Ann. Inst. Statist. Math.
Vol. 50, No. 2, 223–240 (1998)

DATA SPHERING: SOME PROPERTIES AND APPLICATIONS*

JIAN ZHANG

Institute of Systems Science, Academia Sinica, Beijing 100080, China

(Received September 30, 1996; revised May 15, 1997)

Abstract. Centring-then-sphering is a very important pretreatment in data analysis. The purpose of this paper is to study the asymptotic behavior of the sphering matrix based on the square root decomposition (SRD for short) and its applications. A sufficient condition is given under which SRD has nondegenerate asymptotic distribution. As examples, some commonly used and affine equivariant estimates of the dispersion matrix are shown to satisfy this condition. The case when the population dispersion matrix varies is also treated. Applications to projection pursuit (PP) are presented. It is shown that for elliptically symmetric distributions the PP index after centring-then-sphering is independent of the underlying population location and dispersion.

Key words and phrases: Sphering, dispersion matrix, invariance, asymptotics, projection pursuit.

1. Introduction

In general, there may be two types of structure in data set: the linear (or ellipsoidal) and the nonlinear. Usually, only the former can be captured by the dispersion matrix. The latter is what the data analysts seek for in the exploratory data analysis (EDA). An efficient way to EDA is to separate off location/scale and to investigate structure beyond linear correlation and elliptic shapes in separation. Centring and, then, sphering (called centring-then-sphering) data is a simple and intuitive approach to reach this goal (see Tukey and Tukey (1981)). Similarly, in exploratory projection pursuit (PP), the interestingness is an affine invariant notion (see Huber (1985), Section 5.2). Hence an affine invariant projection index is needed. From any projection index, one can construct an affine invariant one by performing a certain centring-then-sphering on data before implementing projection pursuit (Zhang (1993*b*)). Thus, centring-then-sphering provides a convenient way to construct an affine invariant PP index. For convenience, PP analysis (index) based on centred and sphered data is called PP-after-sphering analysis (index) throughout this paper. It was found by Friedman (1987) that a substantial computational saving is attained by his PP-after-sphering analysis. Therefore, centring-then-sphering is a very important pretreatment in data analysis.

* The research is supported partly by National Natural Science Foundation of China.

The mathematical definition of centring-then-sphering was given by Jones and Sibson (1983) and is generalized in the following. Let $x_1, \ldots, x_n$ be p-vectors observed from population x with location μ and dispersion matrix Σ. Denote $X_n = (x_1, \ldots, x_n)$. Let $\mu_n = \hat{\mu}(X_n)$ and $\Sigma_n = \hat{\Sigma}(X_n)$ be affine equivariant estimates (i.e., $\hat{\mu}(AX_n + v\mathbf{1}^\tau) = A\hat{\mu}(X_n) + v$, $\hat{\Sigma}(AX_n + v\mathbf{1}^\tau) = A\hat{\Sigma}(X_n)A^\tau$, for every nonsingular $p \times p$ matrix A and p-vector v) based on X_n, where $\hat{\mu}$ and $\hat{\Sigma}$ are functionals defined on the $p \times n$ matrices. Let C^+ denote the Moore-Penrose generalized inverse of matrix C. Then centring-then-sphering, denoted by S, is a linear transformation of the form

$$S(X_n) = B_n(X_n - \mu_n \mathbf{1}^\tau) \quad \text{with} \quad B_n^\tau B_n = \Sigma_n^+, \tag{1.1}$$

where B_n is a $p \times p$ matrix. For example, let $\Sigma_n = A(n)D(n)A^\tau(n)$ be the spectral decomposition of Σ_n with

$$D(n) = \mathrm{diag}(d(n,1), \ldots, d(n,q), 0, \ldots, 0), \qquad d(n,1) \geq \cdots \geq d(n,q) > 0.$$

Let $D_1(n)$ denote the $q \times q$ diagonal matrix $\mathrm{diag}(d(n,1), \ldots, d(n,q))$. Then we can set $B_n = (D_1^{-1/2}(n), 0_{q\times(p-q)})A^\tau(n)$ in (1.1), where $0_{q\times(p-q)}$ is a zero matrix (cf. Friedman (1987), p. 251).

Suppose, for each $p \times p$ matrix A and p-vector v, that $\hat{\mu}(A(X_n - v\mathbf{1}^\tau)) = A(\hat{\mu}(X_n) - v)$ and $\hat{\Sigma}(A(X_n - v\mathbf{1}^\tau)) = A\hat{\Sigma}(X_n)A^\tau$, then, for S defined by (1.1), we have

$$\hat{\mu}(S(X_n)) = 0, \qquad \hat{\Sigma}(S(X_n)) = \mathrm{diag}(\delta_1, \ldots, \delta_p)$$

with $\mathrm{rank}(\hat{\Sigma}(S(X_n))) = \mathrm{rank}(\hat{\Sigma}(X_n))$ and $\delta_i = 0$ or 1, $1 \leq i \leq p$. That is, centring-then-spherings defined by (1.1) can remove the ellipsoidal structure in X_n. Subtracting $\mu_n \mathbf{1}^\tau$ from X_n is called centring. Transforming $X_n - \mu_n \mathbf{1}^\tau$ by premultiplication by B_n is called sphering. We call B_n a sphering matrix. When $p = 1$, sphering is reduced to the commonly used standardization for data, and B_n is uniquely determined by Σ_n up to factors ± 1. When $p > 1$, B_n is not uniquely determined by Σ_n and the requirement in (1.1). That is, there are a lot of sphering methods. Which is better? The invariance/equivariance of sphering, consistency and asymptotic normality of the corresponding sphering matrix are important criteria, proposed by Li and Zhang (1993), for choosing B_n. In literature, there are three commonly used decomposition methods to determine sphering matrices: lower triangular decomposition, square root decomposition (SRD) and one adopted by J. Friedman (1987). Li and Zhang (1993) gave the invariance/equivariance properties both for a general S and for three commonly used spherings. They also proved that, except for Friedman's sphering matrix, both the lower triangular decomposition and SRD have the asymptotic normality. However, no practical way was given to justify the sufficient condition for the asymptotic normality of SRD in that paper. This is the main topic here.

The paper is arranged as follows. The practical way to verify the sufficient condition for the asymptotic normality of SRD is established in Section 2. As examples, the commonly used estimates of dispersion matrix are shown to satisfy this sufficient condition. The case when the population dispersion matrix varies is also treated in Section 3. In Section 4, we first prove, for the elliptically distributed population, that the distribution of the maximum of each PP-after-sphering index is free from the underlying population location and dispersion. As a result, a proof for the asymptotics of Friedman's PP-after-sphering index is obtained. In view of this result, the P-value associated with this kind of index can be calculated by Sun's tail probability approximation formula (see Sun (1989)) or the bootstrap method (see Zhang (1993a)). The last section is devoted to the proofs of the theorems and propositions which appear in the previous sections.

The following notations are used throughout this paper. Let $\mathrm{DIAG}(A_1, \ldots, A_r)$ and $(A_{ij})_{1\le i,j\le r}$ denote a diagonal block matrix and a partitioned matrix, respectively, i.e.,

$$\mathrm{DIAG}(A_1,\ldots,A_r) = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{pmatrix},$$

$$(A_{ij})_{1\le i,j\le r} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1r} \\ \vdots & \vdots & \vdots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rr} \end{pmatrix}.$$

Two kinds of Vec-functions $\mathrm{Vec}(\cdot)$ and $\mathrm{Vecs}(\cdot)$ are defined by

$$\mathrm{Vec}(T) = (t_{11},\ldots,t_{p1};t_{12},\ldots,t_{p2};\ldots;t_{1p},\ldots,t_{pp})^\tau,$$
$$\mathrm{Vecs}(T) = (t_{11},\ldots,t_{p1};t_{22},\ldots,t_{p2};\cdots;t_{pp})^\tau$$

for any $p\times p$ symmetric matrix $T=(t_{ij})$. Denote by $\mathrm{diag}(b_1,\ldots,b_s)$ a $s\times s$ diagonal matrix. "$\xrightarrow{d}$" stands for convergence in distribution. "$\overset{d}{=}$" means that both sides have the same distribution.

