^{-1}(z)}{\partial X_{ij}^3} \| \leq \frac{M}{n^{5/2}}. \]  \hspace{1cm} (2.30)

This implies that

\[ E|X_{ij}^3 \cdot \frac{1}{n} \frac{\partial^3 \text{tr} B^{-1}(z)}{\partial X_{ij}^3} | \leq \frac{M}{n^{5/2}} E[X_{ij}^3] \leq \frac{M \varepsilon_n}{n^2}. \]  \hspace{1cm} (2.31)

Since all \( X_{ij} \) and \( W_{ij} \) play a similar role in their corresponding matrices, the above argument works for all matrices. Hence we obtain

\[ \left| E\left[ \frac{1}{n} \text{tr} B^{-1}(z) \right] - E\left[ \frac{1}{n} \text{tr} D^{-1}(z) \right] \right| \]
\[ \leq M \sum_{j=1}^{p_1 n} \int_{0}^{1} (1 - \tau)^2 E|\hat{X}_j^3 \partial_j^3 f(Z_j^{(1)}(\tau))|d\tau + \int_{0}^{1} (1 - \tau)^2 E|\hat{G}_j^3 \partial_j^3 f(Z_j^{(2)}(\tau))|d\tau \]
\[ \leq M \varepsilon_n. \]

This ensures that

\[ E\left[ \frac{1}{n} \text{tr} B^{-1}(z) \right] - E\left[ \frac{1}{n} \text{tr} D^{-1}(z) \right] \to 0 \text{ as } n \to \infty. \]

Therefore the proof of Theorem 1 is completed.

3 Conclusion

Canonical correlation coefficients play an important role in the analysis of correlations between random vectors[Anderson (1984)]. Nowadays, investigations of large dimensional random vectors attract a substantial research works, e.g. Fan and Lv (2010). As future works, we plan to develop central limit theorems for the empirical distribution of canonical correlation coefficients and make statistical applications of the developed asymptotic theorems for large dimensional random vectors.
4 Appendix

**Lemma 1** (Burkholder (1973)). Let \( \{X_k, 1 \leq k \leq n\} \) be a complex martingale difference sequence with respect to the increasing \( \sigma \)-field \( \{\mathcal{F}_k\} \). Then, for \( p \geq 2 \),

\[
E \left| \sum_{k=1}^{n} X_k \right|^p \leq K_p \left( E \left( \sum_{k=1}^{n} \left| E(X_k^2 | \mathcal{F}_{k-1}) \right| \right)^{p/2} + E \sum_{k=1}^{n} |X_k|^p \right)
\]

**Lemma 2** (Burkholder (1973)). With \( \{X_k, 1 \leq k \leq n\} \) as above, we have, for \( p > 1 \),

\[
E \left| \sum_{k=1}^{n} X_k \right|^p \leq K_p E \left( \sum_{k=1}^{n} |X_k|^2 \right)^{p/2}.
\]

**Lemma 3** (Lemma B.26 of Bai and Silverstein (2009)). For \( X = (X_1, \ldots, X_n)^T \) i.i.d standardized entries, \( C \) \( n \times n \) matrix, we have, for any \( p \geq 2 \),

\[
E|X^*CX - trC|^p \leq K_p((E|X_1|^4trCC^*)^{p/2} + E|X_1|^{2p}tr(CC^*)^{p/2}).
\]

**Lemma 4** (Theorem A.43 of Bai and Silverstein (2009)). Let \( A \) and \( B \) be two \( n \times n \) symmetric matrices. Then

\[
\|F^A - F^B\| \leq \frac{1}{n} \text{rank}(A - B),
\]

where \( \|f\| = \sup_x |f(x)| \).

**Lemma 5** (Hoeffding (1963)). Let \( Y_1, Y_2, \ldots \) be i.i.d random variables, \( P(Y_1 = 1) = q = 1 - P(Y_1 = 0) \). Then

\[
P(|Y_1 + \cdots + Y_n - nq| \geq n\varepsilon) \leq 2e^{-\frac{n^2\varepsilon^2}{2np^2q}}
\]

for all \( \varepsilon > 0 \), \( n = 1, 2, \ldots \).

**Lemma 6** (Corollary A.41 of Bai and Silverstein (2009)). Let \( A \) and \( B \) be two \( n \times n \) symmetric matrices with their respective ESDs of \( F^A \) and \( F^B \). Then,

\[
L^2(F^A, F^B) \leq \frac{1}{n} tr(A - B)^2.
\]

References


Y. R. Yang and G. M. Pan, Independence Test For High Dimensional Data Based On Regularized Canonical Correlation Coefficients. Preprint (2012)).