

AD-A120 739

ROBUST PROJECTION PURSUIT ESTIMATOR FOR DISPERSION
MATRICES AND PRINCIPAL COMPONENTS(U) HARVARD UNIV
CAMBRIDGE MASS DEPT OF STATISTICS G LI ET AL. NOV 81
PUH-11 N00014-79-C-0512

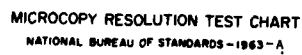
1/1

UNCLASSIFIED

5/2 12/1

ALL

END
PAGE
1
12/1
12/1



ADA 120739

RESEARCH REPORT
RPN - 11

ROBUST PROJECTION PURSUIT ESTIMATOR
FOR DISPERSION MATRICES AND PRINCIPAL COMPONENTS

G. Li and E. Chen*

Department of Statistics, Harvard University

ROBUST PROJECTION PURSUIT ESTIMATOR
FOR DISPERSION MATRICES AND PRINCIPAL COMPONENTS

Guoying Li and Shonglian Chen
November 1981

This work was facilitated in part by National Science Foundation Grant
NSF-79-08485 and Office of Naval Research Contract N00014-79-C-0512.

Department of Statistics
Harvard University
Cambridge, Massachusetts

OCT 26 1982

A

DTIC FILE COPY

This document has been approved
for public release and sale; its
distribution is unlimited.

82 10 26 033

*The authors are on leave from the Institute of Systems Science,
Academia Sinica, Beijing, People's Republic of China.

→ This paper proposes and discusses the ROBUST PROJECTION PURSUIT
ESTIMATOR for dispersion matrices and their principal components. This
estimator finds robust principal components by searching, successively,
for directions which maximize (minimize) a robust estimate of scale; the
estimate of the dispersion matrix is constructed from the estimated
principal components.

These estimators are shown below (under mild conditions) to have a
number of desirable properties. They are orthogonally equivariant and,
within any elliptic underlying density family, asymptotically affinely
equivariant. Furthermore, at elliptic densities, they are consistent and
weakly continuous (i.e., qualitatively robust). Finally, they have good
quantitative robustness - their breakdown point can be as high as $1/2$.

The robust projection pursuit approach is a promising alternative
to other estimators of dispersion matrices. ←

ROBUST PROJECTION PURSUIT ESTIMATOR
FOR DISPERSION MATRICES AND PRINCIPAL COMPONENTS

G. Li and Z. Chen

1. Introduction

Covariance and correlation matrices and their principal components are basic elements in multivariate data analysis. Yet, the classical estimates of these quantities are very sensitive to outliers. In recent years, several robust alternatives have been proposed and studied (see [1], [3], [4], [8], [9], [10], [14], [15], and [16]). The problem seems to be that no one is really satisfactory. The matrix element methods are not orthogonally equivariant, do not even guarantee to produce a positive definite matrix, and are very difficult to study theoretically; while affinely equivariant M-estimates of covariance matrix have quite poor breakdown properties in high dimension.

This paper, together with [2], proposes and discusses a new type of estimator for covariance/correlation matrices and their principal components. It uses a Projection Pursuit (PP) procedure (see [2], [5], [6], [7], [11], [12]) with a robust estimate of scale as the projection index. So we call it the ROBUST PP-ESTIMATOR. Its idea was originally raised by Huber (see [10], p. 200 and pp. 203-204).

The PP procedure deals with high dimensional data: it searches low-dimensional projections which minimize (minimize) an objective function called PROJECTION INDEX.

Suppose that $X \sim P(\Sigma)$ is a p -dimensional random vector and that its location is known in advance and fixed at 0. Our robust PP-estimator works as follows: Let $S(\cdot)$ be a robust estimator for scale, which is usually weakly continuous, and $a \in \mathbb{R}^p$ be a vector. Denote the distribution function of $a^T X$ by F_a^T , or $F(a^T X)$ when needed. The estimators for principal components, in terms of functionals and denoted by $s_i(P)$ and $\lambda_i(P)$ ($1 \leq i \leq p$), are defined, either using the maximising procedure:

$$s_1(P) = \max_{|a|=1} S(F_a^T)$$

$$\lambda_1(P) = s_1 \quad \text{where} \quad |a_1| = 1, \quad S(F_{a_1}^T) = s_1(P),$$

$$s_2(P) = \max_{|a|=1, a \perp a_1} S(F_a^T),$$

$$\lambda_2(P) = s_2 \quad \text{where} \quad |a_2| = 1, \quad a_2 \perp a_1, \quad S(F_{a_2}^T) = s_2(P),$$

$$\dots \dots \dots (1.1)$$

$$s_{p-1}(P) = \max_{|a|=1, a \perp a_1, \dots, a_{p-2}} S(F_a^T),$$

$$\lambda_{p-1}(P) = s_{p-1} \quad \text{where} \quad |a_{p-1}| = 1, \quad a_{p-1} \perp a_1, \dots, a_{p-2}, \quad S(F_{a_{p-1}}^T) = s_{p-1}(P),$$

$$s_p(P) = S(F_{a_p}^T) \quad \text{where} \quad |a_p| = 1, \quad a_p \perp a_1, \dots, a_{p-1},$$

$$\lambda_p(P) = s_p,$$

or using the minimising procedure:



DISTRIBUTION/AVAILABILITY STATEMENTS UNCLASSIFIED SPECIAL	AUTHORITY DATE BY FOR ACTION
---	--

$$S_p(F) = \min_{|a|=1} S(F^a),$$

$$A_p(F) = a_p \quad \text{where} \quad |a_p| = 1, \quad S(F^{a_p}) = S_p(F),$$

.....

$$S_2(F) = \min_{\substack{|a|=1 \\ a_1, \dots, a_3}} \quad (1.1')$$

$$A_2(F) = a_2 \quad \text{where} \quad |a_2| = 1, \quad a_2 \perp a_p, \dots, a_3, \quad S(F^{a_2}) = S_2(F),$$

$$A_1(F) = a_1(F), \quad \text{where} \quad |a_1| = 1, \quad a_1 \perp a_p, \dots, a_2,$$

$$A_1(F) = a_1.$$

Then the estimate for the covariance matrix, denoted by $\xi(F)$, is defined by

$$\xi(F) = \sum_{i=1}^p a_i(F)^2 A_i(F) A_i(F)^T. \quad (1.2)$$

Notice that here we can use "max" ("min") instead of "sup" ("inf") since we assume that $S(\cdot)$ is weakly continuous, thus for any F , $S(F^a)$ is a continuous function of a (see Lemma 4.4 (1) below), so there always exist a_i ($1 \leq i \leq p$) which reach the maximum (minimum) values in (1.1) [(1.1')] on the corresponding regions, respectively.

Obviously,

$$A_1(F) = a_1 \Rightarrow A_1(F) = -a_1 \quad (1 \leq i \leq p). \quad (1.3)$$

So instead of $A_1(F)$ we often consider $A_1(F)A_1(F)^T$. Moreover, from (1.1) [(1.1')] one can see that $A_1(F)A_1(F)^T$, hence $\xi(F)$, may not be uniquely determined, so we account every possible version of $A_1(F)A_1(F)^T$ and $\xi(F)$.

Simulation results in [2] show that these robust M-estimators compare favorably with the best robust estimators which have been proposed before. The robust M-estimators have as good precision (measured by mean squared errors) as these best ones while achieving a higher (empirical) breakdown point.

This paper, as a companion of the simulation study [2], presents some theoretical results. Since weak continuity of $S(\cdot)$ is required, Section 2 shows that the M-estimates, in particular Huber's estimates, of scale are weakly continuous. Then in Section 3, we prove that the M-estimator is orthogonally equivariant, and that at an elliptic underlying distribution it is also asymptotically efficiently equivariant. Section 4 is devoted to the consistency at any distribution belonging to an elliptic probability density family. Qualitative and quantitative robustness are discussed in Section 5. It is shown that the robust M-estimates are weakly continuous at elliptic densities and their breakdown point can be as high as 1/2. Finally, in Section 6, we make some relevant comments.

Since, for both the maximizing and minimizing procedures, the proofs of equivariance, consistency and the weak continuity are all the same, in Sections 3 and 4 and in the first part of Section 5 we discuss only in the case of the maximizing procedure.

2. Further Study of M-Estimators for Scale

It is shown below that the robust M-estimators will have good properties whenever the projection index, an estimator for scale, is weakly continuous and has a high breakdown point. Since M-estimators of scale are both simple and desirable, hence important, estimators, this section discusses their weak continuity and breakdown point.

Let $F(x)$ be one-dimensional distribution function, an M-estimator for scale of F , denoted by $S(F)$, is defined by an implicit equation

$$\int \chi\left(\frac{x}{S(F)}\right) dF(x) = 0.$$

Usually $\chi(t)$ is even. $S(F)$, of course, possesses scale equivariance. Put

$$\lambda(s, F) = \int \chi\left(\frac{x}{s}\right) dF(x) \quad (2.1)$$

$$F(x) = \begin{cases} 0 & x \leq 0 \\ F(x) - F(-x_0) & x > 0. \end{cases} \quad (2.2)$$

So $S(F)$ is a zero point of $\lambda(s, F)$. $F(x)$ is the distribution function of $|x|$ if $x \sim F(x)$.

LEMMA 2.1. Assume that $\chi(t)$ is even and increasing on $t > 0$ and $\lambda(s, F)$ exists for at least one $s \in \mathbb{R}$. Then

- (1) $\lambda(s, F) = \lambda(s, \bar{F})$.
- (2) For any fixed F , $\lambda(s, F)$ is decreasing in s ; for any fixed s , $\lambda(s, F)$ is stochastically increasing in F .
- (3) If $\chi(t)$ is also continuous and bounded then $\lambda(s, F)$ is continuous in s and weakly continuous in F for every s fixed.

Proof. (1) is very easy and (2) is straightforward from, for example, Lemma 2(1) in Lehmann's book (see [13], p. 74).

Now we prove (3). Denote Lévy metric by $d_L(\cdot, \cdot)$. For any small $\epsilon > 0$, let

$$\mathcal{D}_\epsilon(F_0) = \{F : d_L(F, F_0) < \epsilon\}.$$

It is easy to check that

$$F \in \mathcal{D}_{\epsilon/2}(F_0) \Rightarrow F \in \mathcal{D}_\epsilon(F_0).$$

Without losing generality, we assume that there exist x_1, x_2 ($0 < x_1, x_2 < \infty$) satisfying

$$F_0(x_1) = \epsilon, \quad F_0(x_2) = 1 - \epsilon.$$

Let

$$F_1(x) = \begin{cases} 0 & x \leq x_1 + \epsilon \\ F_0(x - \epsilon) - \epsilon & x > x_1 + \epsilon \\ 1 & x = \infty \end{cases} \quad (2.3)$$

$$F_2(x) = \begin{cases} 0 & x \leq 0 \\ F_0(x + \epsilon) + \epsilon & 0 < x \leq x_2 - \epsilon \\ 1 & x > x_2 - \epsilon. \end{cases}$$

Then for any $F \in \mathcal{D}_{\epsilon/2}$, from (2) it follows that

$$\begin{aligned} \lambda(s, F) - \lambda(s, F_0) &\leq \lambda(s, F_1) - \int_{x_1 + \epsilon}^{\infty} \chi\left(\frac{t}{s}\right) dF_0(t - \epsilon) + \epsilon \chi\left(\frac{x_1 + \epsilon}{s}\right) \\ &= \int_{x_1}^{\infty} \chi\left(\frac{t}{s}\right) dF_0(t) + \epsilon \chi\left(\frac{x_1 + \epsilon}{s}\right) - \int_{x_1}^{\infty} \chi\left(\frac{t}{s}\right) dF_0(t) \\ &= \int_{x_1}^{\infty} \chi\left(\frac{t}{s}\right) dF_0(t) + \epsilon \chi\left(\frac{x_1 + \epsilon}{s}\right) < \int_{x_1}^{\infty} \chi\left(\frac{t}{s}\right) dF_0(t) + \epsilon (\chi\left(\frac{x_1 + \epsilon}{s}\right) - \chi(0)). \end{aligned}$$

$$\begin{aligned}
\lambda(u, P) - \lambda(u, P_0) &= \lambda(u, P_2) - \lambda(u, P_0) = \int_0^{P_2 - P_0} \chi\left(\frac{t}{u}\right) dP_0(t) + (P_2 - P_0)\chi(0) \\
&= \int \chi\left(\frac{t-u}{u}\right) dP_0(t) - \int_0^u \chi\left(\frac{t-u}{u}\right) dP_0(t) - \int_{P_2}^u \chi\left(\frac{t-u}{u}\right) dP_0(t) + (P_2 - P_0)\chi(0) \\
&> \int \chi\left(\frac{t-u}{u}\right) dP_0(t) - P_0(u)\left(\chi\left(\frac{t}{u}\right) - \chi(0)\right) - \varepsilon(\chi(u) - \chi(0)).
\end{aligned}$$

Thus for any $0 < u < \varepsilon$,

$$\begin{aligned}
|\lambda(u, P) - \lambda(u, P_0)| &< \lambda(u, P_1) - \lambda(u, P_2) = \\
&< \int \left[\chi\left(\frac{t-u}{u}\right) - \chi\left(\frac{t-u}{u}\right) \right] dP_0(t) + 2\varepsilon(\chi(u) - \chi(0)) + P_0(u)\left(\chi\left(\frac{t}{u}\right) - \chi(0)\right) \\
&\rightarrow 0 \quad (u \rightarrow 0).
\end{aligned}$$

THEOREM 2.1. Assume that $\chi(t)$ is even, continuous, bounded and increasing on $t > 0$ and that $0 < S(P_0) < \infty$ is uniquely defined. Then $S(P)$ is weakly continuous at P_0 .