2. Sufficient conditions for asymptotic normality of SRD

In the following, let q_i, $1\le i\le r$ be positive integers such that $p=\sum_{i=1}^r q_i$. Set $s_0=0$, $s_k=\sum_{i=1}^k q_i$, $1<k<r$.

2.1 *Case of the fixed population dispersion matrix*

In this subsection, we assume that the population dispersion matrix Σ has a spectral decomposition

$$\Sigma = ADA^\tau, \quad D=\mathrm{DIAG}(\lambda_1 I_{q_1},\ldots,\lambda_r I_{q_r}), \quad \lambda_1>\cdots>\lambda_r\ge 0, \tag{2.1}$$

and the estimator Σ_n of Σ based on X_n has a spectral decomposition

$$\Sigma_n = A(n)D(n)A^\tau(n), \quad D(n)=\mathrm{diag}(d(n,1),\ldots,d(n,p)) \tag{2.2}$$
$$d(n,1)\ge\cdots\ge d(n,p)\ge 0,$$

where A and $A(n)$ are $p \times p$ orthogonal matrices. Define $q^* = \max\{1 \le q \le p : d(n,q) > 0\}$, where, for convenience, let $\max \emptyset = 0$. Then, for $l \neq 0$, Σ_n^l and Σ^l are, respectively, defined by

$$\Sigma_n^l = A(n)\,\mathrm{diag}(d(n,1)^l, \ldots, d(n,q^*)^l, 0, \ldots, 0)A(n)^\tau;$$
$$\Sigma^l = A\,\mathrm{DIAG}(\lambda_1^l I_{q_1}, \ldots, \lambda_{q_r}^l I_{q_r})A^\tau, \quad \text{when} \quad \lambda_r > 0,$$
$$\Sigma^l = A\,\mathrm{DIAG}(\lambda_1^l I_{q_1}, \ldots, \lambda_{q_{r-1}}^l I_{q_{r-1}}, 0)A^\tau, \quad \text{when} \quad \lambda_r = 0.$$

In the above, $\Sigma^{-1/2}$ and $\Sigma_n^{-1/2}$ can be used to sphere the population and the data respectively. Obviously, for every $l \neq 0$, Σ^l can be used as a parametric description of the dispersion instead of Σ. For example, like the standard deviation of a random variable, the standard dispersion matrix $\Sigma^{1/2}$ may be more useful than Σ in analyzing the signal-noise ratio of multivariate observations. Currently, for the general $l \neq 0$, the practical applications of Σ^l are not found. However, the asymptotic result for the general case may be helpful to the state analysis of a random linear system where the polynomial matrix function plays a fundamental role.

Li and Zhang (1993) gave a theorem about the weak convergence of the estimator of Σ^l. In that theorem, they assumed:

(C21) Σ is positive definite.

(C22) $c_n \to +\infty$ and there exists a $p \times p$ symmetric random matrix $U = (U_{ij})_{1 \le i,j \le r}$ such that $c_n(\Sigma_n - \Sigma) \xrightarrow{d} V = AUA^\tau$, where U_{ij} is $q_i \times q_j$, $1 \le i,j \le r$.

(C23) U_{kk} has distinct eigenvalues, $1 \le k \le r$.

The problem how to verify (C23) was left open in that paper. In practice, some eigenvalues of Σ (or Σ_n), comparing with the others, are extremely small. We usually modify Σ (or Σ_n) through replacing these eigenvalues by zeros. We refer this to the ill-conditioned case. It is obvious that (C21) may not be true in the ill-conditioned case. In the following, a practical way to justify (C23) is provided and the condition (C21) is slightly weakened by (C21′). To begin with, we let r_0 be the positive integer such that $s_{r_0} = \mathrm{rank}(\Sigma)$. Denote

(C21′) $P(\mathrm{rank}(\Sigma_n) \le s_{r_0}) \to 1$.

(C23′) $\mathrm{Vecs}(U_{kk})$ has a Lebesgue density, $1 \le k \le r_0$.

THEOREM 2.1. *Let W be an $s \times s$ symmetric random matrix. Suppose $\mathrm{Vecs}(W)$ has a Lebesgue density function. Then the eigenvalues of W are distinct almost surely.*

Theorem 2.1 shows that we can verify (C23) by checking (C23′).

Remark 2.1. (i) Suppose $Y = (y_1, \ldots, y_s)^\tau$ has s-dimensional normal distribution with covariance Cov. Then Y has a Lebesgue density iff Cov is positive definite.

(ii) Let W be $p \times p$ symmetric random matrix and $\mathrm{Vecs}(W)$ have a Lebesgue density. Then, for any principal submatrix, say W^*, of W, $\mathrm{Vecs}(W^*)$ has a Lebesgue density, too.

(iii) Assume that Y has s-dimensional elliptically contoured distribution with dispersion matrix Cov. Then Y has a Lebesgue density iff Cov is positive definite and $\|Y^\tau \operatorname{Cov}^{-1} Y\|$ has a Lebesgue density.

Remark 2.1 shows that, for a symmetric random matrix U with a normal (or some elliptically contoured) distribution, justifying (C23′) is reduced to showing the positive definiteness of the associated covariance (or dispersion matrix).

COROLLARY OF THEOREM 2.1. *For $l \neq 0$ and under conditions* (C21′), (C22) *and* (C23′), $c_n(\Sigma_n^l - \Sigma^l) \xrightarrow{d} V^{(l)} = AU^{(l)}A^\tau$, *where* $U^{(l)} = (U_{ij}^{(l)})_{1\le i,j\le r}$, $U_{ii}^{(l)} = l\lambda_i^{l-1}U_{ii}$ *for* $i = 1,\ldots,r_0$, $U_{ii}^{(l)} = 0_{q_i\times q_i}$ *for* $i = r_0+1,\ldots,r$ *and* $U_{ij}^{(l)} = \frac{\lambda_i^l - \lambda_j^l}{\lambda_i - \lambda_j}U_{ij}$. *Especially,* $c_n(\Sigma_n^{-1/2} - \Sigma^{-1/2}) \xrightarrow{d} V^{(-1/2)}$.

Remark 2.2. We conjecture that, when Σ is positive definite, the weak convergence of $c_n(\Sigma_n - \Sigma)$ implies that of $c_n(\Sigma_n^l - \Sigma^l)$, $l \neq 0$. The next proposition supports this conjecture in part.

PROPOSITION 2.1. *Assume that* $\Sigma = \lambda_0 I_p$, $\lambda_0 > 0$, Σ_n *is an estimator of* Σ. *If* $c_n(\Sigma_n - \Sigma) \xrightarrow{d} V$, *then, for* $l \in \{\pm m/k : k, m = 1, 2, \ldots\}$, $c_n(\Sigma_n^l - \Sigma^l) \xrightarrow{d} l\lambda_0^{l-1}V$.

The propositions below are also helpful for justifying condition (C23′). To describe them, we introduce the following notations.

