# ABOUT PRINCIPAL COMPONENTS UNDER SINGULARITY

*José A. Díaz-García and Raúl Alberto Pérez-Agamez*

Comunicación Técnica No I-05-11/08-09-2005
(PE/CIMAT)

# About principal components under singularity

José A. Díaz-García

*Universidad Autónoma Agraria Antonio Narro*
*Department of Statistics and Computation*
*25315 Buenavista, Saltillo*
*Coahuila, MÉXICO*

Raúl Alberto Pérez-Agamez

*Universidad Nacional de Colombia, sede Medellín*
*Department of Statistics*
*Medellín, COLOMBIA.*

**Abstract**

In the present work some topics in the context of principal components are studied when the matrix of covariances is singular. In particular, the maximum likelihood estimates of the eigenvalues and the eigenvectors of the sample covariance matrix are obtained. Also, the likelihood ratio statistics for proving the hypothesis of sphericity and the equality of some eigenvalues are found.

*Key words:* Singular random matrix, sphericity test, likelihood ratio test, singular distribution, eigenvalue.
*PACS:* 62H15, 62H12, 62H25.

## 1 Introduction

Recently, several works have appeared in the context of singular random matrix distributions and compatible subjects; see Uhlig (1994), Díaz-García and Gutiérrez (1997), Díaz-García and González-Farías (1999), Díaz-García and Gutiérrez-Jáimez (2005), Díaz-García and González-Farías (2005a), Díaz-García and González-Farías (2005a), among others. Although, since 1968, Khatri had already proposed and solved several problems related with that

*Email addresses:* `jadiaz@uaaan.mx` (José A. Díaz-García),
`raperez@matematicas.udea.edu.co` (Raúl Alberto Pérez-Agamez ).

topic, in particular, about the multivariate general linear model, see Khatri (1968).

A central problem in the analysis of multivariate data is the reduction of the dimensionality, this is: if it is possible to represent, with certain precision, the values of $m$ variables, by a smaller subgroup of $k < m$ variables, then it would have been possible to reduce the dimension of the problem at the cost of a little loss of information. That is the object of principal components analysis. This is, given $N$ observations of $m$ variables, we ask for the possibility of representing sensibly this information with a smaller number of variables, which are defined as linear combinations of the original ones. When the study simply entails to make a descriptive analysis, the possible linear dependence between the variables is not significant. But when it is tried to make inference the situation changes completely.

Formally, if $\mathbf{X}' = (X_1 \vdots \cdots \vdots X_N)$ denotes the observation matrix $\mathbf{X} \in \Re^{N\times m}$, then $\mathbf{X}$ has a distribution respect to the Lebesgue measure $\Re^{N\times m} (\equiv \Re^{Nm})$. Nevertheless, if there is a linear dependence between the columns, for example $r < m$ are linearly independent, then $\mathbf{X}$ now has density with respect to the Hausdorff measure, see Díaz-García *et al.* (1997) or Díaz-García and González-Farías (2005a). Still more, if it is assumed that the sample comes from a population with $m-$dimensional normal ditribución, $\mathbf{X} \sim \mathcal{N}_{N\times m}^{N,r}(\mathbf{1}\boldsymbol{\mu}', I_N \otimes \mathbf{\Sigma})$, where $\otimes$ denotes the Kronecker product and $r$ denotes the rank of the matrix $\mathbf{\Sigma}$; i.e. if $\mathbf{X}$ has singular matrix normal distribution (see Díaz-García *et al.* (1997) and Khatri (1968)), whose density exists with respect to the Hausdorff measure; then, in such case, as much the density as the measure are not unique and their explicit forms will depend on the base and the set of coordinates selected for the subspace in where such density exists, see Díaz-García and González-Farías (2005a). An analogous situation appears with the distribution of the sample covariance matrix $S$, in that case $nS$ has a $m-$ dimensional singular Wishart distribution of rank $r$, $nS \sim \mathcal{W}_m^r((N-1), \mathbf{\Sigma})$ whose density also exists with respect to the Hausdorff measure defined in the corresponding subspace, see Díaz-García *et al.* (1997), Díaz-García and Gutiérrez (1997) and Díaz-García and González-Farías (2005b).

