98-F-5

Shrinkage to Smooth Non-Convex Cone: Principal Component Analysis as Stein Estimation

Akimichi Takemura
Faculty of Economics, University of Tokyo

Satoshi Kuriki
The Institute of Statistical Mathematics

January 1998

SHRINKAGE TO SMOOTH NON-CONVEX CONE : PRINCIPAL COMPONENT ANALYSIS AS STEIN ESTIMATION

Akimichi Takemura and Satoshi Kuriki

Faculty of Economics, University of Tokyo
Bunkyo-ku, Tokyo 113-0033, Japan
and
The Institute of Statistical Mathematics
Minato-ku, Tokyo 106-8569, Japan

Key Words: matrix mean, reduced rank, second fundamental form, singular value decomposition, Weyl's tube formula.

ABSTRACT

In Kuriki and Takemura (1997a) we established a general theory of James-Stein type shrinkage to convex sets with smooth boundary. In this paper we show that our results can be generalized to the case where shrinkage is toward smooth non-convex cones. A primary example of this shrinkage is descriptive principal component analysis, where one shrinks small singular values of the data matrix. Here principal component analysis is interpreted as the problem of estimation of matrix mean and the shrinkage of the small singular values is regarded as shrinkage of the data matrix toward the manifold of matrices of smaller rank.

1. INTRODUCTION

In Kuriki and Takemura (1997a) we established a general theory of James-Stein type shrinkage to convex sets with smooth boundary using techniques of differential geometry. Tools developed in Kuriki and Takemura (1997a) allow us to investigate shrinkage to much more general sets than the affine subspaces extensively studied in existing literature on Stein estimation. Justification of James-Stein type shrinkage to general non-convex manifolds is very important in many fields including smoothing and nonlinear regression.

In the case of convex sets the rate of shrinkage was expressed in terms of the second fundamental form at the projection point. The second fundamental form is a local property of the boundary of a convex set and similar result is expected to hold for more general manifolds. However convexity is a global property of the set and it guarantees the uniqueness of projection in the whole sample space. In the case of more general manifolds, this global uniqueness of projection is not guaranteed and certain regularity conditions have to be imposed for avoiding this difficulty. In this paper we develop a theory of Stein shrinkage to smooth non-convex cones satisfying some regularity conditions on projection and treat principal component analysis as a primary example of our theory.

In usual practice of principal component analysis, one ignores small eigenvalues of the sample covariance matrix and interprets only the large eigenvalues and the associated eigenvectors. In terms of the original data matrix, this practice amounts to obtaining the singular value decomposition of the (mean adjusted) data matrix and ignoring small singular values. It seems difficult to justify the simple practice of just ignoring small singular values from decision theoretic viewpoint. However if the small singular values are shrunk rather than ignored, then the method of descriptive principal component analysis can be regarded as shrinkage of the data matrix toward the manifold of matrices of smaller rank.

In this paper we regard principal component analysis as estimation of matrix mean. Let $X : n \times p$ be distributed according to the matrix normal distribution $N_{n\times p}(M, I_n \otimes I_p)$. We estimate the matrix mean M with the squared error loss

$$L(M, \hat{M}) = \|\hat{M} - M\|^2 = \mathrm{tr}(\hat{M} - M)'(\hat{M} - M).$$

Let the singular value decomposition of X be written as

$$X = \sum_{i=1}^{p} l_i g_i h_i', \tag{1.1}$$

where $l_1 \geq l_2 \geq \cdots \geq l_p \geq 0$ are the singular values of X. g_i corresponds to the i-th principal component score vector and h_i corresponds to the i-th principal coefficient vector.

In actual practice of principal component analysis, the singular value decomposition is applied to the mean adjusted data matrix. However by considering the orthogonal complement of general mean space and transforming the problem into canonical form, there is no loss of generality in considering the singular value decomposition of X itself.

Let $0 < q < p$ be fixed and define

$$\mathcal{M} = \mathcal{M}_q = \{X \mid \operatorname{rank} X = q\}. \tag{1.2}$$

$\mathcal{M}$ is a smooth manifold of dimension $d = q(n + p - q)$. Furthermore $\mathcal{M}$ is a cone which is not convex. It is well known that when $l_q > l_{q+1}$ the closest point $X_{\mathcal{M}}$ in $\mathcal{M}$ from X,

$$\|X - X_{\mathcal{M}}\| = \min_{Y \in \mathcal{M}} \|X - Y\|,$$

is unique and given by

$$X_{\mathcal{M}} = \sum_{i=1}^{q} l_i g_i h_i'.$$

Let

$$E_{\mathcal{M}} = X - X_{\mathcal{M}} = \sum_{i=q+1}^{p} l_i g_i h_i'$$

and

$$s_1 = \|X_{\mathcal{M}}\|, \quad s_2 = \|E_{\mathcal{M}}\|.$$

We shall consider estimators of the form

$$\hat{M} = (1 - \phi_1(s_1, s_2))X_{\mathcal{M}} + (1 - \phi_2(s_1, s_2))E_{\mathcal{M}}. \tag{1.3}$$

$X_{\mathcal{M}}$ itself is a particular estimator of this class with $\phi_1 = 0$, $\phi_2 = 1$. The simple practice of ignoring the last $p - q$ singular values in the principal component analysis can be interpreted as estimating M by $X_{\mathcal{M}}$. For theoretical convenience we assume that q is fixed . In practice the number of principal components retained is data dependent. In Section 5 we shall discuss how our theory may be generalized to the case where q is data dependent.