Let $A \otimes B$ be the Kronecker product of matrices A and B, and $E(i,j)$ a $p\times p$ matrix with a one in (i,j) position and zeros elsewhere. Denote $I_{(p,p)} = \sum_{i=1}^p \sum_{j=1}^p E(i,j) \otimes E(j,i)$. Suppose that A, B and C are, respectively, $s\times t$, $t\times u$, $u\times v$ matrices, and that D and E are $p\times p$ matrices. The following three equalities (cf. Fang and Zhang (1993)) will be used in this paper without reference.

$$\operatorname{Vec}(ABC) = (C^\tau \otimes A)\operatorname{Vec}(B), \qquad I_{(p,p)}\operatorname{Vec}(D) = \operatorname{Vec}(D^\tau),$$
$$I_{(p,p)}(D\otimes E) = (E\otimes D)I_{(p,p)}.$$

Let Z, W be two $p\times p$ random symmetric matrices with finite second moments. Suppose there exist constants ϕ_1 and ϕ_2 with $\phi_1 \ge 0$ and $\phi_2 \ge -2\phi_1/p$ such that the covariance

$$\operatorname{Cov}(\operatorname{Vec}(Z)) = \phi_1(I_{p^2} + I_{(p,p)}) + \phi_2\operatorname{Vec}(I_p)(\operatorname{Vec}(I_p))^\tau, \tag{2.3}$$
$$\operatorname{Cov}(\operatorname{Vec}(W)) = \phi_1(I_{p^2} + I_{(p,p)})(\Sigma\otimes\Sigma) + \phi_2\operatorname{Vec}(\Sigma)(\operatorname{Vec}(\Sigma))^\tau, \tag{2.4}$$

where ϕ_1 and ϕ_2 are independent of Σ. Clearly, $\operatorname{rank}\{\operatorname{Cov}(\operatorname{Vec}(Z))\} = p(p+1)/2$ provided $\phi_1 > 0$ and $p\phi_2 + 2\phi_1 > 0$.

Tyler (1982) shows that the covariance of a rotationable type distribution with respect to Σ with finite second moments is of the form (2.4).

PROPOSITION 2.2. *Assume* $\operatorname{Vec}(W)$ *has normal distribution with zero mean and covariance* $\operatorname{Cov}(\operatorname{Vec}(W))$ *given in* (2.4) *where* $\phi_1 > 0$ *and* $p\phi_2 + 2\phi_1 > 0$. *Then* *there exist another normal random symmetric matrix* U_{11} *of order* $\mathrm{rank}(\Sigma)(\mathrm{rank}(\Sigma)+1)/2$, *orthogonal matrix* Q *and diagonal matrix* D *such that* $\Sigma = QDQ^\tau$ *and* $W \overset{d}{=} Q\,\mathrm{DIAG}(U_{11}, 0)Q^\tau$, *where* $D = \mathrm{DIAG}(D_1, 0)$, $\mathrm{rank}(D_1) = \mathrm{rank}(\Sigma)$. *Furthermore,* $\mathrm{Vecs}(U_{11})$ *has a Lebesgue density.*

PROPOSITION 2.3. *Let* W *and* W_1 *be two* $p \times p$ *random symmetric matrices satisfying* $W \overset{d}{=} QW_1Q^\tau$, *where* Q *is a nonrandom orthogonal matrix. Then* $\mathrm{Vecs}(W)$ *has a Lebesgue density iff* $\mathrm{Vecs}(W_1)$ *has a Lebesgue density.*

Remark 2.3. If Σ is positive definite and $\mathrm{Vecs}(V)$ (V defined in (C22)) has a Lebesgue density, then, using Proposition 2.3 and Remark 2.1(ii), we can verify (C23) in many cases.

2.2 *Examples*

The examples below illustrate the applications of Theorem 2.1 and the propositions just introduced.

Example 1. *Sample covariance.* Let x_i, $1 \le i \le n$, be an i.i.d. random sample from an elliptically contoured distribution with location μ, dispersion Σ and finite fourth moments. Suppose $\mathrm{Cov}(x_1) = \Sigma$. Let

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad S_n = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^\tau.$$

Case 1. Σ is positive definite, x_1 has a Lebesgue density

$$f(y) = \det(\Sigma)^{-1/2} g\{(y-\mu)^\tau \Sigma^{-1}(y-\mu)\}. \tag{2.5}$$

Murihead and Waternaux (cf. Tyler (1982)) show $\sqrt{n}(S_n - \Sigma) \overset{d}{\to} N$, where N is normal with zero mean and a covariance given by (2.4). In addition, it is directly shown that $\phi_1 = 1+\kappa$, $\phi_2 = \kappa$ and $1+\kappa = (p/(p+1))ER^4/(ER^2)^2$ where

$$R^2 \text{ has density } \frac{\pi^{p/2}}{\Gamma(p/2)} y^{p/2-1} g(y). \tag{2.6}$$

Since R^2 is nondegenerate, we have $ER^4 > (ER^2)^2 > 0$. This implies $\phi_1 > 0$ and $p\phi_2 + 2\phi_1 > 0$. By Propositions 2.2, 2.3 and the corollary of Theorem 2.1, we get the desired limit distribution of $\sqrt{n}(S_n^l - \Sigma^l)$ for $l \ne 0$.

Case 2. $\mathrm{rank}(\Sigma) = p_1 < p$, $\|x_1^\tau \Sigma^+ x_1\|^2 = R^2$ has a Lebesgue density, where Σ^+ is the Moore-Penrose inverse.

In this setting, there exist an orthogonal matrix Q and a nonsingular diagonal matrix such that $\Sigma = Q\,\mathrm{DIAG}(D_1, 0)Q^\tau$. Let $y_1 = (y_{11}^\tau, y_{21}^\tau)^\tau = Q^\tau(x_1 - \mu)$, y_{11} is a vector of order $\mathrm{rank}(\Sigma)$. From $\mathrm{Cov}(y_1) = Q^\tau\Sigma Q = \mathrm{DIAG}(D_1, 0)$, it follows that $y_1 = (y_{11}^\tau, 0)^\tau$, $R^2 = \|y_{11}^\tau D_1^{-1/2} y_{11}\|$ and y_{11} has an elliptically contoured distribution. It follows from Remark 2.1(iii) that y_{11} has a Lebesgue density $f(t_1) = \det(D_1)^{-1/2} g\{t_1^\tau D_1^{-1} t_1\}$ for some function $g(\cdot)$. Note that

$$\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^\tau - \mathrm{DIAG}(D_1, 0)\right\} \xrightarrow{d} \mathrm{DIAG}(U_{11}, 0).$$

From Case 1, we deduce that $\mathrm{Vecs}(U_{11})$ has a Lebesgue density. Thus, applying the corollary of Theorem 2.1, we derive the limit distribution of $\sqrt{n}(S_n^l - \Sigma^l)$ for $l \neq 0$.

In the rest of this subsection, we let $X_n = (x_1, \ldots, x_n)$ be an i.i.d. random sample from an elliptically contoured distribution with the density given in (2.5).

Example 2. *Maximum likelihood estimate*. Suppose g in (2.5) has the first order derivative. Set $h(t) = tg'(t)/g(t)$. Let R^2 be a random vector defined as in (2.6). Suppose $Eh^2(R^2) < +\infty$. The maximum likelihood estimates (μ_n, Σ_n) for the location and scatter (μ, Σ) are the solutions to the equations

$$\Sigma_{i=1}^{n} u(t_i)(x_i - \mu_n) = 0, \qquad \Sigma_{i=1}^{n} u(t_i)(x_i - \mu_n)(x_i - \mu_n)^\tau = n\Sigma_n,$$

where $t_i = (x_i - \mu_n)^\tau \Sigma_n^{-1}(x_i - \mu_n)$ and $u(t) = -2g'(t)/g(t)$. Tyler (1982) shows that $\sqrt{n}(\Sigma_n - \Sigma) \xrightarrow{d} N$, where N is normal with covariance given by (2.4) and

$$\phi_1 = \frac{p(p+2)}{4E\{h^2(R^2)\}}, \qquad \phi_2 = \frac{-2\phi_1(1-\phi_1)}{2 + p(1-\phi_1)}.$$

It is easy to see $Eh^2(R^2) > 0$ which implies $\phi_2 > -2\phi_1/p$. Invoking Proposition 2.2, Proposition 2.3 and the corollary of Theorem 2.1, the limit distribution of $\sqrt{n}(\Sigma_n^l - \Sigma^l)$ follows for $l \neq 0$.