Proof. Assume that $P_n \xrightarrow{w} P_0$. Put

$$u_i = S(P_i) \quad i = 0, 1, 2, \dots$$

Then u_0 is the only zero point of $\lambda(u, P_0)$ and there exist u' and u'' ($u' < u_0 < u''$) such that, from the monotonicity of $\lambda(\cdot, P_0)$ and uniqueness of $S(P_0)$,

$$\lambda(u'', P_0) < 0 < \lambda(u', P_0).$$

From Lemma 2.1 it follows that for any $\varepsilon > 0$, there exist $t_0 = u' < t_1 < \dots < t_n = u''$, and $n > 0$, such that

$$\begin{aligned}
0 &< \lambda(t_{i-1}, P_0) - \lambda(t_i, P_0) < \frac{\varepsilon}{2} \quad i = 1, \dots, n \\
|\lambda(t_1, P_n) - \lambda(t_1, P_0)| &< \frac{\varepsilon}{2} \quad n \geq n, i = 0, \dots, n.
\end{aligned} \tag{2.4}$$

and also

$$\lambda(u'', P_n) < 0 < \lambda(u', P_n) \quad n \geq n. \tag{2.5}$$

Hence, for any $u, u' < u < u''$, there exists an i , such that

$$t_{i-1} < u < t_i.$$

Then (2.4) gives that

$$\begin{aligned}
\lambda(u, P_n) - \lambda(u, P_0) &< \lambda(t_{i-1}, P_n) - \lambda(t_{i-1}, P_0) + \lambda(t_i, P_n) - \lambda(u, P_0) \\
&< \frac{\varepsilon}{2} + |\lambda(t_{i-1}, P_n) - \lambda(t_{i-1}, P_0)| < \varepsilon \\
\lambda(u, P_n) - \lambda(u, P_0) &> \lambda(t_1, P_n) - \lambda(t_1, P_0) + \lambda(t_1, P_n) - \lambda(u, P_0) \\
&> -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} = -\varepsilon.
\end{aligned}$$

That is, $\{\lambda(u, P) | u' < u < u''\}$ is weakly equicontinuous at P_0 .

Now we claim that $u_n \rightarrow u_0$ ($n \rightarrow \infty$). If not, because of (2.5), we have

$$u' < u_n < u''.$$

Thus, there must be a subsequence of u_n , say u_{n_k} , such that

$$u_{n_k} \rightarrow \bar{u} \neq u_0.$$

On the other hand, because of the weak equicontinuity of $\lambda(u, P)$ at P_0 ,

$$\lambda(\bar{u}, P_0) = \lim_{n \rightarrow \infty} \lambda(u_{n_k}, P_0) = \lim_{n \rightarrow \infty} [\lambda(u_{n_k}, P_0) - \lambda(u_{n_k}, P_{n_k})] = 0.$$

This contradicts with the uniqueness of $S(P_0)$. Hence

$$S(P_n) = a_n \rightarrow S(P_0) \quad (n \rightarrow \infty) \quad \square$$

Now we consider Huber's M-estimates of scale. That is the choice

$$\chi(t) = \begin{cases} t^2 - \beta & |t| \leq h \\ h^2 - \beta & |t| > h \end{cases} \quad (2.6)$$

with $h > 0$ and

$$\beta = \int_{|t| \leq h} t^2 d\theta + h^2 \int_{|t| > h} d\theta \quad (2.7)$$

where $\theta(t)$ is the standard Normal distribution function. Obviously,

$0 < \beta < h^2$ and $\chi(t)$ is even, bounded, continuous and increasing on $t > 0$.

So it satisfies every condition in Lemma 2.1. Furthermore, we have

THEOREM 2.2. Assume that $V(0) = F(0_+) - F(0) < 1 - \frac{\beta}{h^2}$, then

- (1) Huber's $S(P)$ is uniquely defined and $0 < S(P) < \infty$;
- (2) $S(\cdot)$ is weakly continuous at P .

Proof. Put

$$u_0 = \sup\{u | u > 0, P(u_0) = V(0) = \beta\}.$$

For any $u < \frac{u_0}{h}$,

$$\begin{aligned} \lambda(u, P) &= \int_{|u| < hu} \left(\frac{t}{h}\right)^2 dP(t) + h^2 \int_{|u| > hu} dP(t) - \beta \\ &= \int_{|u| > u_0} h^2 dP(t) - \beta = h^2(1 - V(0)) - \beta > 0. \end{aligned} \quad (2.8)$$

If $\frac{u_0}{h} < a_1 < a_2$, then, notice that $P(u_0 < |u| < a_1 h) > 0$.

$$\begin{aligned} \lambda(a_2, P) &= \int_{u_0 < |u| < a_1 h} \left(\frac{t}{a_2}\right)^2 dP(t) \\ &\quad + \int_{a_1 h < |u| < a_2 h} \left(\frac{t}{a_2}\right)^2 dP(t) + h^2 \int_{|u| > a_2 h} dP(t) - \beta \\ &< \int_{u_0 < |u| < a_1 h} \left(\frac{t}{a_1}\right)^2 dP(t) + \int_{|u| > a_1 h} h^2 dP(t) - \beta \\ &= \lambda(a_1, P). \end{aligned} \quad (2.9)$$

It is clear that

$$\lim_{u \rightarrow \infty} \lambda(u, P) = -\beta < 0. \quad (2.10)$$

Now (1) follows immediately from (2.8)-(2.10). And (2) is a necessary consequence of (1) and Theorem 2.1. $\square$

The breakdown point for M-estimates of scale has been worked out in the usual sense (see [10], p. 110). But we need to add something else to serve our purpose. For completeness, we discuss it from the beginning.

Suppose that $\chi(t)$ is even and increasing on $t > 0$. Then Lemma 2.1 holds. Put

$$\begin{aligned} 0 \chi_1 &= \chi(\infty) - \chi(0), \\ \lambda(\infty, P) &= \lim_{u \rightarrow \infty} \lambda(u, P). \end{aligned} \quad (2.11)$$

Consider the gross error model

$$P = (1-\epsilon)P_0 + \epsilon H.$$

Since " $S(F) < \infty$ for any H " implies that $\lambda(\infty, F) < 0$ for any H , we have

$$\lambda(\infty, (1-c)P_0 + c\delta_H) = (1-c)\lambda(0) + c\lambda(\infty) < 0,$$

where δ_H is a probability measure putting a pointmass at H . This is equivalent to

$$c < -\frac{\lambda(0)}{\lambda(\infty)} = c_1. \quad (2.12)$$

It is easy to show that if $|\lambda| < \infty$, then

$$c < -\frac{\lambda(0)}{\lambda(\infty)} \Rightarrow S(F) < \infty \text{ for any } H.$$

Conversely, if $c > -\frac{\lambda(0)}{\lambda(\infty)}$, then

$$H = \delta_H \Rightarrow \lambda(\infty, F) > 0 \Rightarrow S(F) = \infty.$$

Sometime we hope that $S(F) > 0$ when F_0 is not degenerate (into 0). In this case, if $S(F) > 0$ for any H , then

$$\lambda(0, (1-c)P_0 + c\delta_H) - (1-c)[P_0(0)\lambda(0) + (1-P_0(0))\lambda(\infty)] + c\lambda(0) > 0,$$

i.e.,

$$c < \frac{\lambda(\infty)}{\lambda(0)} + \frac{P_0(0)\lambda(0)}{(1-P_0(0))\lambda(\infty)} = c_2. \quad (2.13)$$

Also

$$c < c_2 \Rightarrow \lambda(0, (1-c)P_0 + c\delta_H) > 0 \Rightarrow S(F) > 0 \text{ for any } H.$$

On the other hand,

$$c > c_2 \Rightarrow \lambda(0, (1-c)P_0 + c\delta_H) < 0 \Rightarrow S((1-c)P_0 + c\delta_H) = 0.$$

As we will see in the end of Section 5, it is important that $S(F)$ does not misbehave as a projection index for principal component estimation. That

means, if F_0 is degenerate (into 0) then $S(F) = 0$ for any H . This implies that

$$\lambda(0, (1-c)P_0 + c\delta_H) = (1-c)\lambda(0) + c\lambda(\infty) < 0, \quad (2.14)$$

that is,

$$c < -\frac{\lambda(0)}{\lambda(\infty)} = c_3 (= c_1). \quad (2.15)$$

Also

$$c < c_3 \Rightarrow \lambda(0, F) < 0 \text{ for any } H \Rightarrow S(F) = 0 \text{ for any } H.$$

Conversely, if

$$c > c_3$$

then

$$\lambda(0, (1-c)P_0 + c\delta_H) = \lambda(0, (1-c)\delta_0 + c\delta_H) > 0 \Rightarrow S((1-c)\delta_0 + c\delta_H) > 0.$$

As Huber pointed out (see [10], p. 110) that we can usually disregard the second contingency, so we conclude that for the c -contamination model, the breakdown point of H -estimates for scale is

$$c^* = -\frac{\lambda(0)}{\lambda(\infty)}.$$

In the case of Huber's choice,

$$c^* = \frac{1}{2}.$$

Now we see obviously that for any even $\lambda(t)$, the breakdown point of an H -estimator for scale is as high as $1/2$.

3. Equivariance

The study of robust estimators of covariance poses a problem: How to provide a covariance estimator which behaves under coordinate changes the same way as the classical estimator does. It is important to establish equivariance properties of a new estimator in order to show that it is, truly, an estimator of dispersion. This section shows that the robust PP-estimators do have equivariance properties.

As it is mentioned in Section 1 that this paper considers the pure dispersion problem, i.e., assume that location is known and fixed at 0. Also, we assume throughout that $S(\cdot)$ is weakly continuous.

Recalling that $A_1(P)$ and $\hat{\Sigma}$ are not uniquely determined, we use them to denote any version of those quantities.