In Section 4 we give a sufficient condition for the minimaxity of estimators of the form (1.3). The minimax estimation of M has been discussed in literature: Efron and Morris (1972) gave a minimax estimator of M by empirical Bayes method. Stein (1973) gave an unbiased estimator of risk as well as an estimator which improves upon the Efron-Morris estimator. Stein's results were generalized by Zheng (1986). Ghosh and Shieh (1991) treated the problem under some more general losses. For the the estimation of the matrix mean in the MANOVA model, see Bilodeau and Kariya (1989), Konno (1991a, b), and Honda (1991). In this paper, we do not intend to give an impression that our minimaxity result in Section 4 is a substantial contribution to the existing literature in view of the decision theoretic problem of estimating matrix mean. Rather our objective is to clarify geometry of shrinkage toward $\mathcal{M}_q$ in a general framework.

In Section 2 we develop a general theory of shrinkage toward a smooth non-convex cone. We derive an unbiased estimator of the risk of the estimator of the form (1.3) in a general setting and give a sufficient condition for its minimaxity. Actually the technical results in Section 2 are direct generalizations of the corresponding results in Kuriki and Takemura (1997a), and we refer most of the proofs to those in Kuriki and Takemura (1997a). In Section 3 we establish some geometric properties of $\mathcal{M}_q$. In Section 4 we give a class of minimax estimators of the form (1.3) for the principal component analysis based on the results in Sections 2 and 3. In Section 5 we give some discussion on possible extensions of the results in the present paper.

2. SHRINKAGE TO SMOOTH NON-CONVEX CONE

In this section we develop a general theory of shrinkage toward a smooth non-convex cone. For simplicity we take R^p as the sample space $\mathcal{X}$ and consider estimating the mean vector μ of the multivariate normal distribution $N_p(\mu, I_p)$ with the ordinary squared error loss. Note that the set of $n \times p$ matrices of Section 1 can be identified with R^{np} together with the standard inner product.

Let $\mathcal{M} \subset \mathcal{X}$ be a manifold of class C^2 and we assume that $\mathcal{M}$ forms a cone, that is, $\mathcal{M} = c\mathcal{M}$, $\forall c > 0$. Let

$$d = p - m = \dim \mathcal{M},$$

where m is the codimension of $\mathcal{M}$. Let $S^{p-1} \subset R^p$ be the unit sphere and write

$$\mathcal{M}_1 = \mathcal{M} \cap S^{p-1}.$$

$\mathcal{M}_1$ is a $(d-1)$-dimensional manifold of class C^2.

If there exists a unique closest point $x_{\mathcal{M}}$ in $\mathcal{M}$ from $x \in R^p$ satisfying

$$\|x - x_{\mathcal{M}}\| = \min_{y \in \mathcal{M}} \|x - y\|,$$

then we call $x_{\mathcal{M}}$ the *global projection point* of x onto $\mathcal{M}$. We also say that x has the global projection point if $x_{\mathcal{M}}$ is uniquely defined for x. When $x_{\mathcal{M}}$ is defined we call

$$e_{\mathcal{M}} = x - x_{\mathcal{M}}$$

the residual. The tangent space of $\mathcal{M}$ at $x \in \mathcal{M}$ is denoted by $T_x(\mathcal{M})$ and its orthogonal complement is denoted by $T_x^{\perp}(\mathcal{M})$. Since $\mathcal{M}$ is a cone, $x_{\mathcal{M}}$ itself belongs to the tangent space $T_{x_{\mathcal{M}}}(\mathcal{M})$. It follows that $e_{\mathcal{M}} \in T_{x_{\mathcal{M}}}^{\perp}(\mathcal{M})$ and $x_{\mathcal{M}}$ is orthogonal to $e_{\mathcal{M}}$.

Suppose that x has the global projection point $x_{\mathcal{M}}$ and let

$$s_1 = \|x_{\mathcal{M}}\|, \quad s_2 = \|e_{\mathcal{M}}\|,$$
$$u = x_{\mathcal{M}}/s_1 \in \mathcal{M}_1, \quad v = e_{\mathcal{M}}/s_2.$$

Then x is expressed as

$$x = x_{\mathcal{M}} + e_{\mathcal{M}} = s_1 u + s_2 v. \tag{2.1}$$

Actually u and v have to be expressed in terms of local coordinates. $u \in \mathcal{M}_1$ can be expressed in terms of $(d-1)$-dimensional local coordinate vector θ as $u = u(\theta)$. Now v is the vector on the unit sphere S^{m-1} in $T_{x_{\mathcal{M}}}^{\perp}(\mathcal{M})$. Note that $T_{x_{\mathcal{M}}}(\mathcal{M}) = T_u(\mathcal{M})$ and $T_{x_{\mathcal{M}}}^{\perp}(\mathcal{M}) = T_u^{\perp}(\mathcal{M})$ since $\mathcal{M}$ is a cone. Therefore v can be expressed in terms of θ and additional $(m-1)$-dimensional local coordinate vector τ as $v = v(\theta, \tau)$. Differentiability of these local coordinates can be shown as in Section 2 of Kuriki and Takemura (1997a). For simplicity we omit local coordinates because there is no possibility of confusion.

We have considered (2.1) for x which has the global projection point. We need a regularity condition which guarantees that sufficiently many points x have the global projection point. Our regularity condition is given in Assumption 2.1 below. In order to state our regularity condition, choose arbitrary $u \in \mathcal{M}_1$ and $v \in S^{m-1} \subset T_u^\perp(\mathcal{M})$ and consider x of the form (2.1) for $s_1 > 0$ and $s_2 > 0$. We are interested in the range of (s_1, s_2) such that $s_1 u$ is the global projection point of $x = s_1 u + s_2 v$. As we show below, this range is closely related to the radius of curvature at the point u.