Example 3. *Robust estimates*. Take Maronna's M-estimates as an example. Similar results hold also for Huber's M-estimates of location and dispersion matrix (cf. Tyler (1982)), Tyler's distribution free M-estimator (Tyler (1987)), as well as S-estimator (Davies (1987)). Maronna's M-estimates of location and dispersion are the solutions of the equations

$$n^{-1}\sum_{i=1}^{n} u_1(t_i)(x_i - \mu_n) = 0, \qquad n^{-1}\sum_{i=1}^{n} u_2(t_i^2)(x_i - \mu_n)(x_i - \mu_n)^\tau = \Sigma_n,$$

where $t_i^2 = (x_i - \mu_n)^\tau \Sigma_n^{-1}(x_i - \mu_n)$. Under some conditions defined in Maronna (1976), $\Sigma_n \to V_0$, a.s. with

$$V_0 = E\{u_2((x-\mu)^\tau V_0^{-1}(x-\mu)^\tau)\} = \sigma^{-1}\Sigma.$$

Here σ is the solution to the equation $E\psi_2(\sigma R^2) = p$, R^2 is defined as in (2.6). Maronna (1976) and Tyler (1982) show

$$\sqrt{n}(\Sigma_n - V_0) \xrightarrow{d} N(0, \mathrm{Cov}) = N,$$

where Cov $=$ Cov$\{\text{Vec}(N)\}$ is of the form (3.4) with Σ replaced by V_0, and

$$\begin{aligned}&\delta_1 = E\{\psi_2^2(\sigma R^2)\}/(p(p+2)), \qquad \delta_2 = E\{\sigma R^2 \psi_2'(\sigma R^2)\}/p.\\ &\phi_1 = (p+2)^2\delta_1/(2\delta_2+p)^2,\\ &\phi_2 = \delta_2^{-2}[(\delta_1-1) - 2(\delta_2-1)\delta_1(p+(p+4)\delta_2)/(2\delta_2+p)^2].\end{aligned}$$

Since $\psi_2(\sigma R^2)$ is nondegenerate, it follows from Schwarz's inequality that

$$E\psi_2^2(\sigma R^2) > (E\{\psi_2(\sigma R^2)\})^2 = p^2,$$

which implies $(p+2)\delta_1 > p$. Note that $\psi_2' \geq 0$ gives $\delta_2 = E\sigma R^2\psi_2'(\sigma R^2)/p \geq 0$. Hence $(2\delta_2+p)^2 > 0$. Consequently,

$$(p\phi_2 + 2\phi_1)(2\delta_2+p)^2\delta_2^2 = ((p+2)\delta_1 - p)(2\delta_2+p)^2 > 0,$$

i.e., $p\phi_2 + 2\phi_1 > 0$. Appealing to Propositions 2.2 and 2.3, and the corollary of Theorem 2.1, we get the limit distribution of $\sqrt{n}(\Sigma_n^l - V_0^l)$ for $l \neq 0$.

2.3 *Case of the varying population dispersion matrix*

So far in this section we have been concerned only with the fixed population dispersion matrix. However, in some settings, the population dispersion matrix is allowed to vary. For example, in order to study the local power (or bootstrapping approximation) of a statistical test which is based on a sphered data set, we need the corresponding asymptotic properties of sphering matrices when the underlying population dispersion matrix varies. These cases are treated in this subsection.

As before, $X_n = (x_1, \ldots, x_n)$ denotes a random sample from the population x with location v_n and dispersion matrix Ω_n. Let $\Sigma_n = \hat{\Sigma}(X_n)$ be an estimator based on X_n. The p ordered eigenvalues of Σ_n and $D(n)$ are defined in (2.2). Let $\Omega_n = G_n\Delta_n G_n^\tau$ be a spectral decomposition of Ω_n, where G_n is orthogonal and $\Delta_n = \text{DIAG}(\Delta(n,1), \ldots, \Delta(n,r))$ with $\Delta(n,k) = \text{diag}(\delta(n, s_{k-1}+1), \ldots, \delta(n, s_k))$, $1 \leq k \leq r$. Assume that $\Omega_n \to \Sigma$ and Σ is positive definite. Without loss of generality, it is assumed that $\Sigma = \text{DIAG}(\lambda_1 I_{q_1}, \ldots, \lambda_r I_{q_r})$ in this subsection. Put

$$\bar{\delta}_k(n) = \frac{1}{q_k}\sum_{i=s_{k-1}+1}^{s_k} \delta(n,i), \qquad \Lambda(n,k) = c_n(\Delta(n,k) - \bar{\delta}_k(n) I_{q_k}), \qquad 1 \leq k \leq r,$$

and

$$\Lambda(n) = \text{DIAG}(\Lambda(n,1), \ldots, \Lambda(n,r)).$$

As a consequence of $\Omega_n \to \Sigma$, $\bar{\delta}_k(n) \to \lambda_k$, $1 \leq k \leq r$.

THEOREM 2.2. *Assume that* (i) $\Lambda(n) \to \Lambda = \text{diag}(r_1, \ldots, r_p) = \text{DIAG}(\Lambda_1, \ldots, \Lambda_r)$, $G_n \to I_p$; (ii) $c_n \to +\infty$, $c_n(\Sigma_n - \Omega_n) \xrightarrow{d} U = (U_{st})_{1\leq s,t\leq r}$ *and* Vecs(U) *with a Lebesgue density*; (iii) $\Omega_n \to \Sigma$ *and* Σ *is positive definite. Then, for* $l \neq 0$,

$$c_n(\Sigma_n^l - \Omega_n^l) \xrightarrow{d} U^{(l)} = (U_{st}^{(l)})_{1\leq s,t\leq r},$$

with

$$U_{ss}^{(l)} = l\lambda_s^{l-1}U_{ss}, \qquad 1 \leq s \leq r; \qquad U_{st}^{(l)} = \frac{\lambda_s^l - \lambda_t^l}{\lambda_s - \lambda_t}U_{st}, \qquad s \neq t.$$

3. Applications to PP

As stated in Section 1, centring-then-sphering is a useful pretreatment in data analysis, especially in exploratory projection pursuit. Let $Q(\cdot)$ be a n-variate function for measuring interestingness of a projection. Let S be a centring-then-sphering method based on an i.i.d. sample $X_n = (x_1, \ldots, x_n)$. Then the PP-after-sphering index is of the form $Q(a^\tau S(X_n))$, where a is a direction. Let

$$Q_M(X_n, S) = \sup_{\|a\|=1} Q(a^\tau S(X_n)). \tag{3.1}$$

$Q_M(X_n, S)$ can be used to develop a significant test for the departure from "uninteresting" model in the PP setting (see Sun (1989)). Using Friedman's sphering and Legendre's polynomials, Friedman (1987) defined a kind of PP-after-sphering indices which are named FSPs by us. To get the P-values of the maxima of PSPs, Sun (1989) obtained the limit distributions for these maxima when the population is $N(\mu, \Sigma)$. In her proof, she required $B_n = B + 0_p(1/\sqrt{n})$ (B_n is a sphering matrix and see Sun (1989), p. 82) which, however as shown in Li and Zhang (1993), does not hold in general when Σ has an eigenvalue of multiplity bigger than one. In this section, by virtue of the affine invariance of PP index and Theorem 2.6 in Li and Zhang (1993), we prove Sun's result without the condition $B_n = B + 0_p(1/\sqrt{n})$ under that the population has an elliptically contoured distribution.