Under the condition of linear dependence between the variables, we will derive the maximum likelihood estimates of the eigenvalues and the eigenvectors of sample covariance matrix; also, we will find the joint distribution of the eigenvalues, see theorems, 1 and 2, respectively. The likelihood ratio statistics are obtained for proving the hypothesis of sphericity and the equality of some eigenvalores; see theorems 4 y 5, respectively.

## 2 Principal components

In this section some ideas on principal components are extended to the case in that the matrix of covariances is singular. But first let us consider the following notations:

The set of matrices $H_1 \in \Re^{m\times r}$ such that $H_1'H_1 = I_r$ is called the Stiefel manifold, and it is denoted as $\mathcal{V}_{r,m}$. In particular, $\mathcal{V}_{m,m}$ defines the group of orthogonal matrices, which we will denote by $\mathcal{O}(m)$.

Let us suppose a random sample $X_1,\dots,X_N$, $N = n+1$, such that $X_i \sim \mathcal{N}_m^r(\boldsymbol{\mu},\mathbf{\Sigma})$, $\mathbf{\Sigma} \geq \mathbf{0}$ where the rank of $\mathbf{\Sigma} = r$ and $\boldsymbol{\mu} \in \Re^m$ . Then the mean $\bar{X}$ and the matrix $S$ of sample covariances are given by

$$A = nS = \sum_{i=1}^{N}(X_i-\bar{X})(X_i-\bar{X})' = \mathbf{X}'\left(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N'\right)\mathbf{X} \tag{1}$$

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N}X_i \tag{2}$$

respectively; where $\mathbf{X'} = (X_1\vdots\cdots\vdots X_N)$ and $\mathbf{1}_N' = (1,\dots,1)$. This is, the maximum likelihood estimates in the singular case agrees with the estimates in the non singular case, see Rao (1973, p. 532). Something important to notice is that in the singular case the estimates $S$ and $\bar{X}$, no longer they exist with respect to the Lebesgue measure in $\Re^{m(m+1)/2}$ and $\Re^m$ respectively, but with respect to the measures of Hausdorff $(dS)$ and $(d\bar{X})$ defined in Díaz-García and González-Farías (2005a).

Now, let $l_1,\dots,l_r$ be the nonzero eigenvalues of $S$. Then, these are different with probability one and they estimate to the nonzero eigenvalues $\lambda_1 \geq \cdots \geq \lambda_r$ of $\mathbf{\Sigma}$. Besides, let $Q = (q_1\vdots\cdots\vdots q_m) = (Q_1\vdots Q_2) \in \mathcal{O}(m)$ such that $Q_1 \in \mathcal{V}_{r,m}$ and $Q_2 \in \mathcal{V}_{m-r,m}$ (a function of $Q_1$), with $q_i$ the normalized eigenvalues of $S$ such that

$$Q'SQ = \begin{pmatrix} Q_1' \\ Q_2' \end{pmatrix} S(Q_1\vdots Q_2) = \begin{pmatrix} Q_1'SQ_1 & Q_1'SQ_2 \\ Q_2'SQ_1 & Q_2'SQ_2 \end{pmatrix} = \begin{pmatrix} L_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

This is

$$S = Q\begin{pmatrix} L_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}Q'$$

with $L_r = \operatorname{diag}(l_1,\dots,l_r)$ or alternatively $S = Q_1L_rQ_1'$. Note that the eigenvectors $q_i$ are estimates of the eigenvectors $h_i$ of $\mathbf{\Sigma}$, where $H = (h_1\vdots\cdots\vdots h_m) =$

$(H_1 \vdots H_2) \in \mathcal{O}(m)$ such that

$$H'\mathbf{\Sigma}H = \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} \mathbf{\Sigma}(H_1 \vdots H_2) = \begin{pmatrix} H_1'\mathbf{\Sigma}H_1 & H_1'\mathbf{\Sigma}H_2 \\ H_2'\mathbf{\Sigma}H_1 & H_2'\mathbf{\Sigma}H_2 \end{pmatrix} = \begin{pmatrix} \Delta_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

with $H_1 \in \mathcal{V}_{r,m}$ and $H_2 \in \mathcal{V}_{m-r,m}$ (a function of $H_1$). Observe that if $h_{1j} \geq 0$ for every $j = 1, \dots, m$, then the representation of $\mathbf{\Sigma} = H_1 \Delta_r H_1'$ is unique if $\lambda_1 > \cdots > \lambda_r > 0$. Similarly with the probability one, the representation $S = Q_1 L_r Q_1'$ is unique if $q_{1j} \geq 0$ for every $j = 1, \dots, m$.