For $x \in \mathcal{M}$ and $y \in T_x^\perp(\mathcal{M})$ let $H(x, y)$ denote the second fundamental form of $\mathcal{M}$ at x with respect to the normal vector y. Then $H(x, y) = s_2 H(x, v)$, where $s_2 = \|y\|$ and $v = y/s_2$. Furthermore since $\mathcal{M}$ is a cone we have

$$H(x, v) = \frac{1}{s_1} H(u, v), \qquad s_1 = \|x\|, \ u = x/s_1.$$

Therefore it suffices to consider $H(u, v)$ for $\|u\| = \|v\| = 1$. If $H(u, v)$ is nonnegative definite, define

$$\gamma(u, v) = \infty.$$

In this case, $\mathcal{M}$ looks locally like the boundary of a convex set and v can be regarded as an outward normal vector of the convex set at u. Otherwise, let $-\lambda_{\min}$ denote the the minimum eigenvalue of $H(u, v)$, which is negative, and define

$$\gamma(u, v) = \frac{1}{\lambda_{\min}}.$$

In this case, at the boundary $s_2 = s_1 \gamma(u, v)$ the projection onto $\mathcal{M}$ becomes ambiguous even in the local sense.

Define two dimensional cone $\mathcal{C}(u, v)$ by

$$\mathcal{C}(u, v) = \{ x \mid x = s_1 u + s_2 v, \ 0 < s_1, \ 0 < s_2 < s_1 \gamma(u, v) \}.$$

We call the pair of directions (u, v) ($u \in \mathcal{M}_1$, $v \in S^{m-1} \subset T_u^\perp(\mathcal{M})$) *regular with respect to projection onto* $\mathcal{M}$ or *projection regular* (p.r.) if every $x \in$

$\mathcal{C}(u,v)$ has the global projection point $x_{\mathcal{M}}$ which coincides with $s_1 u$. Let $\mathcal{X}_{\mathcal{M}}$ denote the union of $\mathcal{C}(u,v)$ for all projection regular (u,v):

$$\mathcal{X}_{\mathcal{M}} = \bigcup_{(u,v):\text{p.r.}} \mathcal{C}(u,v).$$

Now we make the following regularity assumption on $\mathcal{M}$.

Assumption 2.1 *The interior* $\operatorname{int}\mathcal{X}_{\mathcal{M}}$ *of* $\mathcal{X}_{\mathcal{M}}$ *is not empty.*

From now on we consider shrinking x in the region $\operatorname{int}\mathcal{X}_{\mathcal{M}}$ only. Because we are now considering x in the open set $\operatorname{int}\mathcal{X}_{\mathcal{M}}$, where the correspondence

$$x \leftrightarrow (s_1, u, s_2, v)$$

is one-to-one, we can consider the Jacobian of this correspondence. This Jacobian is the basic technical tool for our theory and was originally derived in Weyl (1939). A simpler proof is given in Appendix A of Kuriki and Takemura (1997a).

Lemma 2.1

$$\begin{aligned} dx &= |I_d + H(x_{\mathcal{M}}, e_{\mathcal{M}})|\; s_1^{d-1} ds_1\; du\; s_2^{m-1} ds_2\; dv \\ &= \left|I_d + \frac{s_2}{s_1} H(u,v)\right|\; s_1^{d-1} ds_1\; du\; s_2^{m-1} ds_2\; dv, \end{aligned}$$

where du *denotes the volume element of* $\mathcal{M}_1$*,* dv *denotes the volume element of* $S^{m-1} \subset T_u^{\perp}(\mathcal{M})$*, and* $H(u,v)$ *is the second fundamental form of* $\mathcal{M}$ *at* u *with respect to the direction* v.

Remark 2.1 Since $\mathcal{M}$ is a cone, $H(u,v)$ has at least one eigenvalue equal to 0 with the associated eigenvector u. Therefore $|I_d + (s_2/s_1)H(u,v)|s_1^{d-1}$ is a polynomial in (s_1, s_2).

Let x be distributed according to $N_p(\mu, I_p)$. From Lemma 2.1 the conditional density of (s_1, s_2) given (u,v) is written as follows. We omit the proof since it is a direct generalization of Lemma 2.2 of Kuriki and Takemura (1997a).

Lemma 2.2 *On* $\mathrm{int}\,\mathcal{X}_{\mathcal{M}}$ *the conditional density of* (s_1, s_2) *given* (u, v) *is written as*

$$f(s_1, s_2 \mid u, v) \propto \exp\big\{ -\frac{1}{2}(s_1^2 + s_2^2) + s_1 u'\mu + s_2 v'\mu \big\} \\ \times \big|I_d + \frac{s_2}{s_1} H(u, v)\big| s_1^{d-1} s_2^{m-1}.$$

As already stated in Section 1, we consider estimator of the form

$$\begin{aligned} \hat{\mu} &= (1 - \phi_1(s_1, s_2))x_{\mathcal{M}} + (1 - \phi_2(s_1, s_2))e_{\mathcal{M}} \\ &= \big(1 - \frac{c_1(s_1, s_2)}{s_1^2}\big)x_{\mathcal{M}} + \big(1 - \frac{c_2(s_1, s_2)}{s_2^2}\big)e_{\mathcal{M}} \end{aligned} \tag{2.2}$$

for $x \in \mathrm{int}\,\mathcal{X}_{\mathcal{M}}$. For $x \notin \mathrm{int}\,\mathcal{X}_{\mathcal{M}}$ we let $\hat{\mu} = x$. (ϕ_1, ϕ_2) or equivalently (c_1, c_2) in (2.2) can depend on u, v.

We now derive an unbiased estimator of risk for estimator of this class under the following boundary condition for integration by parts.