Let $\{l_j(\cdot) : 1 \le j \le J\}$ be the Legendre polynomials on $[-1, 1]$ defined as

$$l_0(v) = 1, \qquad l_1(v) = v,$$
$$l_j(v) = [(2j+1)v l_{j-1}(v) - (j-1) l_{j-2}(v)]/j, \qquad j \ge 2.$$

It is known

$$\int_{-1}^{1} l_j(t) l_i(t) dt = \begin{cases} 0, & j \ne i; \\ \dfrac{2}{2j+1}, & \text{otherwise.} \end{cases}$$

Let x_i, $1 \le i \le n$ are i.i.d. p-variate observations from an elliptically distributed random vector $x = EC(\mu, \Sigma)$, where Σ is positive definite and x has zero probability at zero. It follows from Fang and Zhang ((1993), p. 92) that there exists a spherically distributed random vector y with a common marginal density $\psi(t)$ (the corresponding distribution function $\Psi(t)$) such that $x = \mu + Ay$. Let $\mu_n \hat{=} \hat{\mu}(X_n)$, $\Sigma_n = \hat{\Sigma}(X_n) > 0$ be estimators of μ and Σ based on X_n as before. In this subsection we assume that Σ_n and Σ are positive definite. Let $\Sigma_n^{-1} = B_n^\tau B_n$, $\Sigma^{-1} = B^\tau B$ be the decompositions of Σ_n^{-1}, and Σ^{-1}, respectively, where B_n and B are $p \times p$ matrices. Set

$$V_n = \sqrt{n}(B_n - B), \qquad \eta_n = \sqrt{n}(\mu_n - \mu).$$

For convenience, write

$$z^{B_n}(u) \hat{=} B_n(u - \mu_n), \qquad z^B(u) \hat{=} B(u - \mu), \qquad z(u) \hat{=} u - \mu,$$
$$g(\alpha, u) = 2\Psi(\alpha^\tau u) - 1, \qquad \bar{l}_j^{B_n}(\alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_j(g(\alpha, z^{B_n}(x_i))).$$

The sample version and abstract version of Friedman type PP-after-sphering index are, respectively, defined as

$$\hat{I}_J^{B_n}(\alpha) = \frac{1}{2n}\sum_{j=1}^{J}(2j+1)(\bar{l}_j^{B_n}(\alpha))^2 \qquad \text{(the sample version)},$$

$$I_J^B(\alpha) = \frac{1}{2}\sum_{j=1}^{J}(2j+1)^2(El_j(g(\alpha, z^B(x))))^2 \qquad \text{(the abstract version)}.$$

It is concluded from Theorem 2.3 in Li and Zhang (1993) that:

LEMMA 3.1. *For X_n fixed, $Q_M(X_n, S)$ defined in (3.1) is a constant functional in S. Specifically, for Friedman's PP-after-sphering index, we have that, for any $p \times p$ matrices A_n and B_n with $B_n^\tau B_n = A_n^\tau A_n = \Sigma_n^{-1}$, $\sup_{\|\alpha\|=1} \hat{I}_J^{B_n}(\alpha) = \sup_{\|\alpha\|=1} \hat{I}_J^{A_n}(\alpha)$. Similarly, for any $p \times p$ matrices A and B with $B^\tau B = A^\tau A = \Sigma^{-1}$, $\sup_{\|\alpha\|=1} I_J^B(\alpha) = \sup_{\|\alpha\|=1} I_J^A(\alpha)$.*

THEOREM 3.1. *For the elliptically distributed population with nonsingular dispersion matrix, the distribution of $Q_M(X_n, S)$ is free from the population location and dispersion.*

The next theorem is a generalization of Sun's result (cf. Sun (1989), pp. 76–89). Note first that P_n and P below denote the empirical distribution based on X_n and the distribution of x, respectively. The following assumptions are used in this theorem.

(C31) There exists matrix-valued functional $M(u) = (m_{ij}(u))_{p\times p}$, defined on R^p, such that

$$\sqrt{n}(\Sigma_n - \Sigma) = \sqrt{n}(P_n - P)M(\cdot) + o_p(1) \hat{=} (\sqrt{n}(P_n - P)m_{ij}(\cdot))_{p\times p} + o_p(1)$$

which converges weakly. $\|\operatorname{Vec}(M(u))\|$ has a finite second moment.

For instance, if Σ_n is the sample covariance matrix based on X_n, then $M(u) = (u-\mu)(u-\mu)^\tau$.

(C32) There exists R^p-valued functional $n(u) = (n_i(x))_{p\times 1}$ defined on R^p, such that

$$\sqrt{n}\eta_n = \sqrt{n}(P_n - P)n(\cdot) + o_p(1) \hat{=} (\sqrt{n}(P_n - P)n_i(\cdot))_{p\times 1} + o_p(1),$$

which converges weakly. $\|n(u)\|$ has a finite second moment.

(C33) $\psi(t)$ is bounded and continuous, $E\|x\| < +\infty$.

Set

$$\beta_{j_1}(\alpha) = 2El_j'(g(\alpha, z^B(x)))\psi(\alpha^\tau z^B(x)),$$
$$\beta_{j_2}(\alpha) = 2El_j'(g(\alpha, z^B(x)))\psi(\alpha^\tau z^B(x))z^B(x), \qquad 1 \le j \le J.$$

Obviously, (C33) implies that $\beta_{j_1}(\alpha)$ and $\beta_{j_2}(\alpha)$ are well defined and finite.

LEMMA 3.2. *Let*

$$\beta_{j_1} = \begin{cases} 0, & j \text{ is even}, \\ 2\int l_j'(2\Psi(t)-1)\psi^2(t)dt, & \text{otherwise}, \end{cases}$$

$$\beta_{j_2} = \begin{cases} 0, & j \text{ is odd}, \\ 2\int l_j'(2\Psi(t)-1)\psi^2(t)tdt\alpha, & \text{otherwise}. \end{cases}$$

Then $\beta_{j_1}(\alpha) \equiv \beta_{j_1}$ *and* $\beta_{j_2}(\alpha) = \beta_{j_2}\alpha$.

Write

$$f_j(\alpha,\cdot) = \begin{cases} l_j(g(\alpha,z(\cdot))) - \frac{1}{2}(\alpha^\tau M(\cdot)\alpha^\tau - 1)\beta_{j_1}, & \text{when } j \text{ is even}, \\ l_j(g(\alpha,z(\cdot))) - \alpha^\tau n(\cdot)\beta_{j_2}, & \text{otherwise}, \end{cases}$$

and

$$\mathcal{F} = \{f_j(\alpha,\cdot) : 1 \le i \le J, \|\alpha\| = 1\}.$$

THEOREM 3.2. *Let* $B_n^\tau B_n = \Sigma_n^{-1}$ *and* $B^\tau B = \Sigma^{-1}$ *be decompositions of* Σ_n^{-1} *and* Σ^{-1}, *respectively. Then, under* (C31), (C32) *and* (C33),

$$n \sup_{\|\alpha\|=1} \hat{I}_J^{B_n}(\alpha) \xrightarrow{d} G = \frac{1}{2} \sup_{\|\alpha\|=1} \sum_{j=1}^{J} (2j+1)(h_j(\alpha))^2 \tag{3.2}$$

where W *is a* P *bridge indexed by* $\mathcal{F}$ *(see Pollard* (1984), p. 149 *for the definition of* P-*bridge)* $h_j(\alpha) = W(f_j(\alpha,\cdot))$, $1 \le j \le J$; *and* G *is free from both* μ *and* Σ.

4. Proofs

The proof of Theorem 2.1 is based on the following lemma due to Okamoto (1973).

LEMMA 4.1. *If* $f(t_1,\ldots,t_m)$ *is a polynomial in real variables* $t_1,\ldots,t_m$, *which is not identically zero, then the subset* $N_m = \{(t_1,\ldots,t_m) \mid f(t_1,\ldots,t_m) = 0\}$ *of the Euclidean m-space* R^m *has Lebesgue measure zero.*

PROOF OF THEOREM 2.1. The idea behind our proof is similar to that used in Okamoto (1973). Observe that the eigenvalues, denoted by $\delta_1,\ldots,\delta_s$, of W are the roots of the equation

$$g(\lambda) = \det(W - \lambda I_s) = 0,$$

and $g(\lambda)$ can be written as

$$g(\lambda) = \sum_{i=0}^{s} a_i(\mathrm{Vecs}(W))\lambda^i = b_0(\lambda-\delta_1)\cdots(\lambda-\delta_s),$$

where $a_i(\text{Vecs}(W))$ is a polynomials in the elements of $\text{Vecs}(W)$, $b_0 \neq 0$. Let $D(W) = \prod_{i<j}(\delta_i - \delta_j)^2$. By a well-known theorem in algebra (cf. Va der Waerden (1949), p. 82), $D(W)$ can be written as a polynomial in the elements of $\text{Vecs}(W)$, and, in addition, the eigenvalues of W are distinct iff $D(W) \neq 0$. So to complete the proof, we only need to prove $P_r(D(W) \neq 0) = 1$. According to Lemma 4.1, it suffices to prove that the function $D(T)$ in the elements of $\text{Vecs}(T)$ is not identically zero. This is clear because we can choose $T = \text{diag}(t_{11}, \ldots, t_{pp})$ with $t_{11} > \cdots > t_{pp}$, then $D(T) \neq 0$.