THEOREM 3.1. Robust PP-estimates $S_1(P)$, $A_1(P)$ ($i = 1, 2, \dots, p$) and $\hat{\Sigma}(P)$ are all orthogonal equivariant. I.e., let P be any orthogonal matrix, $\underline{x} \sim P_1(\underline{z})$ and $\underline{y} \sim P_2(\underline{z})$; then

$$\begin{aligned} S_1(P_2) &= S_1(P_1) \\ A_1(P_2) &= PA_1(P_1) \\ \hat{\Sigma}(P_2) &= P\hat{\Sigma}(P_1)P^T \quad i = 1, 2, \dots, p. \end{aligned}$$

Proof. Since

$$\begin{aligned} S_1(P_1) &= \max_{\|u\|=1} S(\mathcal{E}(u^T \underline{x})) = \max_{\|u\|=1} S(\mathcal{E}((Pu)^T \underline{y})) \\ &= \max_{\|v\|=1} S(\mathcal{E}(v^T \underline{y})) = S_1(P_2). \end{aligned} \quad (3.1)$$

Thus

$$S_1(P_2) = S_1(P_1) = S(\mathcal{E}(A_1(P_1)^T \underline{x})) = S(\mathcal{E}((PA_1(P_1))^T \underline{y})),$$

i.e.,

$$A_1(P_2) = PA_1(P_1). \quad (3.2)$$

From (3.1) and (3.2), it follows that

$$\begin{aligned} S_2(P_1) &= \max_{\|u\|=1} S(\mathcal{E}(u^T \underline{x})) = \max_{\|u\|=1} S(\mathcal{E}((Pu)^T \underline{y})) \\ &= \max_{\|v\|=1} S(\mathcal{E}(v^T \underline{y})) = S_2(P_2). \end{aligned}$$

$$S_2(P_2) = S_2(P_1) = S(\mathcal{E}(A_2(P_1)^T \underline{x})) = S(\mathcal{E}((PA_2(P_1))^T \underline{y})).$$

$$\Rightarrow A_2(P_2) = PA_2(P_1).$$

And so on, we have

$$\begin{aligned} S_i(P_2) &= S_i(P_1) \\ A_i(P_2) &= PA_i(P_1) \quad i = 1, 2, \dots, p. \end{aligned}$$

Therefore,

$$\begin{aligned} \hat{\Sigma}(P_2) &= \frac{1}{p} \sum_{i=1}^p S_i(P_2) A_i(P_2) A_i(P_2)^T \\ &= \frac{1}{p} \sum_{i=1}^p S_i(P_1) A_i(P_1) A_i(P_1)^T P^T \\ &= P \hat{\Sigma}(P_1) P^T. \end{aligned}$$

Notice that a random sample $\underline{x}_1, \dots, \underline{x}_n$ and the associated empirical distribution $P_n(\underline{z}) = \frac{1}{n} \sum_{i=1}^n \delta_{\underline{x}_i}(\underline{z})$ change, under a coordinate transformation

P , into $PX_1, \dots, PX_n$ and

$$G_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{PX_i}(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(P^{-1}x) = P_n(P^{-1}x),$$

respectively, and that

$$x \sim P_n(x) \Rightarrow PX \sim P_n(P^{-1}x).$$

So from Theorem 3.1, it follows that as an estimator from sample

$$\begin{aligned} S_1(x_1, \dots, x_n) &= S_1(P_n) & i = 1, \dots, p, \\ A_1(x_1, \dots, x_n) &= A_1(P_n) & i = 1, \dots, p, \\ \hat{g}(x_1, \dots, x_n) &= \hat{g}(P_n) \end{aligned}$$

are indeed orthogonal equivariant, just like classical estimators.

Suppose that $F(x, V)$ belongs to a p -dimension elliptic probability density family, i.e., $F(x, V)$ has a density $f(x, V)$

$$f(x, V) = (\det V^{-1}) f_0(V^{-1}x), \quad (3.3)$$

where V is a nonsingular $p \times p$ matrix and $f_0(x)$ is spherically symmetric (and nondegenerate, of course), i.e.,

$$f_0(x) = f_0(|x|). \quad (3.4)$$

To prove the affine equivariance within an elliptic density family, we need

LEMMA 3.1. Assume that p -dimensional random vector X has a spherically symmetric probability density $f_0(x) = f_0(x_1, \dots, x_p) = f_0(|x|)$. Let

$$q(x) = \int f_0(x, x_2, \dots, x_p) dx_2 dx_3 \dots dx_p \quad (3.5)$$

be the marginal density. Then $q(x) \sim q(x)$ for any unit vector q .

Proof. For any unit q fixed, let P be an orthogonal matrix such that

$$P = \begin{pmatrix} q \\ \vdots \end{pmatrix},$$

and put $y = Px = (y_1, \dots, y_p)^T$.

Since y has a probability density

$$f_y(y) = (\det P^{-1}) f_0(P^{-1}y) = f_0(|y^{-1}y|) = f_0(y),$$

it is immediate that

$$q^T x = y_1 \sim q(x).$$

Let $F_0(x)$ be probability distribution function of $f_0(x)$ and let $G(x)$ be $q(x)$'s (see (3.5)). From Lemma 3.1, it is straightforward that

$$\begin{aligned} S_1(P_n) &= \min_{|a|=1} S(S^T a^T x) = S(q) & i = 1, 2, \dots, p, \\ &= \int A_1(P_n), \dots, A_{p-1}(P_n) \end{aligned}$$

and any set of orthonormal vectors $a_1, \dots, a_p$ can be $A_1(P_n), \dots, A_p(P_n)$. Without losing generality, we assume that

$$S(q) = 1. \quad (3.6)$$

Then,

$$\hat{g}(P_n) = \frac{1}{n} \sum_{i=1}^n A_1(P_n)^T A_1(P_n) A_1(P_n)^T = S(q) = 1. \quad (3.7)$$

Also, Lemma 3.1, together with the scale equivariance of $S(\cdot)$, yields that for any $q \in R^p$, $q \neq 0$,

$$S(S^T q^T x) = S\left(\frac{q^T x}{|q|}\right) = |q|. \quad (3.8)$$

hence, for any number of the elliptic family $F_{\underline{V}}(\underline{x}) = F(\underline{x}, \underline{V})$ [see (3.3) and (3.4)], put

$$\begin{aligned} \underline{Z} &= \underline{V} \underline{Z}, \\ \underline{Z} &= \underline{V} \underline{V}^T, \end{aligned} \quad (3.9)$$

$\underline{Z}$ is usually called the pseudo-covariance matrix of $\underline{Z}$. Let $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ be the eigenvalues of $\underline{Z}$ and $\underline{a}_1, \dots, \underline{a}_p$ be any set of eigenvectors corresponding to $\lambda_1, \dots, \lambda_p$, respectively. (3.8) gives us that for any $\underline{a} \in \mathbb{R}^p$,

$$s(\underline{r}_{\underline{V}}^2) = s(\underline{Z}(\underline{a}^T \underline{a})) = s(\underline{Z}(\underline{V}^T \underline{a})^T \underline{a}) = |\underline{V}^T \underline{a}|^2.$$

Therefore, simple algebra gives

$$\begin{aligned} s_1(\underline{r}_{\underline{V}}) &= \max_{|\underline{a}|=1} s(\underline{r}_{\underline{V}}^2) = \max_{|\underline{a}|=1} |\underline{V}^T \underline{a}|^2 = |\underline{V}^T \underline{a}_1|^2 = \lambda_1, \\ \underline{a}_1(\underline{r}_{\underline{V}}) &= \underline{a}_1, \\ s_2(\underline{r}_{\underline{V}}) &= \max_{\substack{|\underline{a}|=1 \\ \underline{a} \perp \underline{a}_1}} s(\underline{r}_{\underline{V}}^2) = |\underline{V}^T \underline{a}_2|^2 = \lambda_2, \\ \underline{a}_2(\underline{r}_{\underline{V}}) &= \underline{a}_2, \\ &\dots \dots \dots \\ s_p(\underline{r}_{\underline{V}}) &= \lambda_p, \\ \underline{a}_p(\underline{r}_{\underline{V}}) &= \underline{a}_p. \end{aligned} \quad (3.10)$$

Consequently,

$$\underline{Z}(\underline{r}_{\underline{V}}) = \sum_{i=1}^p s_i^2(\underline{r}_{\underline{V}}) \underline{a}_i(\underline{r}_{\underline{V}}) \underline{a}_i(\underline{r}_{\underline{V}})^T = \sum_{i=1}^p \lambda_i \underline{a}_i \underline{a}_i^T = \underline{Z}.$$

So far, we have proved

THEOREM 3.2. Within any elliptic probability density family, robust M-estimator for covariance matrix $\underline{Z}(\underline{r})$, as a functional, is affinely equivariant. In detail, assume that $\underline{X} \sim F(\underline{x})$ belongs to an elliptic density family and that $\underline{V}$ is any nonsingular $p \times p$ matrix. Denote the distribution of $\underline{V} \underline{X}$ by $G(\underline{x})$. Then

$$\underline{Z}(\underline{x}) = \underline{V} \underline{Z}(\underline{r}) \underline{V}^T.$$

Under an affine transformation $\underline{V}$, an empirical distribution $F_n(\underline{x}) = \frac{1}{n} \sum_{i=1}^n \delta_{\underline{x}_i}(\underline{x})$ transforms into $G_n(\underline{x}) = \frac{1}{n} \sum_{i=1}^n \delta_{\underline{V} \underline{x}_i}(\underline{x}) = \frac{1}{n} \sum_{i=1}^n \delta_{\underline{x}_i}(\underline{V}^{-1} \underline{x})$. Since $F_n(\underline{x})$ is generally not generated by a spherical distribution, $F_n(\underline{x})$ and $G_n(\underline{x})$ do not belong to any elliptic distribution family. Therefore, as an estimate from sample, $\underline{Z}(\underline{x}_1, \dots, \underline{x}_n) = \underline{Z}(\underline{r}_n)$ is not affinely equivariant. However, if the underlying distribution of F_n is a member of an elliptic family, then $\underline{Z}(\underline{x}_1, \dots, \underline{x}_n) = \underline{Z}(\underline{r}_n)$ is asymptotically affinely equivariant. Actually, suppose $\underline{x}_1, \dots, \underline{x}_n$ come from an elliptic distribution $F(\underline{x}, \underline{V}_1)$, according to Theorem 4.4,

$$\begin{aligned} \underline{Z}(\underline{x}_1, \dots, \underline{x}_n) &= \underline{Z}(\underline{r}_n) \rightarrow \underline{Z}_1 = \underline{Z}(\underline{r}(\underline{x}_1, \underline{V}_1)) \quad \text{P. a. s.} \\ \underline{Z}(\underline{V} \underline{x}_1, \dots, \underline{V} \underline{x}_n) &= \underline{Z}(\underline{r}_n) \rightarrow \underline{Z}_2 = \underline{Z}(\underline{r}(\underline{x}_1, \underline{V}_2)) \quad \text{P. a. s.} \end{aligned}$$

then, from Theorem 3.2, it follows that

$$\underline{Z}_2 = \underline{V} \underline{Z}_1 \underline{V}^T.$$

4. Consistency at Elliptic Probability Density Family

When the data come from an elliptic probability distribution, for which pseudocovariance has an intuitive interpretation — the shape of the underlying ellipses, it is possible to show that the robust PP-estimators give consistent estimates. The idea of the proof is simple: it combines the continuity property of the projection index $S(\cdot)$ (discussed in Section 3 for M-estimates) with a compactness argument. But putting this into operation in high dimension and for any underlying pseudo-covariance is quite complicated. To cope with this problem, we introduce some lemmas first, then start with some special underlying pseudo-covariances to achieve general consistency results.

LEMMA 4.1. Assume that

(1) Ω is a compact set in R^p , $J_n(a)$ ($n = 0, 1, 2, \dots$) are continuous on Ω .

(2) $\lim_{n \rightarrow \infty} J_n(a) = J_0(a)$ uniformly in a on Ω .

(3) For $\Omega_n \subset \Omega$ ($n = 0, 1, 2, \dots$), there exist a_n such that

$$J_n(a_n) = \max_{a \in \Omega_n} J_n(a) \quad n = 0, 1, 2, \dots$$

(4) There also exist $p \times p$ orthogonal matrices P_n such that

$$\Omega_n = P_n \Omega_0 = \{P_n a \mid a \in \Omega_0\} \quad n = 1, 2, \dots$$

$$P_n \rightarrow I \quad (n \rightarrow \infty)$$

(here we use $\|A\| = \text{trace}(AA^T)$ as the norm of matrix A).