Assumption 2.2 c_1 *and* c_2 *in (2.2) are continuous in* (s_1, s_2). c_1 *is piecewise differentiable in* s_1 *for each* s_2 *and* c_2 *is piecewise differentiable in* s_2 *for each* s_1. *Furthermore*

$$\lim_{s_1 \to s_2/\gamma(u,v), \infty} \frac{c_1(s_1, s_2)}{s_1} f(s_1, s_2 \mid u, v) = 0,$$

$$\lim_{s_2 \to 0, s_1\gamma(u,v)} \frac{c_2(s_1, s_2)}{s_2} f(s_1, s_2 \mid u, v) = 0.$$

Under Assumption 2.2 an unbiased estimator of the risk difference

$$\Delta R = R(\mu, \hat{\mu}) - R(\mu, x) = R(\mu, \hat{\mu}) - p$$

is given in the following lemma.

Lemma 2.3 *An unbiased estimator of the risk difference* $\widehat{\Delta R}$ *is given by*

$$\begin{aligned} \widehat{\Delta R} &= \sum_{i=1}^{2} \big(\phi_i^2 s_i^2 - 2\frac{\partial \phi_i}{\partial s_i} s_i - 2\phi_i s_i \frac{\partial}{\partial s_i} \log P(s_1, s_2) - 2\phi_i\big) \\ &= \sum_{i=1}^{2} \big(\frac{c_i^2}{s_i^2} - 2\frac{c_i}{s_i^2}(q_i(s_1, s_2) - 1) - 2\frac{1}{s_i}\frac{\partial c_i}{\partial s_i}\big), \end{aligned}$$

where

$$P(s_1, s_2) = \left|I_d + \frac{s_2}{s_1} H(u,v)\right| s_1^{d-1} s_2^{m-1},$$

$$q_i(s_1, s_2) = s_i \frac{\partial}{\partial s_i} \log P(s_1, s_2), \quad i = 1, 2.$$

Proof is omitted since this is a direct generalization of Lemma 3.1 of Kuriki and Takemura (1997a).

Based on Lemma 2.3 we can give a sufficient condition for the minimaxity of estimators of the form (2.2) as follows.

Lemma 2.4 *An estimator of the form (2.2) satisfying Assumption 2.2 is minimax if* $\widehat{\Delta R} \le 0$ *on* $\mathrm{int}\, \mathcal{X}_{\mathcal{M}}$.

It seems that, given a particular form of the second fundamental form $H(u, v)$, many minimax estimators satisfying the condition of Lemma 2.4 can be constructed. However it is desirable to have a simpler sufficient condition which guarantees the condition of Lemma 2.4.

Let d_0, d_+, d_- denote the number of zero eigenvalues, positive eigenvalues, and negative eigenvalues of $H(u, v)$, respectively. In Remark 2.1 we mentioned that $d_0 \ge 1$. Let the positive eigenvalues of $H(u, v)$ be denoted by $0 < \lambda_{+1} \le \cdots \le \lambda_{+d_+}$ and the negative eigenvalues of $H(u, v)$ be denoted by $0 > -\lambda_{-1} \ge \cdots \ge -\lambda_{-d_-}$, where $\lambda_{-d_-} = \lambda_{\min} = 1/\gamma(u, v)$.

For (u, v) such that $H(u, v)$ is nonnegative definite, the estimator for the convex case in Theorem 3.1 of Kuriki and Takemura (1997a) can be directly applied. More precisely let $t_i = s_1/(s_1 + s_2\lambda_{+i})$, $i = 1, \ldots, d_+$, and let

$$\begin{aligned} c_1(s_1, s_2) &= \begin{cases} d_0 - 2 + \sum_{i=1}^{d_+} t_i, & \text{if } d_0 \ge 2, \\ d_0 - 2 + \sum_{i=1}^{d_+} t_i + \prod_{i=1}^{d_+} (1 - t_i), & \text{if } d_0 = 1, \end{cases} \\ c_2(s_1, s_2) &= \begin{cases} m - 2 + \sum_{i=1}^{d_+} (1 - t_i), & \text{if } m \ge 2, \\ m - 2 + \sum_{i=1}^{d_+} (1 - t_i) + \prod_{i=1}^{d_+} t_i, & \text{if } m = 1, \end{cases} \end{aligned} \tag{2.3}$$

then as in Theorem 3.1 of Kuriki and Takemura (1997a) it can be shown that $\widehat{\Delta R} \le 0$. The quantities on the right hand side of (2.3) can be motivated by a geometric notion of average codimension as discussed in Section 3.2 of Kuriki and Takemura (1997a).

When $H(u,v)$ has negative roots, it is difficult to give a simple explicit estimator as in (2.3). We present a sufficient condition which covers the case of the principal component analysis in Section 4. However it is not as satisfactory as (2.3). We consider (u,v) such that $0 < d_- \le d_+$ and

$$\lambda_{+i} \le \lambda_{-i} \le \lambda_{+(i+d_+-d_-)}, \qquad i = 1, \dots, d_-. \tag{2.4}$$

Then we have the following theorem.

Theorem 2.1 *Consider (u,v) such that $0 < d_- \le d_+$ and (2.4) is satisfied. Suppose that $d_0 \ge 3$, $m \ge 3$. Write $d_1 = d_0 - 2$, $d_2 = m - 2$. If $\partial c_i / \partial s_i \ge 0$, $0 \le c_i \le 2d_i$, $i = 1, 2$, and*

$$\frac{c_1}{c_2} \ge \frac{1}{\gamma^2(u,v)},$$

then $\widehat{\Delta R} \le 0$.