PROOF OF REMARK 2.1. (i) and (ii) are obvious, and (iii) immediately follows from Corollary 1 in Fang and Zhang ((1993), p. 84).

LEMMA 4.2. *Assume that* (i) D *is given in* (2.1); (ii) $W = (W_{ij})_{1\le i,j\le r}$ *is a* $p \times p$ *(non-random) symmetric matrix,* W_{ij} *is* $q_i \times q_j$ *matrix, and for* $1 \le k \le r_0$, W_{kk} *has distinct eigenvalues* $h(s_{k-1}+1) > \cdots > h(s_k)$; (iii) $c_n \to +\infty$, *nonrandom* $S(n)$ *is a* $p \times p$ *nonnegative definite matrix, with* $\text{rank}(S(n)) \le \text{rank}(D)$ *and* $c_n(S(n) - D) \to W$. *Then, for any* $l \neq 0$,

$$c_n(S(n)^l - D^l) \to W^{(l)} = (W_{ij}^{(l)})_{1\le i,j\le r},$$

where $W_{ii}^{(l)} = l\lambda_i^{l-1}W_{ii}$ *for* $i = 1, \ldots, r_0$, $W_{ii}^{(l)} = 0_{q_i\times q_i}$ *for* $i = r_0+1, \ldots, r$ *and* $W_{ij}^{(l)} = \frac{\lambda_i^l - \lambda_j^l}{\lambda_i - \lambda_j} W_{ij}$, $i \neq j$.

PROOF OF LEMMA 4.2. A slight modification of the proof of Lemma 2.5 in Li and Zhang (1993) gives the desired assertion.

PROOF OF THE COROLLARY OF THEOREM 2.1. It is a direct consequence of Theorem 2.1 and Lemma 4.2 and Almost Sure Representation Theorem (Pollard (1984), p. 71).

PROOF OF PROPOSITION 2.1. When $k = m = 1$, we have

$$c_n(\Sigma_n^{-1} - \lambda_0^{-1}I_p) = -c_n\lambda_0^{-1}(\Sigma_n - \lambda_0 I_p)\Sigma_n^{-1}$$

which converges weakly to $-\lambda_0^{-2}V$ by the assumption.

When $\max\{m, k\} \ge 2$, taking $l = -1/k$ as an example, we have

$$\begin{aligned}
&c_n(\Sigma_n^{-1/k} - \lambda_0^{-1/k}I_p)\\
&= -c_n\lambda_0^{-1/k}(\Sigma_n - \lambda_0 I_p)\\
&\quad \cdot (\Sigma_n^{(k-1)/k} + \Sigma_n^{(k-2)/k}\lambda_0^{1/k} + \cdots + \Sigma_n^{1/k}\lambda_0^{(k-2)/k}I_p)^{-1}\Sigma_0^{-1/k}\\
&= -c_n\lambda_0^{-1/k}(\Sigma_n - \lambda_0 I_p)k^{-1}\lambda_0^{-(k-1)/k}\lambda_0^{-1/k}(1 + o_p(1))\\
&\xrightarrow{d} -k^{-1}\lambda_0^{-(k+1)/k}V - l\lambda_0^{l-1}V.
\end{aligned}$$

The following two lemmas are required in the proofs of Propositions 2.2 and 2.3.

LEMMA 4.3. *In (2.4), if $\phi_1 > 0$, $p\phi_2 + 2\phi_1 > 0$, then*

$$\mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(W))\} = \mathrm{rank}(\Sigma)(\mathrm{rank}(\Sigma)+1)/2 = \mathrm{rank}\{\mathrm{Cov}(\mathrm{Vecs}(W))\}.$$

PROOF. Clearly, $\mathrm{Cov}\{\mathrm{Vecs}(W)\}$ can be obtained by deleting the $[(j-1)p+i]$-th, $1 \le i < j \le p$ rows and columns of $\mathrm{Cov}\{\mathrm{Vec}(W)\}$, which are the copies of $[(i-1)p+j]$-th, $1 \le i < j \le p$, respectively. So $\mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(W))\} = \mathrm{rank}\{\mathrm{Cov}(\mathrm{Vecs}(W))\}$. Let $\Sigma = Q\,\mathrm{DIAG}(D_1^2, 0)Q^\tau$ $(= \{Q\,\mathrm{DIAG}(D_1,0)\} \cdot \{\mathrm{DIAG}(D_1,0)Q^\tau\})$ be the spectral decomposition of Σ. Then,

$$\begin{aligned}\mathrm{Cov}\{\mathrm{Vec}(W)\} = (Q \otimes Q^\tau)&(\mathrm{DIAG}(D_1,0) \otimes \mathrm{DIAG}(D_1,0))\\ &\times [\phi_1(I_{p^2} + I_{(p,p)}) + \phi_2\,\mathrm{Vec}(I_p)(\mathrm{Vec}(I_p))^\tau]\\ &\times (\mathrm{DIAG}(D_1,0) \otimes \mathrm{DIAG}(D_1,0))(Q^\tau \otimes Q^\tau).\end{aligned}$$

So to show $\mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(W))\} = \mathrm{rank}(\Sigma)(\mathrm{rank}(\Sigma)+1)/2$, we only need to show it for $\Sigma = \mathrm{DIAG}(D_1^2, 0)$. To this end, we partition Z (defined in Subsection 2.3), according to the order of D_1, as

$$Z = \begin{pmatrix} Z_{11}, & Z_{12} \\ Z_{21}, & Z_{22} \end{pmatrix}$$

and let

$$Y \doteq \mathrm{DIAG}(D_1,0)Z\,\mathrm{DIAG}(D_1,0) = \mathrm{DIAG}(D_1 Z_{11} D_1, 0), \qquad EZ = 0.$$

Note that $\mathrm{rank}\{\mathrm{Cov}(\mathrm{Vecs}(Z))\} = \mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(Z))\} = p(p+1)/2$ results in the positive definiteness of $\mathrm{Cov}(\mathrm{Vecs}(Z))$. Thus $\mathrm{Cov}(\mathrm{Vecs}(Z_{11}))$ is also positive definite. In particular,

$$\begin{aligned}\mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(D_1 Z_{11} D_1))\} &= \mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(Z_{11}))\}\\ &= \mathrm{rank}(\Sigma)(\mathrm{rank}(\Sigma)+1)/2.\end{aligned}$$

Combining this with $\mathrm{Cov}(\mathrm{Vec}(W)) = \mathrm{Cov}(\mathrm{Vec}(Y))$, we deduce

$$\begin{aligned}\mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(W))\} &= \mathrm{rank}\{\mathrm{Cov}(\mathrm{Vec}(D_1 Z_{11} D_1))\}\\ &= \mathrm{rank}(\Sigma)(\mathrm{rank}(\Sigma)+1)/2.\end{aligned}$$

LEMMA 4.4. *For any $p \times p$ orthogonal matrix Q and $p \times p$ random symmetric V, there exists $[p(p+1)/2] \times [p(p+1)/2]$ nonsingular matrix G such that $\mathrm{Vecs}(QVQ^\tau) = G\,\mathrm{Vecs}(V)$.*