Assuming that does not exist multiplicity between the eigenvalores of the matrix $\mathbf{\Sigma}$, we have the following results.

**Theorem 1** *The maximum likelihood estimates of* $\Delta_r$ *and* $H_1$ *are* $\widehat{\Delta}_r = (n/N)L_r$ *and* $\widehat{H}_1 = (\widehat{h}_1, \dots, \widehat{h}_r) = Q_1$ *where* $\widehat{h}_i$ *is an eigenvector of the maximum likelihood estimate* $\widehat{\mathbf{\Sigma}} = (n/N)S$ *of* $\mathbf{\Sigma}$.

*Proof.* From Díaz-García *et al.* (1997) it is obtained that

$$\begin{aligned} dL(\boldsymbol{\mu}, \mathbf{\Sigma}) &= \frac{1}{(2\pi)^{Nr} \prod_{i=1}^{r} \lambda_i^{N/2}} \operatorname{etr}\left\{-\tfrac{1}{2}\mathbf{\Sigma}^-(\mathbf{X} - \mathbf{1}\boldsymbol{\mu}')'(\mathbf{X} - \mathbf{1}\boldsymbol{\mu})\right\}(d\mathbf{X}) \\ &= (2\pi)^{-Nr} \prod_{i=1}^{r} \lambda_i^{-N/2} \operatorname{etr}\left\{-\tfrac{1}{2}\mathbf{\Sigma}^-[A + N(\bar{X} - \boldsymbol{\mu})(\bar{X} - \boldsymbol{\mu})']\right\}(d\mathbf{X}), \end{aligned}$$

where $\operatorname{etr}(\cdot) = \exp(\operatorname{tr}(\cdot))$. So

$$L(\boldsymbol{\mu}, \mathbf{\Sigma}) = (2\pi)^{-Nr} \prod_{i=1}^{r} \lambda_i^{-N/2} \operatorname{etr}\left\{-\tfrac{1}{2}\mathbf{\Sigma}^- A - \frac{N}{2}(\bar{X} - \boldsymbol{\mu})'\mathbf{\Sigma}^-(\bar{X} - \boldsymbol{\mu})]\right\}.$$

Observing that for each $\mathbf{\Sigma}$, $L(\boldsymbol{\mu}, \mathbf{\Sigma})$ is maximum when $\boldsymbol{\mu} = \bar{X}$,

$$L(\bar{X}, \mathbf{\Sigma}) = (2\pi)^{-Nr} \prod_{i=1}^{r} \lambda_i^{-N/2} \operatorname{etr}\left\{-\frac{1}{2}\mathbf{\Sigma}^- A\right\}$$

Simply it reduces to maximize the function

$$h(\mathbf{\Sigma}) = \log L(\bar{X}, \mathbf{\Sigma}) = -Nr\log(2\pi) - \frac{N}{2}\sum_{i=1}^{r} \log \lambda_i - \frac{1}{2} \operatorname{tr}\left\{\mathbf{\Sigma}^- A\right\}.$$