Proof $P(s_1, s_2)$ can be written as

$$P(s_1, s_2) = s_1^{d_0-1} s_2^{m-1} \prod_{i=1}^{d_+} (s_1 + s_2\lambda_{+i}) \prod_{i=1}^{d_-} (s_1 - s_2\lambda_{-i}).$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^{2} \Big(\frac{c_i^2}{s_i^2} - 2\frac{c_i}{s_i^2}(q_i - 1)\Big) \\
&\quad = \frac{c_1(c_1 - 2d_1)}{s_1^2} - 2\frac{c_1}{s_1^2}\Big(\sum_{i=1}^{d_+} \frac{s_1}{s_1 + s_2\lambda_{+i}} + \sum_{i=1}^{d_-} \frac{s_1}{s_1 - s_2\lambda_{-i}}\Big) \\
&\quad + \frac{c_2(c_2 - 2d_2)}{s_2^2} - 2\frac{c_2}{s_2^2}\Big(\sum_{i=1}^{d_+} \frac{s_2\lambda_{+i}}{s_1 + s_2\lambda_{+i}} - \sum_{i=1}^{d_-} \frac{s_2\lambda_{-i}}{s_1 - s_2\lambda_{-i}}\Big).
\end{aligned}$$

Now by (2.4),

$$\sum_{i=1}^{d_+} \frac{s_1}{s_1 + s_2\lambda_{+i}} \ge \sum_{i=1}^{d_-} \frac{s_1}{s_1 + s_2\lambda_{+i}} \ge \sum_{i=1}^{d_-} \frac{s_1}{s_1 + s_2\lambda_{-i}}.$$

Hence

$$\sum_{i=1}^{d_+} \frac{s_1}{s_1 + s_2\lambda_{+i}} + \sum_{i=1}^{d_-} \frac{s_1}{s_1 - s_2\lambda_{-i}} \ge \sum_{i=1}^{d_-} \frac{2s_1^2}{s_1^2 - s_2^2\lambda_{-i}^2}.$$

Similarly using $\lambda_{-i} \le \lambda_{+(i+d_+-d_-)}$, $i = 1, \ldots, d_-$, we obtain

$$\sum_{i=1}^{d_+} \frac{s_2\lambda_{+i}}{s_1 + s_2\lambda_{+i}} - \sum_{i=1}^{d_-} \frac{s_2\lambda_{-i}}{s_1 - s_2\lambda_{-i}} \ge -\sum_{i=1}^{d_-} \frac{2s_2^2\lambda_{-i}^2}{s_1^2 - s_2^2\lambda_{-i}^2}.$$

It follows that

$$\sum_{i=1}^{2} \Big(\frac{c_i^2}{s_i^2} - 2\frac{c_i}{s_i^2}(q_i - 1)\Big) \le \frac{c_1(c_1 - 2d_1)}{s_1^2} + \frac{c_2(c_2 - 2d_2)}{s_2^2} - 4\sum_{i=1}^{d_-} \frac{c_1 - c_2\lambda_{-i}^2}{s_1^2 - s_2^2\lambda_{-i}^2}.$$

Note that if $c_1 \ge c_2/\gamma^2(u,v) = c_2\lambda_{-d_-}^2$ then $c_1 - c_2\lambda_{-i}^2 \ge 0$ for $1 \le i \le d_-$. It is now easy to see that under the assumption of the theorem $\widehat{\Delta R} \le 0$. Q.E.D.

3. GEOMETRY OF THE SET OF MATRICES OF REDUCED RANK

In this section we derive some basic results on geometry of $\mathcal{M} = \mathcal{M}_q$ in (1.2). These results are of some independent interest. Recall that the singular value decomposition of $X : n \times p$ is written as (1.1). As mentioned in Section 1, $\mathcal{M}$ has the special property that every X for which $l_q > l_{q+1}$ has the global projection point $X_{\mathcal{M}}$. Furthermore from Lemma 3.4 below it follows that

$$\text{int } \mathcal{X}_{\mathcal{M}} = \{X \mid l_q > l_{q+1}\}.$$

Therefore int $\mathcal{X}_{\mathcal{M}}$ coincides with the whole sample space $\mathcal{X} = R^{np}$ except for a null set.

Now we determine some differential geometric properties of $\mathcal{M}$. First we consider the tangent space.

Lemma 3.1 *Let g_j, $p+1 \le j \le n$, be $n \times 1$ vectors such that the matrix $(g_1, \ldots, g_n)$ is $n \times n$ orthogonal. Then the tangent space of $\mathcal{M}$ and its orthogonal complement at $X_{\mathcal{M}}$ are given by*

$$T_{X_{\mathcal{M}}}(\mathcal{M}) = \text{Span}(g_i h_j', \min(i,j) \le q)$$

and

$$T_{X_{\mathcal{M}}}^{\perp}(\mathcal{M}) = \text{Span}(g_i h_j', \min(i,j) > q),$$

respectively. Their dimensions are $d = q(n+p-q)$ and $m = (n-q)(p-q)$, respectively.

Proof is easy and omitted.

The Jacobian of the singular value decomposition (1.1) is known to be

$$dX = \prod_{i=1}^{p} l_i^{n-p} \prod_{1\le i<j\le p} (l_i^2 - l_j^2) \prod_{i=1}^{p} dl_i \bigwedge_{1\le i\le p, i<j\le n} g_j' dg_i \bigwedge_{1\le i<j\le p} h_j' dh_i, \quad (3.1)$$

where $dX = \prod_{i=1}^{n}\prod_{j=1}^{p} dx_{ij}$ is the Lebesgue measure of R^{np} ((8.8) of James (1954)). We show here that the right hand side of (3.1) can be split into three parts. Proofs of the following three lemmas are similar to those of Takemura and Kuriki (1995) or Kuriki and Takemura (1997b) and we only give sketches.