PROOF. Let $Q = (q_1, \ldots, q_p)$, $V = (v_{ij})_{p\times p}$, q_i is p-vector. Then

$$(4.1)\qquad \begin{cases} Q \otimes Q = (q_1 \otimes q_1, q_1 \otimes q_2, \cdots, q_p \otimes q_p), \\ \mathrm{Vec}(QVQ^\tau) = (Q \otimes Q)\,\mathrm{Vec}(V) = \sum_{i,j} (q_j \otimes q_i) v_{ji} \\ \qquad = \sum_{i>j} (q_j \otimes q_i + q_i \otimes q_j) v_{ji} + \sum_{i=1}^{p} (q_i \otimes q_i) v_{ii} \\ \qquad = G_1\,\mathrm{Vecs}(V) \end{cases}$$

with

$$G_1 = (q_1 \otimes q_1, q_1 \otimes q_2 + q_2 \otimes q_1, \cdots, q_p \times q_p).$$

Observe that $G_1^\tau G_1 = I_{p(p+1)/2}$, and, in the rows of G_1, the $[(j-1)p+i]$-th is a copy of the $[(i-1)p+j]$-th, $1 \le i < j \le p$. Let G be the remainder of G_1 after deleting the $[(j-1)p+i]$-th, $1 \le i < j \le p$, rows. Then, $\mathrm{rank}(G) = \mathrm{rank}(G_1) = p(p+1)/2$, and (4.1) yields $\mathrm{Vecs}(QVQ^\tau) = G\,\mathrm{Vecs}(V)$.

PROOF OF PROPOSITION 2.2. Let Z be $p \times p$ normal random matrix with zero mean and covariance $\mathrm{Cov}(\mathrm{Vec}(Z))$ given by (2.3). Partition Z, according to D, as

$$Z = \begin{pmatrix} Z_{11}, & Z_{12} \\ Z_{21}, & Z_{22} \end{pmatrix}.$$

Set

$$U \hat{=} \begin{pmatrix} U_{11}, & U_{12} \\ U_{21}, & U_{22} \end{pmatrix} = DZD = \mathrm{DIAG}(D_1 Z_{11} D_1, 0).$$

Then

$$\mathrm{Cov}(\mathrm{Vec}(W)) = \mathrm{Cov}(\mathrm{Vec}(QUQ^\tau)) = (Q \otimes Q^\tau)\,\mathrm{Cov}(\mathrm{Vec}(U))(Q^\tau \otimes Q^\tau).$$

Hence, by Lemma 4.3, we have $W \overset{d}{=} QUQ^\tau$, and

$$\mathrm{rank}(\mathrm{Cov}(\mathrm{Vecs}(U_{11}))) = \mathrm{rank}(\mathrm{Cov}(\mathrm{Vecs}(U))) = \mathrm{rank}(\Sigma)(\mathrm{rank}(\Sigma)+1)/2.$$

Consequently, $\mathrm{Cov}\{\mathrm{Vecs}(U_{11})\}$ is positive definite. From Remark 2.1(i), it is shown that $\mathrm{Vecs}(U_{11})$ has a Lebesgue density.

PROOF OF PROPOSITION 2.3. It directly follows from Lemma 4.4.

PROOF OF THEOREM 2.2. The proof is analogous in character to that of Theorem 2.5 in Li and Zhang (1993) (or Theorem 1 in Anderson (1963)). Without loss of generality, we assume that $\Omega_n = \Delta_n$, $G_n = I_p$. It is shown by Lemma 3.2 that the eigenvalues of $U_{kk} + \Lambda_k$ are distinct almost surely. Appealing to Almost Sure Representation Theorem (Pollard (1984), p. 71), to finish the proof, we only need to show the following lemma.

LEMMA 4.5. *Assume that* (i) Δ_n, $\Lambda(n)$ *and* Λ *are given as in Theorem* 2.2; (ii) $W = (W_{ij})_{1\le i,j\le r}$ *is a* $p \times p$ *(nonrandom) symmetric matrix,* W_{ij} *is* $q_i \times q_j$ *matrix,* $\sum_{k=1}^r q_k = p$, *and for* $1 \le k \le r$, $W_{kk} + \Lambda_k$ *has distinct eigenvalues,* (iii) $c_n \to +\infty$, *nonrandom* S_n *is a* $p \times p$ *nonnegative definite matrix and* $c_n(S_n - \Delta_n) \to W$. *Then, for any* $l \ne 0$,

$$c_n(S_n^l - \Delta_n^l) \to W^{(l)} = (W_{ij}^{(l)})_{1\le i,j\le r},$$

where $W_{ii}^{(l)} = l\lambda_i^{l-1} W_{ii}$, $W_{ij}^{(l)} = \frac{\lambda_i^l - \lambda_j^l}{\lambda_i - \lambda_j} W_{ij}$, $i \ne j$.

PROOF OF LEMMA 4.5. Since the proof is similar to that of Lemma 2.5 in Li and Zhang (1993), and, therefore, is omitted.

PROOF OF THEOREM 3.1. Let $\Sigma^{-1} = L^\tau L$ be the lower triangular decomposition of Σ^{-1}. Since, under lower triangular transformation, the lower triangular sphering is invariant (see Li and Zhang (1993)), i.e., if S is the lower triangular sphering, then $S(L(X_n - \mu\mathbf{1}^\tau)) = S(X_n)$. And it easily follows from the assumption that the distribution $L(X_n - \mu\mathbf{1}^\tau)$ is free from μ and Σ. Thus, in view of Lemma 3.1, the desired assertion follows.

PROOF OF THEOREM 3.2. In view of Theorem 3.1, without loss of generality, we assume that $B_n = \Sigma_n^{-1/2}$ and $B = \Sigma = I_p$. Then the proof is splitted into the following three lemmas.

LEMMA 4.6. *Write* $W_n \hat{=} \{W_n(f) : f \in \mathcal{F}\} \hat{=} \{\sqrt{n}(P_n - P)f : f \in \mathcal{F}\}$. *Then, under* (C31), (C32) *and* (C33), $W_n \to W$.

LEMMA 4.7. *For* $1 \le j \le J$, $\|\alpha\| = 1$, *let* $h_{n_j}(\alpha) \hat{=} \sqrt{n}(P_n - P)f_j(\alpha, \cdot)$. *Under* (C31), (C32) *and* (C33),

$$n \sup_{\|\alpha\|=1} \hat{I}_J(\alpha) = \frac{1}{2} \sup_{\|\alpha\|=1} \sum_{j=1}^{J} (2j+1)(h_{n_j}(\alpha))^2 + o_p(1). \tag{4.2}$$

LEMMA 4.8. *Under conditions* (C31), (C32) *and* (C33), (3.2) *holds.*

PROOF OF LEMMA 4.6. Note that $l_j'(t)$, $1 \le j \le J$ are either symmetric or asymmetric functions in t according to j being odd and even. This together with $B(x - \mu)$ following a p-variate spherical distribution shows that $El_j(g(\alpha, z^B(x))) = 0$ and $\beta_{j_1}(\alpha) = \beta_{j_1}$. To show $\beta_{j_2}(\alpha) = \beta_{j_2}\alpha$, we choose submatrix Q_1 such that $Q \hat{=} (\alpha, Q_1)$ is a $p \times p$ orthogonal matrix. Let $u = (u_1, u_2^\tau)^\tau = Q^\tau z^B(x)$, u_1 is real. Then

$$\begin{aligned}
\beta_{j_1}(\alpha) &= 2El_j'(2\Psi(t) - 1)\psi^2(t)dt, \\
\beta_{j_2}(\alpha) &= 2El_j'(2\Psi(u_1) - 1)\psi(u_1)(\alpha u_1 + Q_1 u_2) \\
&= \begin{cases} 0, & j \text{ is odd}, \\ 2\int l_j'(2\Psi(t) - 1)\psi^2(t)t\,dt\,\alpha, & \text{otherwise}. \end{cases}
\end{aligned}$$

The proof is finished.