Then

$$J_n(a_n) \rightarrow J_0(a_0) \quad (n \rightarrow \infty).$$

Proof. Notice that

$$P_n a_0 \in \Omega_n, \quad P_n^{-1} a_n \in \Omega_0, \quad n = 1, 2, \dots$$

From assumption (3) we have

$$J_n(P_n a_0) - J_0(a_0) < J_n(a_n) - J_0(a_0) < J_n(a_n) - J_0(P_n^{-1} a_n).$$

Now we need only show that the two extreme sides of this inequality converge to zero when $n \rightarrow \infty$. Actually, by assumption (4), (1) and (2), it follows that

$$P_n a_0 \rightarrow a_0 \quad (n \rightarrow \infty)$$

$$\begin{aligned} |J_n(P_n a_0) - J_0(a_0)| &< |J_n(P_n a_0) - J_0(P_n a_0)| \\ &+ |J_0(P_n a_0) - J_0(a_0)| \rightarrow 0 \quad (n \rightarrow \infty). \end{aligned}$$

Similarly,

$$|J_n(a_n) - J_0(P_n^{-1} a_n)| \rightarrow 0 \quad (n \rightarrow \infty).$$

LEMMA 4.2. Under the conditions of Lemma 4.1, if a_0 is the unique maximum of $J_0(a)$ on $\bar{\Omega}_0$, the closure of Ω_0 , then

$$a_n \rightarrow a_0 \quad n \rightarrow \infty.$$

Proof. Put

$$\tilde{a}_n = P_n^{-1} a_n.$$

By the assumption (4) in Lemma 4.1, we know that

$$\begin{aligned} \bar{a}_n &\in \bar{a}_0 \\ \lim_{n \rightarrow \infty} (\bar{a}_n - \bar{a}_0) &= 0. \end{aligned}$$

Assume that

$$\bar{a}_n \rightarrow \bar{a}_0 \quad n \rightarrow \infty.$$

Since $\bar{a}_n \in \bar{a}_0 \subset \Omega$ and Ω is compact, there must be a subsequence of $\bar{a}_n, \bar{a}_{n'}$, such that $\lim_{n' \rightarrow \infty} \bar{a}_{n'} = \bar{a}$. Thus $\lim_{n' \rightarrow \infty} \bar{a}_{n'} = \bar{a}$ and $\bar{a} \in \bar{a}_0$. Then assumptions (1)-(3) give

$$\mathcal{J}_0(\bar{a}) = \lim_{n' \rightarrow \infty} \mathcal{J}_0(\bar{a}_{n'}) = \lim_{n' \rightarrow \infty} \mathcal{J}_n(\bar{a}_{n'}) \geq \lim_{n' \rightarrow \infty} \mathcal{J}_n(\bar{a}_0) = \mathcal{J}_0(\bar{a}_0).$$

This contradicts that $\bar{a}_0$ is the only maximum of $\mathcal{J}_0(\cdot)$ on $\bar{a}_0$. Hence

$$\bar{a}_n \rightarrow \bar{a}_0 \quad (n \rightarrow \infty).$$

LEMMA 4.1.

(1) Assume $\bar{a}, \bar{a}_1 \in \mathbb{R}^p$, $|\bar{a}| = |\bar{a}_1| = 1$, $\bar{a}_1 \neq \bar{a}$, then there exists a rotation matrix P such that

$$\begin{aligned} \bar{a}_1 &= P\bar{a} \\ \bar{g} &= P\bar{g} \quad \text{for any } \bar{g} \perp \bar{a}, \bar{a}_1 \\ |\bar{g} - \bar{g}| &\leq |\bar{a}_1 - \bar{a}| \quad \forall \bar{g}. \end{aligned} \quad (4.1)$$

(2) Assume that $\bar{a}_n, \bar{a} \in \mathbb{R}^p$, $|\bar{a}_n| = |\bar{a}| = 1$, and that P_n is a rotation matrix as in (1) such that

$$\bar{a}_n = P_n \bar{a} \quad n = 1, 2, \dots$$

Then $P_n \rightarrow I \quad (n \rightarrow \infty) \iff \bar{a}_n \rightarrow \bar{a} \quad (n \rightarrow \infty).$

Proof.

(1) Without losing generality, we assume that $\bar{a}, \bar{a}_1$ are on the same coordinate plane formed by the first two axes. Let θ be the angle between $\bar{a}$ and $\bar{a}_1$, we define a matrix P by

$$P = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & I_{p-2} \end{pmatrix}. \quad (4.2)$$

It is easy to check that P is the rotation matrix we want, i.e., it satisfies (4.1).

(2) Let θ_n be the angle between $\bar{a}$ and $\bar{a}_n$ ($n = 1, 2, \dots$). From (1) we know that

$$P_n = \begin{pmatrix} \cos \theta_n & -\sin \theta_n & 0 \\ \sin \theta_n & \cos \theta_n & 0 \\ 0 & 0 & I_{p-2} \end{pmatrix}.$$

Obviously,

$$\bar{a}_n \rightarrow \bar{a} \iff \theta_n \rightarrow 0 \iff P_n \rightarrow I. \quad (4.3)$$

LEMMA 4.4. Assume that $X \sim F_{\theta}(n)$ is a p -dimension random vector and $F_n(n)$ ($n = 1, 2, \dots$) are the empirical distribution of F_0 , and that H -estimate $\hat{H}(\cdot)$ for scale is weakly continuous. Then

(1) for every n ($n = 0, 1, 2, \dots$) fixed, $\hat{H}_n^2$ is weakly continuous in $\bar{a}$, $\hat{H}(\hat{H}_n^2)$ is continuous in $\bar{a}$.

(2) $\mathcal{W} = \{F_n^a | a \in \mathbb{R}^p, |a| = 1, n = 0, 1, 2, \dots\}$ is a.s. weakly compact;

(3) if $|a_n| = |a_{n-1}| = 1$, $(a_n - a_{n-1}) \rightarrow 0$ ($n \rightarrow \infty$), then

$$S(F_0^{a_n}) - S(F_0^{a_{n-1}}) \rightarrow 0 \quad (n \rightarrow \infty)$$

$$S(F_n^{a_n}) - S(F_n^{a_{n-1}}) \rightarrow 0 \quad (n \rightarrow \infty) \quad \text{P. s. a. s.}$$

$$S(F_n^{a_n}) - S(F_n^{a_{n-1}}) \rightarrow 0 \quad (n \rightarrow \infty) \quad \text{P. s. a. s.}$$

Proof.

(1) For any n ($n = 0, 1, 2, \dots$) fixed, assume random vector $\underline{Y} \sim F_n$; thus, according to our convention, $\underline{a}^T \underline{Y} \sim F_n^a$. If $\underline{a}_t \rightarrow \underline{a}$ ($t \rightarrow \infty$), then

$$\underline{a}_t^T \underline{Y} \rightarrow \underline{a}^T \underline{Y} \quad (t \rightarrow \infty) \quad \text{everywhere.}$$

Thus

$$F_n^{a_t} \rightarrow F_n^a(W) \quad (t \rightarrow \infty).$$

By the weak continuity of $S(\cdot)$, we have

$$S(F_n^{a_t}) \rightarrow S(F_n^a) \quad (t \rightarrow \infty).$$

(2) We have to show that any sequence $\{a_t | t=1, 2, \dots\} \subset \mathcal{W}$ has a subsequence which a.s. converges according to weak topology. Suppose that

$$a_t = F_{n_t}^{a_t} \in \quad t = 1, 2, \dots$$

If there are only finite different n_t in $\{n_t | t=1, 2, \dots\}$, denote these n_t which are the same by m , then choose a subsequence of a_t , namely $a_{t'}$, such

that $a_{t'} \rightarrow a$ ($t' \rightarrow \infty$) and $a_{t'} = m$. From (1) we obtain that

$$G_{t'} = F_{n_t}^{a_{t'}} \rightarrow F_m^a \quad \text{a.s.} \quad (t' \rightarrow \infty).$$

Now assume that $\{n_t | t=1, 2, \dots\}$ has infinitely many different elements. We can choose a subsequence of a_t , say $a_{t'}$, such that

$$a_{t'} \rightarrow \underline{a} \quad \text{and} \quad n_{t'} \rightarrow \infty \quad (t' \rightarrow \infty).$$

Then, let d_a is a metric which measures weak topology, we know

$$d_a(G_{t'}, F_0^a) \leq d_a(F_{n_{t'}}^{a_{t'}}, F_0^{a_{t'}}) + d_a(F_0^{a_{t'}}, F_0^a). \quad (4.4)$$

Because $F_0^{a_{t'}} \rightarrow F_0^a(W)$ ($t' \rightarrow \infty$), the second term of the right-hand side of (4.4) vanishes when $t' \rightarrow \infty$. Denote the Kolmogorov metric by $d_K(\cdot, \cdot)$.

From the Glivenko-Cantelli theorem, i.e.,

$$P[d_K(F_n, W) > 0, n \rightarrow \infty] = 1$$

uniformly in W (F_n are empirical distributions of W), it follows that

$$d_a(F_{n_{t'}}^{a_{t'}}, F_0^{a_{t'}}) \leq d_K(F_{n_{t'}}^{a_{t'}}, F_0^{a_{t'}}) \rightarrow 0 \quad \text{a.s.} \quad (t' \rightarrow \infty). \quad (4.5)$$

Hence, the first term of the right-hand side of (4.4) also converges to zero when $t' \rightarrow \infty$. Thus

$$G_{t'} \rightarrow F_0^a(W) \quad \text{a.s.} \quad (t' \rightarrow \infty).$$

(3) As a result of (1), $S(F_0^a)$ is uniformly continuous on $\{a | |a|=1, a \in \mathbb{R}^p\}$. It follows immediately that

$$|a_n - a_{n-1}| \rightarrow 0 \quad (n \rightarrow \infty) \Rightarrow S(F_0^{a_n}) - S(F_0^{a_{n-1}}) \rightarrow 0 \quad (n \rightarrow \infty).$$

To complete (3), we have only to show that

$$S(F_n^Q) - S(F_0^Q) \rightarrow 0 \quad (n \rightarrow \infty) \quad P. \text{ s. a. s. } .$$

But

$$S(F_n^Q) - S(F_0^Q) = [S(F_n^Q) - S(F_0^Q)] + [S(F_0^Q) - S(F_0^Q)] .$$

We know already that the second term of the right-hand side vanishes when $n \rightarrow \infty$, as for the first term, since $S(\cdot)$ is weakly continuous, $\mathcal{W}$ is a.s. weakly compact, so $S(\cdot)$ is uniformly continuous a.s. on $\mathcal{W}$; for verifying

$$S(F_n^Q) - S(F_0^Q) \rightarrow 0 \quad n \rightarrow \infty \quad P. \text{ s. a. s. }$$

we need only

$$d_s(F_n^Q, F_0^Q) \rightarrow 0 \quad n \rightarrow \infty \quad P. \text{ s. a. s. } .$$

This is straightforward from (4.5).

COROLLARY. Put

$$\Omega = \{g \in \mathbb{R}^p, |g| = 1\} . \quad (4.6)$$

$S(F_n^Q)$ weakly converges (P. s. a. s.) to $S(F_0^Q)$ ($n \rightarrow \infty$) uniformly in g on Ω .

Roughly speaking, Lemmas 4.1 and 4.2 say that under certain conditions, the maximum values and the maximum points of a convergent sequence converge to the maximum value and the maximum point of the limiting function. This would be the basic idea of the consistency of our robust PP-estimates, since $S_1(F_n^Q)$, $A_1(F_n^Q)$, $S_1(F)$ and $A_1(F)$ are maximum values and maximum points.

Lemmas 4.3 and 4.4 will essentially provide the assumptions in Lemmas 4.1 and 4.2 for $S(F_n^Q)$ and $S(F^Q)$. Especially when the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ are all different, it is easy to make $A_1(F)$ ($1 \leq i \leq p$) unique on certain regions; thus Lemma 4.2 holds for $S(F_n^Q)$ and $S(F^Q)$. Then the consistency of $S_1(F_n^Q)$ and $A_1(F_n^Q)$, hence $\hat{S}_1(F_n^Q)$, are almost straightforward.