Lemma 3.2 *The volume element of $\mathcal{M}$ at $X_{\mathcal{M}} = \sum_{i=1}^{q} l_i g_i h_i'$ is given by*

$$\prod_{i=1}^{q} l_i^{n+p-2q} \prod_{1\le i<j\le q} (l_i^2 - l_j^2) \prod_{i=1}^{q} dl_i \bigwedge_{1\le i\le q, i<j\le n} g_j' dg_i \bigwedge_{1\le i\le q, i<j\le p} h_j' dh_i. \quad (3.2)$$

Lemma 3.3 *The volume element of $T_{X_{\mathcal{M}}}^{\perp}(\mathcal{M})$ at $E_{\mathcal{M}} = X - X_{\mathcal{M}} = \sum_{i=q+1}^{p} l_i g_i h_i'$ is given by*

$$\prod_{i=q+1}^{p} l_i^{n-p} \prod_{q+1\le i<j\le p} (l_i^2 - l_j^2) \prod_{i=q+1}^{p} dl_i \bigwedge_{q+1\le i\le p, i<j\le n} g_j' dg_i \bigwedge_{q+1\le i<j\le p} h_j' dh_i. \quad (3.3)$$

Proof of Lemmas 3.2 and 3.3 Let $L_1 = \mathrm{diag}(l_1, \ldots, l_q)$, $G_1 = (g_1, \ldots, g_q)$, $G_2 = (g_{q+1}, \ldots, g_n)$, $H_1 = (h_1, \ldots, h_q)$, and $H_2 = (h_{q+1}, \ldots, h_p)$. The exterior derivative of $X_{\mathcal{M}} = G_1 L_1 H_1'$ is given by

$$dX_{\mathcal{M}} = dG_1 L_1 H_1' + G_1 dL_1 H_1' + G_1 L_1 dH_1'.$$

Consider an orthogonal transformation which preserves the inner product:

$$dY = (dy_{ij}) = \begin{pmatrix} G_1' \\ G_2' \end{pmatrix} dX_{\mathcal{M}} \, (H_1 \; H_2)$$
$$= \begin{pmatrix} G_1' dG_1 L_1 + dL_1 + L_1 dH_1' H_1 & L_1 dH_1' H_2 \\ G_2' dG_1 L_1 & O \end{pmatrix}.$$

This implies that

$$
\begin{array}{ll}
dy_{ii} = dl_i & (1 \le i \le q) \\
dy_{ij} = \begin{cases} -l_j g_j' dg_i + l_i h_j' dh_i \\ l_i h_j' dh_i \end{cases} & \begin{array}{l} (1 \le i < j \le q) \\ (1 \le i \le q,\ q+1 \le j \le p) \end{array} \\
dy_{ji} = \begin{cases} l_i g_j' dg_i - l_j h_j' dh_i \\ l_i g_j' dg_i \end{cases} & \begin{array}{l} (1 \le i < j \le q) \\ (1 \le i \le q,\ q+1 \le j \le n) \end{array}
\end{array}
$$

and all the other $dy_{ij} = 0$. We can now obtain (3.2) by taking the exterior product $\bigwedge_{\min(i,j) \le q} dy_{ij}$.

Let $L_2 = \operatorname{diag}(l_{q+1}, \ldots, l_p)$, and $G_{21} = (g_{q+1}, \ldots, g_p)$, $G_{22} = (g_{p+1}, \ldots, g_n)$. The exterior derivative of $E_{\mathcal{M}} = G_{21} L_2 H_2'$ is given by

$$
dE_{\mathcal{M}} = dG_{21} L_2 H_2' + G_{21} dL_2 H_2' + G_{21} L_2 dH_2'.
$$

Noting that $G_1' dG_{21} = O$ and $H_1' dH_2 = O$, we have

$$
\begin{aligned}
d\tilde{Y} = (d\tilde{y}_{ij}) &= \begin{pmatrix} G_1' \\ G_{21}' \\ G_{22}' \end{pmatrix} dE_{\mathcal{M}}\ (H_1\ H_2) \\
&= \begin{pmatrix} O & O \\ O & G_{21}' dG_{21} L_2 + dL_2 + L_2 dH_2' H_2 \\ O & G_{22}' dG_{21} L_2 \end{pmatrix},
\end{aligned}
$$

which implies

$$
\begin{array}{ll}
d\tilde{y}_{ii} = dl_i & (q+1 \le i \le p) \\
d\tilde{y}_{ij} = -l_j g_j' dg_i + l_i h_j' dh_i & (q+1 \le i < j \le p) \\
d\tilde{y}_{ji} = \begin{cases} l_i g_j' dg_i - l_j h_j' dh_i \\ l_i g_j' dg_i \end{cases} & \begin{array}{l} (q+1 \le i < j \le p) \\ (q+1 \le i \le p,\ p+1 \le j \le n) \end{array}
\end{array}
$$

and all the other $d\tilde{y}_{ij} = 0$. We can now obtain (3.3) by taking the exterior product $\bigwedge_{\min(i,j) > q} d\tilde{y}_{ij}$. Q.E.D.

Lemma 3.4 *The second fundamental form $H(X_{\mathcal{M}}, E_{\mathcal{M}})$ has the following $2q(p-q)$ nonzero eigenvalues:*

$$
\pm \frac{l_j}{l_i}, \qquad 1 \le i \le q,\ q+1 \le j \le p.
$$

Proof Define the matrices G's, L's, and H's as in the proof of Lemmas 3.2 and 3.3. Then, $X_{\mathcal{M}} = G_1 L_1 H_1'$, $E_{\mathcal{M}} = X - X_{\mathcal{M}} = G_{21} L_2 H_2'$. We adopt a new coordinate system of $\mathcal{X}$ by the orthogonal transformation $X \mapsto \begin{pmatrix} G_1' \\ G_2' \end{pmatrix} X \ (H_1 \ H_2)$. In the new coordinate system, we can write

$$X_{\mathcal{M}} = \begin{pmatrix} L_1 & O \\ O & O \end{pmatrix}, \qquad E_{\mathcal{M}} = \begin{pmatrix} O & O \\ O & \begin{pmatrix} L_2 \\ O \end{pmatrix} \end{pmatrix} \in T^{\perp}_{X_{\mathcal{M}}}(\mathcal{M}),$$

where $\begin{pmatrix} L_2 \\ O \end{pmatrix}$ is $(n-q) \times (p-q)$, and

$$T^{\perp}_{X_{\mathcal{M}}}(\mathcal{M}) = \left\{ Y = \begin{pmatrix} O & O \\ O & Y_{22} \end{pmatrix} \;\middle|\; Y_{22} : (n-q) \times (p-q) \right\}.$$