PROOF OF LEMMA 4.7. From the multivariate central limit theorem, it follows that, for each finite subset $\{f_1, \ldots, f_k\} \subset \mathcal{F}$, $\{W_n(f_i) : 1 \le i \le k\}$ converges weakly to a multivariate normal random vector. It is analogous to the proof of Theorem 2.1 in Zhang (1993*a*) to show that the entropy condition in the equicontinuity lemma (Pollard (1984), p. 150) holds for $\mathcal{F}$. Invoke the central limit theorem of empirical processes (Pollard (1984), p. 157), we establish the desired assertion.

The following statement is useful in proving Lemma 4.7.

LEMMA 4.9. *For* $\|\alpha\| = 1$, $p \times p$ *nonsingular matrix* A, p-*vector* v *and* $1 \le j \le J$, *define*

$$g_{1j}(\alpha, A, v, x_i) = l_j'(2\Psi(\alpha^\tau A(x_i - v)) - 1)\psi(\alpha^\tau A(x_i - v)),$$
$$g_{2j}(\alpha, A, v, x_i) = g_{1j}(\alpha, A, v, x_i)(x_i - v).$$

Let

$$K_{jn_1} = \frac{1}{n}\sum_{i=1}^{n} g_{1j}(\alpha, A_n, \mu_n, x_i), \qquad K_{jn_2} = \frac{1}{n}\sum_{i=1}^{n} g_{2j}(\alpha, A_n, \mu_n, x_i),$$

where $A_n = I_p + o_p(1)$, $\mu_n = \mu + o_p(1)$. *Then, for* $1 \le j \le J$,

$$\sup_{\|\alpha\|=1} |K_{n1}(\alpha) - \beta_{j1}(\alpha)| = o_p(1), \qquad \sup_{\|\alpha\|=1} |K_{n2}(\alpha) - \beta_{j2}(\alpha)| = o_p(1). \tag{4.3}$$

PROOF. Take the second equality in (4.3) as an example. First, we split K_{n2} into two parts: $K_{n2}^{(1)}$ and $K_{n2}^{(2)}$, where

$$K_{jn_2}^{(1)} = \frac{1}{n}\sum_{i=1}^{n} g_{2j}(\alpha, A_n, \mu_n, x_i) I_{[\|x_i\| \le t_0]},$$
$$K_{jn_2}^{(2)} = \frac{1}{n}\sum_{i=1}^{n} g_{2j}(\alpha, A_n, \mu_n, x_i) I_{[\|x_i\| > t_0]}.$$

Here, $I_{[\cdot]}$ denotes the indicator function of a set. Then, for every $\varepsilon > 0$, by (C33) and the strong large number law, there exists a large t_0 such that

$$\limsup \sup_{\|\alpha\|=1} |K_{n2}^{(2)}(\alpha)| \le \varepsilon, \qquad \sup_{\|\alpha\|=1} |Eg_{2j}(\alpha, I_p, \mu, x) I_{[\|x\| \le t_0]} - \beta_{j_2}(\alpha)| \le \varepsilon.$$

For this fixed t_0, by the uniform continuity of g_{2j} on any bounded and closed region, and the strong large number law, we obtain

$$\limsup \sup_{\|\alpha\|=1} |K_{n2}^{(1)}(\alpha) - Eg_{2j}(\alpha, I_p, \mu, x) I_{[\|x\| \le t_0]}| \le \varepsilon.$$

Hence

$$\limsup \sup_{\|\alpha\|=1} |K_{n2}(\alpha) - \beta_{j_2}(\alpha)| \le 3\varepsilon.$$

The proof is completed.

PROOF OF LEMMA 4.7. Let $G(\theta) = n^{-1/2} \sum_{i=1}^{p} l_j(2\Psi(\alpha^\tau \bar\theta_1(x_i - \bar\theta_2)) - 1)$ with

$$\bar\theta_1 = I_p + \theta(B_n - I_p), \qquad 0 \le \theta \le 1, \qquad \bar\theta_2 = \mu + \theta(\mu_n - \mu), \qquad 0 \le \theta \le 1.$$

By virtue of the mean value theorem, there exist

$$\bar{\theta}_1^* = I_p + \theta^*(B_n - I_p), \qquad \bar{\theta}_2^* = \mu + \theta^*(\mu_n - \mu)$$

such that

$$\begin{aligned} G(1) - G(0) &= \frac{\partial G(\theta^*)}{\partial \theta} \\ &= n^{-1} \sum_{i=1}^{n} 2l_j'(2\Psi(\alpha^\tau \bar{\theta}_1^*(x_i - \bar{\theta}_2^*)) - 1)\psi(\alpha^\tau \bar{\theta}_1^*(x_i - \bar{\theta}_2^*)) \\ &\quad \cdot (\alpha^\tau V_n(x_i - \bar{\theta}_2^*) - \alpha^\tau \bar{\theta}_1^* \eta_n). \end{aligned}$$

Combining this with Lemma 4.9, we have

$$\begin{aligned} G(1) &= G(0) + 2\alpha^\tau V_n E l_j(g(\alpha, z(x)))\psi(\alpha^\tau z(x)) z(x) \\ &\quad - 2\alpha^\tau \eta_n E l_j'(g(\alpha, z(x)))\psi(\alpha^\tau z(x)) + o_p(1), \\ &= \sqrt{n}(P_n - P) l_j(g(\alpha, z(\cdot))) + \alpha^\tau V_n \beta_{j_2}(\alpha) - \alpha^\tau \eta_n \beta_{j_1}(\alpha) + o_p(1), \end{aligned}$$

where $o_p(1)$ is uniform for $\|\alpha\| = 1$. Invoke the proof of Proposition 2.1 we have

$$G(1) = \sqrt{n}(P_n - P)(f_j(\alpha, \cdot)) + o_p(1).$$

Hence,

$$n \sup_{\|\alpha\|=1} \hat{I}_J(\alpha) = \frac{1}{2} \sup_{\|\alpha\|=1} \sum_{j=1}^{J} (2j+1)(h_{n_j}(\alpha))^2 + o_p(1).$$

Proof of Lemma 4.8. The assertion follows directly from Lemma 4.7 and the continuous mapping theorem (Pollard (1984), p. 70).

Acknowledgements

The author thanks the referees for the very helpful comments.

References

Anderson, T. W. (1963). Asymptotic theory for principal component analysis, *Ann. Math. Statist.*, **34**, 122–148.

Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices, *Ann. Statist.*, **15**, 1269–1292.

Fang, K. T. and Zhang, S. T. (1993). *The Generalized Multivariate Analysis*, The Science Publishing House, Beijing.

Friedman, J. H. (1987). Exploratory projection pursuit, *J. Amer. Statist. Assoc.*, **82**, 249–266.

Huber, P. J. (1985). Projection pursuit (with discussions), *Ann. Statist.*, **13**, 435–475.

Jones, M. C. and Sibson, R. (1983). What is projection pursuit? (with discussions), *J. Roy. Statist. Soc. Ser. A*, **150**, 1–36.

Li, G. and Zhang, J. (1993). Sphering and its properties, Tech. Report, Institute of Systems Science, Academia Sinica, Beijing.

Maronna, R. A. (1976). Robust M-estimator of multivariate location and scatter, *Ann. Statist.*, **4**, 51–67.

Okamoto, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample, *Ann. Statist.*, **1**, 763–765.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Sun, J. (1989). P-values in projection pursuit, Tech. Report, No. 104, Department of Statistics, Stanford University.

Tyler, D. E. (1982). Radial estimates and the test for sphericity, *Biometrika*, **69**, 429–436.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter, *Ann. Statist.*, **15**, 234–251.

Tukey, P. A. and Tukey, J. W. (1981). Graphical display of data in three and higher dimensions, *Interpreting Multivariate Data* (ed. V. Barnet), 195–198, Wiley, New York.

Va der Waerden, B. L. (1949). *Modern Algebra I*, Ungar, New York.

Zhang, J. (1993*a*). The asymptotics of two exploratory projection pursuit indices, *Chinese J. Appl. Probab. Statist.*, **9**, 18–25.

Zhang, J. (1993*b*). The maximal affine invariant sphering, Tech. Report, Institute of Systems Science, Academia Sinica, Beijing.