Now, let F_n ($n = 1, 2, \dots$) be the empirical distribution of $F = F(\mathbb{R}^p, V)$, and recall the notations $F(\mathbb{R}^p, V)$, $\mathbb{R}$, λ_1 , and u_1 introduced in Section 3 (see somewhere around (3.3) and (3.9)). To avoid the unnecessary multiple solutions of $A_1(\cdot)$ caused by (1.5), we can reduce the regions over which the maximum values $S_1(\cdot)$ are obtained. Let g be any p -dimension distribution; put

$$u_1 = g_1(g) \quad , \quad a_1 = A_1(g) \quad ,$$

$$D_1 = \{g \in \mathbb{R}^p, |g| = 1\} \quad ,$$

$$D_1 = \{g \in \mathbb{R}^p, |g| = 1, g \perp u_1, \dots, u_{p-1}\} \quad i = 2, \dots, p \quad .$$

Let W_1 be any half of D_1 satisfying

$$g \in W_1 \Rightarrow -g \notin W_1 \quad ,$$

$$W_1 \cup (-W_1) = D_1 \quad (4.7)$$

where $(-W_1) = \{-g \in W_1\}$. For simplicity, we call such a half of a hypersphere a **SEMI HALP**. From (1.3), it follows that W_1 contains at least one maximum point of $S(F^Q)$ over D_1 , and

$$S_1(g) = \max_{W_1} S(F^Q) \quad .$$

Put

$$a_{1n} = A_1(F_n^Q)$$

and let a_{in} be any value of $A_1(F_n)$ in a certain skew half (we will specify it later) of the $(P-1)$ -dimension hypersphere, and X_n be the associated value of $\frac{1}{2}(F_n)$. We are going to prove the consistency step by step.

THEOREM 4.1. Assume that $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$, then

$$\begin{aligned} s_{in} &\rightarrow \sqrt{\lambda_1} \\ a_{in} a_{in}^T &\rightarrow a_1 a_1^T \quad (n \rightarrow \infty), \quad P. \& \text{ a.s.} \\ X_n &\rightarrow X \end{aligned}$$

Proof. First of all, we adjust our problem to Lemma 4.2, as we have mentioned before. Since the eigenvalues $\lambda_1, \dots, \lambda_p$ are all different, $A_1(F)$ has only two possible values $\pm a_1$, and any skew half of hypersphere $S_1 = \{a | a \in R^p, |a|=1, a_1 a_1^T = a_1 a_1^T, \dots, a_{p-1} a_{p-1}^T = a_{p-1} a_{p-1}^T\}$, say Q_1 , contains only one of the two maximum points $\pm a_1$. We choose Q_1 containing a_1 . Lemma 4.2 also requires that the maximum is unique over the closure of that region, so we need to make sure that

$$-a_1 \notin \bar{Q}_1.$$

Specifically, we choose Q_1 such that a_1 is its "center." In detail, let

$$\begin{aligned} Q_p &= \{a_p\}, \\ Q_{p-1} &= \{a | |a|=1, a^T a_1 = 0, i=1, 2, \dots, p-2; a^T a_{p-1} > 0\} \cup Q_p, \\ Q_{p-2} &= \{a | |a|=1, a^T a_1 = 0, i=1, 2, \dots, p-3; a^T a_{p-2} > 0\} \cup Q_{p-1}, \\ &\dots \dots \dots \\ Q_2 &= \{a | |a|=1, a^T a_1 = 0, a^T a_2 > 0\} \cup Q_3, \\ Q_1 &= \{a | |a|=1, a^T a_1 > 0\} \cup Q_2. \end{aligned} \quad (4.8)$$

Obviously, $Q_i (i=1, 2, \dots, p)$ satisfy (4.7) and

$$s_1(F) = \max_{Q_1} S(F^{Q_1}) = S(F^{a_1}) = \sqrt{\lambda_1} \quad i=1, 2, \dots, p, \\ a_1 \in Q_1, \quad -a_1 \notin \bar{Q}_1.$$

hence, a_1 is the unique maximum of $S(F^{Q_1})$ on $\bar{Q}_1$.

For proving $s_{in} \rightarrow S(F^{a_1})$, we have to create the conditions (1)-(4) required by Lemma 4.1. Actually, we have achieved (1) and (2):

(1) Let Q be defined in (4.6); the continuity of $S(F^{Q_n})$ and $S(F^Q)$ in Q on Q follows by Lemma 4.4 (1).

(2) The corollary of Lemma 4.4 has told us that

$$\lim_{n \rightarrow \infty} S(F^{Q_n}) = S(F^Q) \quad P. \& \text{ a.s.}$$

uniformly in Q on Q .

We shall verify assumptions (3) and (4).

For $i=1$, it is obvious: Since $s_1(F) = \max_{Q_1} S(F^{Q_1}) = S(F^{a_1})$ and

$$a_{in} = \max_{Q_1} S(F^{Q_n}),$$

we need only choose $Q_n = Q_1 = Q_1$ and $P_n = 1$; then by Lemmas 4.1 and 4.2, it follows immediately that

$$\begin{aligned} a_{in} &\rightarrow S(F^{a_1}) = \sqrt{\lambda_1} \\ a_{in} &\rightarrow a_1 \end{aligned} \quad (n \rightarrow \infty) \quad P. \& \text{ a.s.}$$

Consider the case of $i=2$. Let P_{1n} be a rotation matrix as in Lemma 4.3 such that

and define

Obviously, Q_{2n} is a skew half of $(\underline{a} | |\underline{a}|=1, \underline{a} \perp \underline{a}_{2n})$ and

Pick any maximum of $s(P_n^G)$ in Q_{2n} , namely $\underline{a}_{2n}$, then for obtaining

$$\begin{aligned} \alpha_{2n} &\rightarrow \sqrt{\lambda_2} \\ \alpha_{2n} &\rightarrow \alpha_2 \end{aligned} \quad (n \rightarrow \infty) \quad \text{P. S. S. S.}$$

we need only show that

$$P_{1,n} \rightarrow I \quad (n \rightarrow \infty) \quad \text{P. S. A. A.}$$

but this follows from $g_{2i_1} \rightarrow g_{2i_1}$ and Lemma 4.3.

Similarly, we can find a rotation matrix P_{2n} as in Lemma 4.3 such that

$$a_{-2n} = p_{2n} p_{1n-2} \dots$$

Define

$$Q_3 = P_2 P_1 Q_1.$$

then

$$s_{jn} = \frac{\max_k s(\frac{a_k}{n})}{Q_{jn}} = s(\frac{a_{jn}}{n})$$

where $a_{1n} \in \mathbb{Q}_{1n}$. And $a_{2n} \rightarrow a_2$ gives

$$P_{2n}P_{1n} \rightarrow I \quad (n \rightarrow \infty) \quad \text{P. G. A. O.}$$

From Lemma 4.1 and Lemma 4.2, we have

$a_{2n} \rightarrow \sqrt{A_2}$ $(n \rightarrow \infty)$ P. 52.5. .

And so on. Hence, we obtain that

$$u_{1n} \rightarrow \sqrt{\lambda_1} \quad (n \rightarrow \infty) \quad P, \text{ a p.p.} \quad i = 1, 2, \dots, p.$$

Taking the different sign of $\lambda_1(F_n)$, $\lambda_1(F)$ into account, we'd better write as

SECRET - SECRET (S - S) P. 6 a. 6. .

Then it follows that

$$I_n = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \frac{1}{\sqrt{1-x^2}} \frac{1}{\sqrt{1-y^2}} \frac{1}{\sqrt{1-z^2}} dx dy dz \rightarrow I \quad (n \rightarrow \infty) \quad \text{P. G. A. S. .}$$

When λ has multiple eigenvalues, we cannot make the associated eigenvectors unique anymore. But we can take another fact: any set of orthonormal basis of an eigen subspace can be the eigenvectors associated with the eigen subspace.

QUESTION 4.2.

(2) Assume that $\lambda_1 = \lambda_2 = \dots = \lambda_n > 0$, then

$$x_{i+1} = \sqrt{x_i} \quad i = 1, 2, \dots, p \quad n \rightarrow \infty$$

$$x_i \rightarrow x \quad n \rightarrow \infty$$

(2) Assume that $\lambda_1 > \lambda_2 > \dots > \lambda_h > \lambda_{h+1} = \lambda_{h+2} = \dots = \lambda_p > 0$, then

$$u_{1n} \rightarrow \sqrt{\lambda_1} \quad (n \rightarrow \infty) \quad P. \& \&O. \quad i = 1, 2, \dots, p,$$

$$u_{1n} u_{1n}^T \rightarrow u_1 u_1^T \quad (n \rightarrow \infty) \quad P. \& \&O. \quad i = 1, 2, \dots, h,$$

$$\sum_{i=h+1}^p u_{1n} u_{1n}^T \rightarrow \sum_{i=h+1}^p u_i u_i^T \quad (n \rightarrow \infty) \quad P. \& \&O.,$$

$$I_n \rightarrow I \quad (n \rightarrow \infty) \quad P. \& \&O.$$

Proof.

(1) Notice that in the case of $\lambda_1 = \lambda_2 = \dots = \lambda_p$,

$$s(r_n^2) = \sqrt{\lambda_1} \quad \text{for any } q, |q| = 1$$

and for any version of $\lambda_1(r_n)$, namely u_{1n} ($i = 1, 2, \dots, p$).

$$\sum_{i=1}^p u_{1n} u_{1n}^T = I = \sum_{i=1}^p u_i u_i^T.$$

Using Lemma 4.4 (3), we have

$$|u_{1n} - \sqrt{\lambda_1}| = |s(r_n^2) - s(r_n^2)| \rightarrow 0 \quad (n \rightarrow \infty) \quad P. \& \&O.$$

Hence,

$$I_n = \sum_{i=1}^p u_{1n} u_{1n}^T \rightarrow \sum_{i=1}^p u_i u_i^T = I \quad (n \rightarrow \infty) \quad P. \& \&O.$$

(2) For $i = 1, 2, \dots, h$, do exactly what we did in Theorem 4.1

and put

$$u_{1n} = v_{1n} \cdot v_{1-1n} \cdot \dots \cdot v_{1n} \quad (i = 1, 2, \dots, h) \quad (4.9)$$

Then we have

$$Q_{1n} = Q_1 \cdot Q_{1n} = v_{1-1n} Q_1 \quad i = 2, \dots, h,$$

$$u_{1n} = \min_{Q_{1n}} s(r_n^2) = s(r_n^{Q_{1n}}) \quad i = 1, 2, \dots, h,$$

$$Q_{1n} = u_{1n} Q_1 \in Q_{1n} \quad i = 1, 2, \dots, h,$$

and

$$u_{1n} \rightarrow \sqrt{\lambda_1}, \quad u_{1n} \rightarrow u_1 \quad (n \rightarrow \infty) \quad i = 1, 2, \dots, h \quad P. \& \&O. \quad (4.10)$$

Put

$$Q_{h+1n} = Q_{h+1n} Q_{h+1n}$$

We can find $Q_{h+1n}, Q_{h+2n}, \dots, Q_{p-1n} \in Q_{h+1n}$ such that

$$Q_{h+1n} = \min_{Q_{h+1n}} s(r_n^2) = s(r_n^{Q_{h+1n}})$$

$$Q_{h+2n} = \min_{Q \in Q_{h+1n}} s(r_n^2) = s(r_n^{Q_{h+2n}})$$

.....

$$Q_{p-1n} = \min_{Q \in Q_{h+1n}, \dots, Q_{p-2n}} s(r_n^2) = s(r_n^{Q_{p-1n}})$$

$$Q_{pn} = s(r_n^{Q_{pn}})$$

where $Q_{h+1n} \in Q_{h+1n}, Q_{h+2n} \in Q_{h+2n}, \dots, Q_{p-1n}$ and $|Q_{pn}| = 1$. Since

$$u_{kn}^{-1} \in Q_{k+1} \quad k = k+1, \dots, p,$$

$$s(r^2) = \sqrt{p} \quad \text{any } r \in Q_{k+1},$$

$$r_{kn} \rightarrow 1 \quad (n \rightarrow \infty) \quad P. \text{ a.s.},$$

we obtain, following Lemma 4.4 (3),

$$|a_{kn} - \sqrt{p}| = |s(r_{kn}^{-1} u_{kn}) - s(r_{kn}^{-1} u_{kn})| \rightarrow 0 \quad (n \rightarrow \infty) \quad P. \text{ a.s.} \quad k = k+1, \dots, p. \quad (4.11)$$

Notice that

$$\sum_{i=k+1}^p (u_{kn}^{-1} u_{kn})^T = \sum_{i=k+1}^p u_i u_i^T.$$

we have

$$\begin{aligned} \sum_{i=k+1}^p u_{kn} u_{kn}^T - \sum_{i=k+1}^p u_i u_i^T &= \sum_{i=k+1}^p [u_{kn} u_{kn}^T - (u_{kn}^{-1} u_{kn})^T] + \\ &+ \sum_{i=k+1}^p [(u_{kn}^{-1} u_{kn})^T - u_i u_i^T] \rightarrow 0 \quad (n \rightarrow \infty) \quad P. \text{ a.s.} \end{aligned} \quad (4.12)$$

Finally, (4.10), (4.11) and (4.12) together yield

$$r_{kn} \rightarrow 1 \quad (n \rightarrow \infty) \quad P. \text{ a.s.} \quad \square$$

As we have seen in Theorem 4.2, when the smallest eigenvalue is the only one of multiplicity r ($r > 1$) (including the case that all the eigenvalues are the same), it is easy to find an asymptotic identity

rotation (that is, u_{p-m}^{-1}) which transforms all the corresponding estimates of these associated eigenvectors into that eigen subspace. But if there are some other eigenvalues of multiplicity, things will be more complicated. We discuss the case that the largest eigenvalue is the only one of multiplicity first.