It is easily shown that a partitioned matrix with $(1,1)$ block $q \times q$, say $Z = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}$, in the neighborhood of $X_{\mathcal{M}}$ is of rank q if and only if $Z_{22} = Z_{21} Z_{11}^{-1} Z_{12}$. Hence, the triplet (Z_{11}, Z_{12}, Z_{21}) can be taken as a local orthonormal coordinate system of $\mathcal{M}$. Consider the Taylor expansion of $\operatorname{tr}(E_{\mathcal{M}}' Z)$ around $Z = X_{\mathcal{M}}$, that is, $(Z_{11}, Z_{12}, Z_{21}) = (L_1, O, O)$. Then,

$$\begin{aligned} \operatorname{tr}(E_{\mathcal{M}}' Z) &= \operatorname{tr}((L_2, O) Z_{21} L_1^{-1} Z_{12}) \\ &\qquad - \operatorname{tr}((L_2, O) Z_{21} L_1^{-1} (Z_{11} - L_1) L_1^{-1} Z_{12}) + \cdots, \end{aligned}$$

and the quadratic term (the leading term) is written as

$$\frac{1}{2} (\operatorname{vec}(Z_{211})', \operatorname{vec}(Z_{12}')') \begin{pmatrix} O & L_1^{-1} \otimes L_2 \\ L_1^{-1} \otimes L_2 & O \end{pmatrix} \begin{pmatrix} \operatorname{vec}(Z_{211}) \\ \operatorname{vec}(Z_{12}') \end{pmatrix},$$

where Z_{211} is a $(p-q) \times q$ matrix consisting of first $p-q$ rows of Z_{21}. (See Muirhead (1982), page 74, for the definition of $\operatorname{vec}(\cdot)$.) This means that the nonzero part of the second fundamental form of $\mathcal{M}$ at $X_{\mathcal{M}}$ with respect to the direction $E_{\mathcal{M}} = X - X_{\mathcal{M}}$ is

$$-\begin{pmatrix} O & L_1^{-1} \otimes L_2 \\ L_1^{-1} \otimes L_2 & O \end{pmatrix},$$

which has eigenvalues $\pm l_j / l_i$, $1 \le i \le q$, $q+1 \le j \le p$. Q.E.D.

Remark 3.1 We can now confirm that the rhs of (3.1) is factored into three parts: the volume element of $\mathcal{M}$ at $X_{\mathcal{M}}$ in (3.2), the volume element of $T^{\perp}_{X_{\mathcal{M}}}(\mathcal{M})$ at $E_{\mathcal{M}} = X - X_{\mathcal{M}}$ in (3.3), and $|I + H(X_{\mathcal{M}}, E_{\mathcal{M}})| = \prod_{i=1}^{q}\prod_{j=q+1}^{p}(1 - l_j^2/l_i^2)$.

The lengths of $X_{\mathcal{M}}$ and $E_{\mathcal{M}}$ are given by

$$s_1 = \|X_{\mathcal{M}}\| = \sqrt{l_1^2 + \cdots + l_q^2}, \quad s_2 = \|E_{\mathcal{M}}\| = \sqrt{l_{q+1}^2 + \cdots + l_p^2}.$$

Define

$$(\tilde{l}_1, \ldots, \tilde{l}_q) = (l_1, \ldots, l_q)/s_1,$$
$$(\tilde{l}_{q+1}, \ldots, \tilde{l}_p) = (l_{q+1}, \ldots, l_p)/s_2.$$

Then from Lemma 3.4, $H(u,v) = H(X_{\mathcal{M}}/s_1, E_{\mathcal{M}}/s_2)$ has $2q(p-q)$ nonzero eigenvalues

$$\pm\frac{\tilde{l}_j}{\tilde{l}_i}, \qquad 1 \le i \le q,\ q+1 \le j \le p,$$

and the radius of curvature is given by

$$\gamma(u,v) = \frac{\tilde{l}_q}{\tilde{l}_{q+1}}.$$

From this radius of curvature it follows that every (u,v) is regular with respect to projection onto $\mathcal{M}$.

4. A CLASS OF MINIMAX ESTIMATORS FOR PRINCIPAL COMPONENT ANALYSIS

It is now straightforward to give a class of minimax estimators in the principal component setting based on the results of Sections 2 and 3.

By Lemma 3.4, $d_+ = d_- = q(p-q)$ and $d_0 = q(n+p-q) - 2q(p-q) = q(n-p+q)$. Let

$$P(s_1, s_2) = s_1^{q(n-p+q)-1} s_2^{(n-q)(p-q)-1} \prod_{1\le i\le q, q+1\le j\le p} (s_1^2\tilde{l}_i^2 - s_2^2\tilde{l}_j^2).$$