THEOREM 4.3. Assume that $\lambda_1 = \lambda_2 = \dots = \lambda_j > \lambda_{j+1} > \dots > \lambda_p > 0$.

Then

$$a_{kn} \rightarrow \sqrt{\lambda_1} \quad (n \rightarrow \infty) \quad P. \text{ a.s.} \quad k = 1, \dots, p$$

$$\sum_{i=1}^j u_{kn} u_{kn}^T \rightarrow \sum_{i=1}^j u_i u_i^T \quad (n \rightarrow \infty) \quad P. \text{ a.s.}$$

$$u_{kn} u_{kn}^T \rightarrow u_i u_i^T \quad (n \rightarrow \infty) \quad P. \text{ a.s.} \quad k = j+1, \dots, p$$

$$r_{kn} \rightarrow 1 \quad (n \rightarrow \infty) \quad P. \text{ a.s.}$$

PROOF. From the argumentation in Theorems 4.1 and 4.2, we can assume that there will be no big problem to verify $a_{kn} \rightarrow \sqrt{\lambda_1}$ ($n \rightarrow \infty$) $P. \text{ a.s.}$, the difficulty lies in the first j eigenvectors.

Denote the linear subspace spanned by $A \in \mathbb{R}^p$ by $L(A)$ and the projection matrix onto a linear subspace L by P_L . Put

$$L_j = L(u_1, \dots, u_j).$$

Notice that any orthonormal bases of L_j can be eigenvectors of the eigen subspace L_j corresponding to λ_1 , and they give the same projection matrix onto L_j as $(u_1, \dots, u_j)$. Our trick is that for every n fixed, we work out a version of $L_j(n)$ in L_j (as we will see later, that is $u_{kn}^{-1} u_{kn}$) related to

the version Q_{1n} of $A_1(P_n)$ such that $A_1(P)$ and $A_1(P_n)$ are as close as possible ($i = 1, 2, \dots, J$).

As before, we can find $Q_{1n} \in Q_1$ such that

$$Q_{1n} = \max_{Q_1} S(P_n^Q) = S(P_n^{Q_{1n}})$$

and from Lemma 4.1, we have

$$Q_{1n} \rightarrow \max_{Q_1} S(P^Q) = \sqrt{\lambda_1} \quad (n \rightarrow \infty) \quad P. \& \text{ a.s.} \quad (4.13)$$

We should choose the first eigenvector in L_1 as close to Q_{1n} as possible.

Then, we pick

$$Q_{1n}^{(1)} = P_{L_1} Q_{1n} / |P_{L_1} Q_{1n}|.$$

Let P_{1n} be the rotation matrix as in Lemma 4.3 such that

$$Q_{1n} = P_{1n} Q_{1n}^{(1)}.$$

We claim that

$$P_{1n} \rightarrow I \quad (n \rightarrow \infty) \quad P. \& \text{ a.s.}$$

If not, i.e.,

$$P(P_{1n} \rightarrow I \text{ (n.s.)}) < 1, \quad (4.14)$$

then we will see a contradiction with (4.13).

Let θ_{1n} ($0 \leq \theta_{1n} \leq \pi/2$) be the angle between Q_{1n} and L_1 (or equivalently, between Q_{1n} and $Q_{1n}^{(1)}$), and $Q_{2n}^{(1)}, \dots, Q_{Jn}^{(1)}$ be such that $\{Q_{1n}^{(1)}, (1 \leq i \leq J)\}$ is a basis of L_1 . Put

$$Q_{2n} = Q_{2n}^{(1)} \cos \theta_{2n} / |Q_{2n}^{(1)} \cos \theta_{2n}|,$$

clearly,

$$Q_{2n} \perp Q_{2n}^{(1)}, \dots, Q_{Jn}^{(1)}.$$

$$Q_{2n} = Q_{2n}^{(1)} \cos \theta_{2n} + Q_{2n} \sin \theta_{2n}.$$

From (4.3) we know that (4.14) is equivalent to

$$P(\lim_{n \rightarrow \infty} \theta_{2n} > 0 \text{ (n.s.)}) < 1.$$

Then, there exists a $\theta_1 > 0$ such that

$$P(\lim_{n \rightarrow \infty} \theta_{2n} > 2\theta_1) = \alpha_1 > 0.$$

Since, for any sample sequence $\omega = (\omega_1, \omega_2, \dots) \in \{\lim_{n \rightarrow \infty} \theta_{2n} > 2\theta_1\}$, there exists a subsequence of $Q_{2n}, Q_{2n}',$ such that $\theta_{2n}' > \theta_1$. Then

$$(S(P_{1n}^{Q_{2n}'}))^2 = Q_{2n}'^2 \perp Q_{1n}',$$

$$= Q_{2n}^{(1)} \cos \theta_{2n}' + Q_{2n}' \sin \theta_{2n}', \quad Q_{2n}^{(1)} \cos \theta_{2n}' + Q_{2n}' \sin \theta_{2n}'$$

$$< \lambda_1 \cos^2 \theta_{2n}' + \lambda_{J+1} \sin^2 \theta_{2n}' = (\lambda_1 - \lambda_{J+1}) \cos^2 \theta_{2n}' + \lambda_{J+1}$$

$$< \lambda_1 \cos^2 \theta_1 + \lambda_{J+1} \sin^2 \theta_1.$$

This yields

$$\lim_{n \rightarrow \infty} S(P_{1n}^{Q_{2n}'}) < \lim_{n \rightarrow \infty} S(P_{1n}^{Q_{2n}'}) < (\lambda_1 \cos^2 \theta_1 + \lambda_{J+1} \sin^2 \theta_1)^{1/2} < \sqrt{\lambda_1}.$$

Therefore,

$$P(\lim_{n \rightarrow \infty} S(P_{1n}^{Q_{2n}'}) < \sqrt{\lambda_1}) > P(\lim_{n \rightarrow \infty} \theta_{2n} > 2\theta_1) = \alpha_1 > 0.$$

This contradicts (4.13). So we have proved that

$$P_{1n} = 1 \quad (n \rightarrow \infty) \quad P. \& \& \& .$$

Put

$$Q_{1n}^{(1)} = Q_1 \quad i = j+1, \dots, p.$$

Clearly, $Q_{1n}^{(1)}, P_{1n}^{(1)}, \dots, P_{1n}^{(p)}$ is a basis of R^p . Define Q_{2n} in the same way as defining Q_2 (see (4.6)), but with $Q_2, \dots, Q_p$ being replaced by $Q_{2n}^{(1)}, \dots, Q_{2n}^{(p)}$. Let

$$L_{2n} = P_{1n} L(Q_{2n}^{(1)}, \dots, Q_{2n}^{(p)})$$

$$Q_{2n} = P_{1n} Q_{2n}.$$

Obviously,

$$Q_{1n} = P_{1n} Q_{1n}^{(1)} \perp L_{2n}, \quad Q_{1n} \perp Q_{2n}. \quad (4.15)$$

Now we see that we have made our problem one dimension lower than before.

Very similarly, we choose $Q_{2n} \in Q_{2n}$ such that

$$Q_{2n} = \min_{Q_{2n}} S(P_{2n}^{(2)}) = S(P_{2n}^{(2)}),$$

define

$$Q_{2n}^{(2)} = P_{1n} Q_{2n} / \|P_{1n} Q_{2n}\|$$

and let:

- (1) P_{2n} be a rotation matrix as in Lemma 4.3 such that

$$Q_{2n} = P_{2n} Q_{2n}^{(2)},$$

- (2) $Q_{2n}^{(2)}, \dots, Q_{2n}^{(p)}$ be such that $Q_{2n}^{(2)}, \dots, Q_{2n}^{(p)}$ is a basis of L_{2n} .

Now we can show that

$$Q_{2n} = \sqrt{1} \quad P_{2n} = 1 \quad (n \rightarrow \infty) \quad P. \& \& \& .$$

Actually, since $Q_{2n}^{(2)} \perp Q_{2n}^{(1)}$, we have

$$S(P_{2n}^{(2)}) = \min_{\substack{Q_{2n}^{(2)} \perp Q_{2n}^{(1)} \\ \|Q_{2n}^{(2)}\|=1}} S(P_{2n}^{(2)}) > S(P_{2n}^{(2)}).$$

Also, from $P_{1n}^{-1} Q_{2n}^{(1)} \in L(Q_{2n}^{(1)}, \dots, Q_{2n}^{(p)}) \subset L_1$ and $P_{1n} = 1$ ($n \rightarrow \infty$) P. & & & .

It follows that

$$S(P_{2n}^{(2)}) = \sqrt{1} - S(P_{2n}^{(2)}) = S(P_{2n}^{(2)}) \rightarrow 0 \quad (n \rightarrow \infty) \quad P. \& \& \& . \quad (4.16)$$

Hence, by Lemma 4.4 (3) we obtain

$$\begin{aligned} \sqrt{1} &= \min_{\|Q_{2n}^{(2)}\|=1} S(P_{2n}^{(2)}) > \lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) = \lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) > \lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) \\ &= \lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) > \lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) = \sqrt{1}. \quad P. \& \& \& . \end{aligned}$$

This implies

$$\lim_{n \rightarrow \infty} Q_{2n} = \lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) = \lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) = \sqrt{1}. \quad P. \& \& \& . \quad (4.17)$$

We are going to show $P_{2n} = 1$ ($n \rightarrow \infty$) P. & & & . Let Q_{2n} be the angle between Q_{2n} and L_{2n} for $Q_{2n}^{(2)} \in Q_{2n} \subset \mathbb{R}^2$. If $\lim_{n \rightarrow \infty} P_{2n} = 1$ (P. & & &) is not true, then by Lemma 4.3 we have

$$V(Q_{2n}) \rightarrow 0 \quad (n \rightarrow \infty) < 1. \quad (4.18)$$

Notice that (4.16) and Lemma 4.4 (3) together give

$$\lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) = \sqrt{\lambda_1}.$$

Using this fact and by some logical argumentation like the case of P_{1n} , (4.18) will lead us to $\lim_{n \rightarrow \infty} S(P_{2n}^{(2)}) < \sqrt{\lambda_1} > 0$. This is a contradiction with (4.17). Therefore,

$$P_{2n} \rightarrow I \quad (n \rightarrow \infty) \quad P. \& \& \& \& .$$

In order to continue our procedure, we need to clarify two important facts:

(A) $u_{2n}^{-1} u_{1n}$ and $u_{2n}^{-1} u_{2n}$ are orthonormal vectors in L_1 , where u_{1n} is defined in (4.9).

(B) Put

$$u_{2n}^{(2)} = P_{1n} u_{2n} \quad k = J+1, \dots, p.$$

then $(u_{1n}, u_{2n}, P_{2n} u_{2n}^{(2)}, \dots, P_{2n} u_{2n}^{(2)})$ is an orthonormal basis of R^p .

Actually, the definition of u_{2n} and $u_{2n}^{(2)}$ implies $u_{2n}^{-1} u_{2n}^{(2)} = P_{1n}^{-1} u_{2n}^{(2)} \in L_1$.

And (4.15) implies $u_{1n} \perp u_{2n}^{(2)}, u_{1n} \perp u_{2n}$. This result and Lemma 4.3 together yield

$$P_{2n} u_{1n} = u_{1n}, \quad u_{2n}^{-1} u_{1n} = P_{1n}^{-1} u_{1n} \in L_1, \quad u_{2n}^{-1} u_{2n} \perp u_{2n}^{-1} u_{2n}^{(2)}.$$

Then the definition of $u_{2n}^{(2)}, \dots, u_{2n}^{(2)}$ and P_{2n} gives that $(u_{1n}, u_{2n}, P_{2n} u_{2n}^{(2)}, \dots, P_{2n} u_{2n}^{(2)})$ is a basis of R^p .

Suppose we have shown $u_{1-1n} = \sqrt{\lambda_1}, P_{1-1n} \rightarrow I \quad (n \rightarrow \infty, P. \& \& \& \& .)$.

Similar to the case of $J=2$, we can find $u_{1n}, L_{1n}, u_{1n}, u_{1n}, u_{2n}^{(1)}$

$(k = 1, J+1, \dots, p)$ and P_{1n} , such that:

(I) u_{1n} is constructed from $u_{1-1n}^{(1-1)}, \dots, u_{2n}^{(1-1)}$ in the way of (4.8),
 $L_{1n} = P_{1-1n} L_{1-1n}^{(1-1)}, \dots, L_{2n}^{(1-1)}, \quad u_{1n} = P_{1-1n} u_{1-1n}$ and $u_{2n}^{(1)} = u_{1-1n} u_{2n}$
 $(k = J+1, \dots, p)$;

(II) $u_{1n} \in L_{1n}, \quad u_{1n} = \max_{u \in L_{1n}} S(P_{1n}^u) = S(P_{1n}^{u_{1n}})$ and

$$u_{1n}^{(1)} = P_{1n} u_{1n} / |P_{1n} u_{1n}|;$$

(III) P_{1n} is a rotation matrix as in Lemma 4.3 such that $u_{1n} = P_{1n} u_{1n}^{(1)}, \quad u_{1n}^{(1)}, u_{1+1n}^{(1)}, \dots, u_{2n}^{(1)}$ is a basis of L_{1n} .

Then we can show that

(a) $u_{1n} = \sqrt{\lambda_1}, \quad P_{1n} \rightarrow I, \quad u_{1n} \rightarrow I \quad (n \rightarrow \infty, P. \& \& \& \& .)$

(b) $(u_{1+1n}^{(1)}, \dots, u_{2n}^{(1)})$ are orthonormal vectors in L_1 ;

(c) $u_{1n}, \dots, u_{1n}, P_{1n} u_{1+1n}^{(1)}, \dots, P_{1n} u_{2n}^{(1)} = (u_{1n}, \dots, u_{1n}, P_{1n} u_{1+1n}^{(1)}, \dots, P_{1n} u_{2n}^{(1)}, u_{2n}^{(1)}, \dots, u_{2n}^{(1)})$ is a basis of R^p , for $i = 2, 3, \dots, J$.

Now we see that (a) has already provided

$$u_{1n} = \sqrt{\lambda_1} \quad (n \rightarrow \infty) \quad P. \& \& \& \& \quad i = 1, 1, \dots, J.$$

Using the method in (4.12), (a) and (b) together yield

$$\sum_{i=1}^J u_{1n} u_{1n}^2 = \sum_{i=1}^J u_{1n}^2 \quad (n \rightarrow \infty) \quad P. \& \& \& \& .$$

As for verifying of

$$\begin{aligned} \sigma_{1n} &\rightarrow \sqrt{\lambda_1} \\ \sigma_{1n}^T \sigma_{1n} &\rightarrow \sigma_1^T \sigma_1 \quad (n \rightarrow \infty) \quad i = 1, 2, \dots, p \quad P. \text{ s. a. s.} \end{aligned}$$

it is almost the same as in Theorem 4.1. Since when $i=1$, (c) tells us: after obtaining $\sigma_{1n}, \sigma_{2n}, \dots, \sigma_{pn}$, despite a rotation U_{1n} (it is asymptotically identity matrix when $n \rightarrow \infty$), we are actually returning to a situation where the eigenvalues $\lambda_{j+1}, \dots, \lambda_p$ are all different. $\square$

Now we can have the general results of consistency.

THEOREM 4.4. Assume that

$$\lambda_1 = \dots = \lambda_{t_1} > \lambda_{t_1+1} = \dots = \lambda_{t_2} > \dots > \lambda_{t_{k-1}+1} = \dots = \lambda_{t_k}$$

where $t_k = p$. Then

$$\begin{aligned} \sigma_{1n} &\rightarrow \sqrt{\lambda_1} \quad (n \rightarrow \infty) \quad P. \text{ s. a. s.} \quad i = 1, 2, \dots, p, \\ \sum_{i=t_k+1}^{t_{k-1}+1} \sigma_{1n}^T \sigma_{1n} &\rightarrow \sum_{i=t_k+1}^{t_{k-1}+1} \sigma_i^T \sigma_i \quad (n \rightarrow \infty) \quad P. \text{ s. a. s.} \quad k = 0, \dots, t-1, \\ I_n &\rightarrow I \quad (n \rightarrow \infty) \quad P. \text{ s. a. s.} \end{aligned} \quad (4.19)$$

Proof. For $i = 1, \dots, t_1$, we do either exactly as in Theorem 4.1 if $t_1 = 1$, or exactly as in Theorem 4.3 if $t_1 > 1$. Then we will have

$$\begin{aligned} \sigma_{1n} &\rightarrow \sqrt{\lambda_1} \quad (i = 1, \dots, t_1) \\ \sigma_{1n} &\rightarrow 0 \quad (i = t_1+1, \dots, t_2) \\ \sigma_{1n} &\rightarrow 0 \quad (i = t_2+1, \dots, t_3) \\ &\vdots \\ \sum_{i=t_k+1}^{t_{k-1}+1} \sigma_{1n}^T \sigma_{1n} &\rightarrow \sum_{i=t_k+1}^{t_{k-1}+1} \sigma_i^T \sigma_i \quad (k = 0, \dots, t-1) \\ I_n &\rightarrow I \quad P. \text{ s. a. s.} \end{aligned} \quad (4.20)$$

For $i = t_k + 1, \dots, t_{k+1}$ ($k = 1, 2, \dots, t-1$), we repeat the same procedure as the one for $i = 1, \dots, t_1$, except we have to deal with $\sigma_{1n} \sigma_{t_k+1}, \dots, \sigma_{1n} \sigma_{t_{k+1}}$, instead of $\sigma_1, \dots, \sigma_p$, and everything is a little more complicated; and, of course, we have to use the results, as in (4.20), which we have obtained. $\square$

Since (1.3) and σ_{1n} is any version of $A_1(W)$ in a show half of the whole $(p-1)$ -dimension hypersphere, (4.19) is actually

$$\begin{aligned} \sigma_1(W_n) &\rightarrow \sigma_1(W) \quad (i = 1, \dots, p) \\ \sum_{i=t_k+1}^{t_{k-1}+1} A_1(W_n) A_1(W_n)^T &\rightarrow \sum_{i=t_k+1}^{t_{k-1}+1} A_1(W) A_1(W)^T \quad k = 0, \dots, t-1 \\ \{W_n\} &\rightarrow \{W\} \quad (n \rightarrow \infty) \quad P. \text{ s. a. s.} \end{aligned} \quad (4.21)$$

Now consider Fisher's choice for $S(\cdot)$. If $F(t)$ is a member of an elliptic probability density family and $F_n(t)$ are the empirical distributions, then, obviously,

$$\begin{aligned} r^2(t) - r^2(t_0) - r^2(t_0) &= 0 \\ r^2(t) - r^2(t_0) - r^2(t_0) &= 0 \quad \text{a. s.} \end{aligned} \quad \text{for any } t \in \mathbb{R}^p, |t| = 1.$$

Hence, according to Theorem 2.2, $S(\cdot)$ is really uniquely defined, finite and nonzero and weakly continuous almost surely on W (W is given in Lemma 4.4 (2)). This means that every condition we need for Theorem 4.1 through Theorem 4.4 is satisfied. Therefore, we have

THEOREM 4.5. If $S(\cdot)$ is Huber's scale estimator and $F(g)$ belongs to an elliptic probability density family, then the robust PP-estimates for covariance matrix and its principal components are consistent (P. & a.s.) in the sense of Theorem 4.4.

Because of consistency, we conclude that although the robust PP-estimates $\hat{\Lambda}_1(P_n)$ and $\hat{\xi}(P_n)$ may not uniquely be determined, they are asymptotically equivalent when the data come from an elliptic distribution.

5. Qualitative and Quantitative Robustness

Qualitative and quantitative robustness (i.e., weakly continuity and breakdown point) is discussed in this section. It is shown first that the robust PP-estimates are weakly continuous.

Assume the p -dimensional probability distribution functions $F_n(g)$ weakly converge to F_0 which belongs to an elliptic density family.

$$F_n \rightarrow F_0 \quad (w).$$

Put

$$\bar{U} = \{e_k^q | e_k \in \mathbb{R}^p, |q| = 1, k = 0, 1, 2, \dots\}.$$

THEOREM 5.1. Assume that $S(\cdot)$ is weakly continuous on $\bar{U}$ and that (4.19) holds. Then robust PP-estimates $\hat{\Lambda}_1(F)$,

$$\sum_{i=1}^{p-1} \hat{\Lambda}_i(F) \hat{\Lambda}_i(F)^T \quad \text{and} \quad \hat{\xi}(F)$$
 are weakly continuous at F_0 .

Proof. Checking all the proofs of Lemma 4.1 through Theorem 4.4, with the convergence P. & a.s. replaced by ordinary convergence, we find that every step will go through except (4.5). What we need in (4.5) is that

$$d_p(F_{n_k}^{S_{k_1}}, F_{n_k}^{S_{k_2}}) \rightarrow 0 \quad (n_k \rightarrow \infty). \quad (5.1)$$

where d_p denotes the weak topology. Denote the Prohorov metric for k -dimensional probability distributions by $d_k(\cdot, \cdot)$. According to the definition of $d_k(\cdot, \cdot)$, we have

$$\begin{aligned}
d_1(r_n^0, r^0) &= \inf(c > 0 | r_n^0(A) < r^0(A^c) + c \text{ for all } A \in \mathcal{B}_1) \\
&< \inf(c > 0 | r_n^0(B) < r^0(B^c) + c \text{ for all } B \in \mathcal{B}_p) \\
&= d_p(r_n^0, r^0) \text{ for any } p \in \mathbb{N}^p, |p| = 1. \quad (5.2)
\end{aligned}$$

where $u^d = \{u | \inf_{y \in u} |u - y| < \delta\}$ ($u \in \mathbb{R}^k$) and $\mathcal{B}_n$ is the Borel- σ -algebra in $\mathbb{R}^k$. Since the Prohorov metric metrizes the weak topology (see [10], Theorem 3.8, p. 28), (5.2) yields that

$$\begin{aligned}
r_n^0(u) &\rightarrow r^0(u) \quad (u) = d_p(r_n^0, r^0) \rightarrow 0 \\
&= d_1(r_n^0, r^0) \rightarrow 0 \text{ uniformly in } u, |u| = 1.
\end{aligned}$$

Hence, (5.1) is also true in the above case. $\square$

COROLLARY 5.1. If $S(\cdot)$ is Huber's choice, then $S(\cdot)$ is weakly continuous on $\mathcal{P}^0$ for any member of an elliptic density family. Hence, Theorem 5.1 holds.