Here we have multiplied $P(s_1,s_2)$ of Lemma 2.3 by $\prod_{i=1}^{q}\tilde{l}_i^{2(p-q)}$ for notational simplicity. Now by Lemma 2.2 the conditional density of (s_1,s_2) is written as

$$f(s_1, s_2) \propto \exp\{ -\frac{1}{2}(s_1^2 + s_2^2) + s_1\nu_1 + s_2\nu_2\}P(s_1, s_2),$$

where $\nu_1 = \sum_{i=1}^{q} \tilde{l}_i g_i' M h_i$, $\nu_2 = \sum_{i=q+1}^{p} \tilde{l}_i g_i' M h_i$. Therefore the unbiased estimator of the risk difference is given by

$$\begin{aligned}\widehat{\Delta R} = &\left(\frac{c_1^2}{s_1^2} - 2(q(n-p+q)-2)\frac{c_1}{s_1^2} - 2\frac{1}{s_1}\frac{\partial c_1}{\partial s_1}\right) \\ &+ \left(\frac{c_2^2}{s_2^2} - 2((n-q)(p-q)-2)\frac{c_2}{s_2^2} - 2\frac{1}{s_2}\frac{\partial c_2}{\partial s_2}\right) \\ &- 4 \sum_{1 \le i \le q, q+1 \le j \le p} \frac{c_1\tilde{l}_i^2 - c_2\tilde{l}_j^2}{s_1^2\tilde{l}_i^2 - s_2^2\tilde{l}_j^2}.\end{aligned}$$

By Lemma 3.4, positive and negative singular values appear in pairs. Therefore from Theorem 2.1 we immediately obtain the following theorem.

Theorem 4.1 *Let $d_1 = q(n-p+q) - 2$, $d_2 = (n-q)(p-q) - 2$. An estimator satisfying $\partial c_i/\partial s_i \ge 0$, $0 \le c_i \le 2d_i$, $i = 1, 2$, and*

$$\frac{c_1}{c_2} \ge \frac{\tilde{l}_{q+1}^2}{\tilde{l}_q^2}$$

is minimax.

We have considered fixed q and estimators of the form (1.3) which have the same shrinkage factors for the first q singular values and for the last $p-q$ singular values. For estimators treating each singular values separately, the unbiased estimator of risk and the result corresponding to Theorem 4.1 are given in Zheng (1986) based on results of Stein (1973).

Remark 4.1 Our derivation of the unbiased estimator of risk is similar to the method of Sheena (1995) and much easier than the method employed in Stein (1973), Zheng (1986), or Konno (1991a).

5. SOME DISCUSSION

In this section we give some discussion on possible extensions of the results of the present paper.

As mentioned several times, for simplicity we considered the case where the number of principal components q was fixed. More flexible approach is to treat each singular value separately. For example in the Efron-Morris estimator (Efron and Morris (1972))

$$\hat{M} = \sum_{i=1}^{p} (1 - \frac{n-p-1}{l_i^2}) l_i g_i h_i',$$

the rate of shrinkage is larger for the smaller singular values. Note also that by taking the positive part $(1-(n-p-1)/l_i^2)^+$ it is justified to ignore singular value l_i if $l_i^2 \leq n-p-1$. In order to treat each singular value separately in our general geometric framework, we need to investigate the structure of nesting of the manifolds $\mathcal{M}_q$ in (1.2) for $q = 1, \ldots, p-1$. This nesting has a remarkable structure that the composition of projection onto $\mathcal{M}_q$ followed by the projection from $\mathcal{M}_q$ onto $\mathcal{M}_{q-1}$ coincides with the direct projection onto $\mathcal{M}_{q-1}$. In this sense projections to $\mathcal{M}_q, q = 1, \ldots, p-1$, are similar to the projections to nested affine subspaces.

In usual practice of principal component analysis, larger singular values are not shrunk. We tried but did not succeed in justifying this practice from decision theoretic viewpoint. If we fix the first q singular values, then the last $p-q$ singular values are restricted to a bounded region. In general it seems very difficult to obtain minimax estimator when shrinkage is restricted to a bounded region. This causes a serious difficulty in justifying Stein shrinkage toward a general smooth manifold. In the setting of the present paper, the cone $\mathcal{C}(u, v)$ is unbounded and the boundedness problem can be avoided by shrinking both s_1 and s_2 as we approach the boundary $s_2 = s_1\gamma(u, v)$.

BIBLIOGRAPHY

Bilodeau, M. and Kariya, T. (1989). Minimax estimators in the normal MANOVA model. *J. Multivariate Anal.*, **28**, 260–270.

Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika*, **59**, 335–347.

Ghosh, M. and Shieh, G. (1991). Empirical Bayes minimax estimators of matrix normal means. *J. Multivariate Anal.*, **38**, 306–318.

Honda, T. (1991). Minimax estimators in the MANOVA model for arbitrary quadratic loss and unknown covariance matrix. *J. Multivariate Anal.*, **36**, 113–120.

James, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.*, **25**, 40–75.

Konno, Y. (1991a). On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.*, **36**, 44–55.

Konno, Y. (1991b). Families of minimax estimators of matrix of normal means with unknown covariance matrix. *J. Japan Statist. Soc.*, **20**, 191–201.

Kuriki, S. and Takemura, A. (1997a). James-Stein type estimator by shrinkage to closed convex sets with smooth boundary. *Discussion Paper 97-F-22*, Faculty of Economics, Univ. of Tokyo. Submitted for publication.

Kuriki, S. and Takemura, A. (1997b). Some geometry of the cone of nonnegative definite matrices and weights of associated $\bar{\chi}^2$ distribution. Submitted for publication.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Sheena, Y. (1995). Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix. *J. Japan Statist. Soc.*, **25**, 35–48.

Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proc. Prague Symp. Asymptotic Statistics*, 345–381.

Takemura, A. and Kuriki, S. (1995). Weights of $\bar{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives. *Discussion Paper 95-F-28*, Faculty of Economics, Univ. of Tokyo.

Weyl, H. (1939). On the volume of tubes. *Am. J. Math.*, **61**, 461–472.

Zheng, Z. (1986). On estimation of matrix of normal mean. *J. Multivariate Anal.*, **18**, 70–82.