Proof. According to Theorem 2.2, we have only to show that

$$r_n^0(0) = r_n^0(u_n) - r_n^0(0) < 1 - \frac{\delta}{h^2} \quad |u_n| = 1, \quad n = 0, 1, 2, \dots \quad (5.3)$$

Actually, since r_0 is continuous, it is of course true that

$$r_0^0(0) = r_0^0(u) = 0 < 1 - \frac{\delta}{h^2} \quad \text{for any } u, |u| = 1.$$

Also from, for example, Lemma 2.2 in [10] (see [10], p. 22), it follows that

$$0 < \lim_{n \rightarrow \infty} r_n^0(0) < \lim_{n \rightarrow \infty} r_n^0(u) < r_0^0(u) = 0.$$

Thus, we obtain that

$$r_n^0(0) = r_n^0(u) \rightarrow 0 \quad (n \rightarrow \infty) \quad \text{for any } u, |u| = 1.$$

So, without losing generality, we can suppose that (5.3) holds for all $n = 0, 1, 2, \dots$

Now we show that the robust r_0 -estimates have much higher breakdown point than the affinely equivariant estimates, and it does not depend on the dimension.

Actually, for any p -dimensional c -contamination model

$$r = (1-c)r_0 + cr^0,$$

the distribution of any one-dimensional projection $r_{\frac{1}{2}}^0$ is

$$r_{\frac{1}{2}}^0 = (1-c)r_{\frac{1}{2}}^0 + cr_{\frac{1}{2}}^0.$$

And from the discussion at the end of Section 2, we have

$$(1) \text{ When } c < c_1 \text{ (see (2.12)),}$$

$$s_1(r) = s(r_{\frac{1}{2}}^0) < \infty \quad \text{for any } u, \quad i = 1, 2, \dots, p.$$

(Of course, $c_1 = c_1(u)$ depends on $u, i = 1, \dots, p$.) Thus, $\hat{s}(r)$ is finite, whether the minimizing or the maximizing method is used.

If the maximizing method is used, let q be any given direction and let S put all its mass at infinity in the direction q , then

$$s_1(r) > s(r^0) = \infty \quad \text{for } c > c_1.$$

Hence, $\hat{s}(r)$ is infinite, but if the minimizing method is used, it may not break down when $c > c_1$, since it may avoid the higher contamination direction.

(2) If F_0 is nondegenerate, then for any direction g , F_0^g is nondegenerate, too. Thus, $c < c_j$ implies that

$$S_1(F) = S(F^g) > 0 \quad \text{for any } g.$$

Therefore, $\xi(F)$ is not singular for any W . Conversely, if $c > c_j$, let $g = \xi_0(g)$; then for any g , $|g| = 1$, $S(F^g) = 0$. So, $S_1(F) = 0$, $(i = 1, \dots, p)$, $\xi(F) = 0$. These are also the same for both maximizing and minimizing methods.

(3) Now assume the F_0 is degenerate in some direction g , i.e.,

$$F_0(g^T \xi = 0) = 1.$$

Consider minimizing method first. If $c < c_j$, then for any W

$$S_p(F) = \min_{|g|=1} S(F^g) \quad S(F^g) = 0.$$

Obviously, if F_0 is degenerate in a k -dimensional linear subspace L ($F_0(\xi \in L) = 0$), then for any W

$$S_p(F) = S_{p-1}(F) = \dots = S_{p-k+1}(F) = 0 \\ \lambda_p(F), \dots, \lambda_{p-k+2}(F) \in L$$

provided $c < c_j$. So $\xi(F)$ does not misbehave at all.

The maximizing method may not pick up these phenomena, unless F_0 is degenerate into 0, i.e., $F_0(\xi \in 0) = 1$. In this case

$$c < c_j = S(F^g) = 0 \quad \text{for any } g.$$

Then, $S_1(F) = 0$ ($i = 1, 2, \dots, p$) and $\xi(F) = 0$ for both maximizing and minimizing methods. But it's not an interesting case.

To summarize, we conclude that robust VP-estimates have at least the same breakdown point as the projection index $S(\cdot)$ has; that is

$$c^* = \frac{S(0)}{S(1)}$$

except in the case where $F_0(g)$ is degenerate to a real linear subspace and the maximizing method is used. And the minimizing method may give an estimate which has a somewhat higher breakdown point than the maximizing method.

Before ending this section, there is one thing we ought to mention. Principal component analysis is a useful tool for reducing variation. But classical principal component estimation may give a totally misleading outcome when just one or two outliers occur. Robust VP-estimators do not have this problem because of their insensitivity to outliers. Especially, from the discussion in (3) above, the minimum procedure would be very helpful for detecting linear substructures of intermediate dimensions in high dimension, because it reports the degeneracy of the data honestly as long as the fraction of contamination is less than c_j .

6. Some Comments

(1) Denote the two estimators for the covariance matrix from maximizing and minimizing procedures by $\hat{\Sigma}_M(F)$ and $\hat{\Sigma}_m(F)$, respectively. We can define another estimator, denoted by $\hat{\Sigma}_A(F)$, by the average of these two:

$$\hat{\Sigma}_A(F) = \frac{1}{2} (\hat{\Sigma}_M(F) + \hat{\Sigma}_m(F)).$$

Since $\hat{\Sigma}_M(F)$ and $\hat{\Sigma}_m(F)$ both are consistent and weakly continuous at any member of an elliptic probability density family, obviously, so is $\hat{\Sigma}_A(F)$. And $\hat{\Sigma}_A(F)$ should also have the same equivariance and breakdown point as $\hat{\Sigma}_M(F)$, which usually has a lower breakdown point than $\hat{\Sigma}_m(F)$ has.

The simulation results show that the average procedure $\hat{\Sigma}_A(F)$ provides, on the whole, better performance than either $\hat{\Sigma}_M(F)$ or $\hat{\Sigma}_m(F)$.

(2) So far, we have not mentioned anything about correlation estimation which is also important in multivariate data analysis. If we denote the elements of $\hat{\Sigma}(F)$ by $\hat{\Sigma}_{ij}(F)$, i.e.,

$$\hat{\Sigma}(F) = (\hat{\Sigma}_{ij}(F))_{p \times p}, \quad (6.1)$$

then the robust PP-estimates for corresponding correlation coefficients and correlation matrix, denoted by $r_{ij}(F)$ and $R(F)$, can be defined by rescaling, i.e.,

$$r_{ij}(F) = \frac{\hat{\Sigma}_{ij}(F)}{[\hat{\Sigma}_{ii}(F)\hat{\Sigma}_{jj}(F)]^{1/2}} \quad (6.2)$$

$$R(F) = (r_{ij}(F))_{p \times p}.$$

Regarding the breakdown point, if F_0 is nondegenerate, then whether $\hat{\Sigma}_{11}(F) > 0$, in this case, is important. From the discussion in (1), (2) and (3) of Section 5, we know that the breakdown point of $r_{1j}(F)$ and $R(F)$ is

$$c^* = \min\{c_1, c_2, c_3\} = \min \left\{ -\frac{\pi(0)}{12\pi}, \frac{\pi(\pi)}{12\pi} + \frac{r_0(0)}{(1-r_0(0))\sin\theta} \right\}$$

for the contamination model $F = (1-\epsilon)F_0 + \epsilon H$.

(3) In Section 5, we mentioned that robust PP principal components themselves can do a much better job for variation reducing than the classical approach. Also, we would say that using robust PP covariance, instead of the classical one, to constitute a "robust Mahalanobis" distance would give, because of its good robustness, much better results in discrimination and clustering et al. multivariate analysis methods.

(4) It should be pointed out that the methods used here apply in some generality to the study of PP-type procedures.

Notice that Lemma 4.1 through Lemma 4.4 have nothing to do with either an elliptic probability density family or the affinity equivariance of the estimates. What we really need are the weak continuity of the projection index $S(\cdot)$ (in Lemma 4.1, 4.2 and 4.4) and the uniqueness of the maximum points of $S(F^0)$ on the corresponding regions (in Lemma 4.2 only). Hence, if the projection index is weakly continuous, then at least the functional for determining the first maximum value (like $S_1(F)$) should be consistent and weakly continuous. Furthermore, if the maximum points of underlying distribution (like eigenvectors g_1 here) are all uniquely defined in some sense, then the functionals for searching maximum values and maximum points (like $S_1(F)$ and $h_1(F)$ here) should be all consistent and weakly continuous.

7. Acknowledgments

We would like to thank Peter J. Huber for exposing us to this direction, as well as for his careful reading of the first draft and helpful comments. We also wish to express appreciation to David Donoho for his help in editing the manuscript.

References

- [1] Chen, H. et al. (1974), Statistical methods for grouping corporation, *Sankhya*, 36, pp. 1-28.
- [2] Chen, Z. and Li, G. (1981), Robust principal components and dispersion matrices via projection pursuit (manuscript).
- [3] Devlin, S.J. et al. (1975), Robust estimation and outlier detection with correlation coefficients, *Biometrika*, 62, pp. 531-545.
- [4] ——— (1981), Robust estimation of dispersion matrices and principal components, *J. Amer. Statist. Ass.* 76, pp. 354-362.
- [5] Friedman, J.H. et al. (1980), Projection pursuit regression, *Tech. Rep. No. 146*, Dept. of Statist., Stanford Univ.
- [6] Friedman, J. H. and Stuetzle, W. (1980), Projection pursuit classification (unpublished).
- [7] Friedman, J. H. and Tukey, J. W. (1974), A projection pursuit algorithm for exploratory data analysis, *IEEE Trans. Computers*, C-23, pp. 881-889.
- [8] Gnanesikan, R. and Kettenring, J. R. (1972), Robust estimates, residuals and outlier detection with multiresponse data, *Biometrics*, 28, pp. 81-124.
- [9] Huber, P.J. (1977), Robust Covariances, in: *Statistical Decision Theory and Related Topics*, II, S.S. Gupta and D.S. Moore, Eds., Academic Press, New York.
- [10] Huber, P.J. (1981a), *Robust Statistics*, John Wiley & Sons, New York.

- [11] Huber, P.J. (1981b), Density estimation and projection pursuit methods (manuscript).
- [12] Huber, P.J. (1981c), Projection pursuit (manuscript).
- [13] Lehmann, E.L. (1959), *Testing Statistical Hypotheses*, John Wiley & Sons, New York.
- [14] Maronna, R.A. (1976), Robust M-estimators of multivariate location and scatter, *Ann. Statist.*, 4, pp. 51-67.
- [15] Mosteller, F. and Tukey, J.W. (1977), *Data Analysis and Regression*, Addison-Wesley, Reading, Mass.
- [16] Schöcherer, H. (1979), *Robuste Kovarianzen*, Ph.D. Thesis, Eidgen. Technische Hochschule, Zurich.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Robust Projection Pursuit Estimator for Dispersion Matrices and Principal Components		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Guoying Li Zhonglian Chen		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0512 NR 042-425
9. PERFORMING ORGANIZATION NAME AND ADDRESS Harvard University Department of Statistics Cambridge, MA		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Code 411SP Arlington, VA 22217		12. REPORT DATE November 1981
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper proposes and discusses the ROBUST PROJECTION PURSUIT ESTIMATOR for dispersion matrices and their principal components. This estimator finds robust principal components by searching, successively, for directions which maximize (minimize) a robust estimate of scale; the estimate of the dispersion matrix is constructed from the estimated principal components. These estimators are shown below (under mild conditions) to have a number of desirable properties. They are orthogonally equivariant and, within any elliptic underlying density family, asymptotically affinely equivariant.		

DD FORM 1473
1 JAN 73EDITION OF 1 NOV 65 IS OBSOLETE
S. N. 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Furthermore, at elliptic densities, they are consistent and weakly continuous (i.e., qualitatively robust). Finally they have good quantitative robustness--their breakdown point can be as high as $1/2$.

The robust projection pursuit approach is a promising alternative to other estimators of dispersion matrices.

S. N 0102- LF- 014- 6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

DATE
FILMED
-8