Let us consider the spectral decompositions $\mathbf{\Sigma} = H_1 \Delta_r H_1'$ and $A = nS = nQ_1 L_r Q_1'$ with $H_1 \in \mathcal{V}_{r,m}$, $Q_1 \in \mathcal{V}_{r,m}$, $\Delta_r = \operatorname{diag}(\lambda_1, \dots, \lambda_r)$ and $L_r = \operatorname{diag}(l_1, \dots, l_r)$ and note that $\mathbf{\Sigma}^- = H_1 \Delta_r^{-1} H_1'$. Then, ignoring the constant,

$$h(\mathbf{\Sigma}) = -\frac{N}{2}\sum_{i=1}^{r} \log \lambda_i - \frac{n}{2} \operatorname{tr}\left\{H_1 \Delta_r^{-1} H_1' Q_1 L_r Q_1'\right\}$$

$$= -\frac{N}{2}\sum_{i=1}^{r}\log\lambda_i - \frac{n}{2}\operatorname{tr}\left\{\Delta_r^{-1}P'_{11}L_rP_{11}\right\} \tag{3}$$

where $P_{11} = Q'_1H_1$. Also, notice that

$$Q'H = \begin{pmatrix} Q'_1 \\ Q'_2 \end{pmatrix}(H_1 \vdots H_2) = \begin{pmatrix} Q'_1H_1 & Q'_1H_2 \\ Q'_2H_1 & Q'_2H_2 \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = (P_1 \vdots P_2) \tag{4}$$

with $P_1 \in \mathcal{V}_{r,m}$. Now, by the problem 9.4 in Muirhead (1982, p. 427) we have that

$$\operatorname{tr}(UP'_1VP_1) \leq \sum_{i=1}^{r} u_iv_i$$

where $U = \operatorname{diag}(u_1,\dots,u_r)$, $u_1 > \cdots > u_r > 0$; $V = \operatorname{diag}(v_1,\dots,v_r)$, $v_1 > \cdots > v_r > 0$ and $P_1 \in \mathcal{V}_{r,m}$, the equality holds only with the $2^r$ matrices $m \times r$ of the form

$$P_1 = \begin{pmatrix} P_{11} \\ P_{21} \end{pmatrix} = \begin{pmatrix} \pm I_r \\ \mathbf{0}_{m-r\times r} \end{pmatrix}.$$

Applying this result to the term where the trace appears in (3), with $U = \Delta_r^{-1}$ and $V = L_r$, we have that this term is maximized with respect to $P_{11}$ when $P_{11} = \pm I_r$. Then the maximum of (3) with respect to $P_{11}$ is

$$h(\lambda_i) = -\frac{N}{2}\sum_{i=1}^{r}\log\lambda_i - \frac{n}{2}\sum_{i=1}^{r}\frac{l_i}{\lambda_i}. \tag{5}$$

Note that, by (4) $QP = H$, i.e. $Q(P_1 \vdots P_2) = (H_1 \vdots H_2)$, then $\widehat{H}_1 = Q\widehat{P}_1$ is a maximum likelihood estimate of $H_1$. Note that in this case, $P_1 \in \mathcal{V}_{r,m}$ is any matrix such that $h_{1,j} \geq 0$ for every $j = 1,\dots,r$, but given that $Q = (Q_1 \vdots Q_2)$ is such that $q_{1,j} \geq 0$ for every $j = 1,\dots,r$; then $P_1$ can be taken such that $P'_1 = (I_r \vdots \mathbf{0})$, thus $\widehat{H}_1 = Q_1$ is a maximum likelihood estimate of $H_1$. In order to conclude the demonstration, we just differentiate (5) with respect to $\lambda_i$, $i = 1,\dots,r$ and we equal to zero, like this

$$\widehat{\lambda}_i = \frac{n\ l_i}{N},\ i = 1,\dots,r$$

then we get the desired result. ■

Now we are interested in finding the joint distribution of the eigenvalues $l_1,\dots,l_r$, $l_1 > \cdots > l_r > 0$; for it, consider the fact that $A = nS \sim \mathcal{W}_m^r(n,\mathbf{\Sigma})$, see Díaz-García *et al.* (1997) or Díaz-García and González-Farías (2005b).

**Theorem 2** *Let* $nS \sim \mathcal{W}_m^r(n, \mathbf{\Sigma})$, $n > m-1$, $\mathbf{\Sigma} \geq \mathbf{0}$ *of rank* $r$. *Then the joint density function of the eigenvalues*$l_1, \dots, l_r$ *of* $S$, *can be expressed in the form*

$$f_{l_i}(l_1, \dots, l_r) = \frac{\pi^{mr/2} \left(\dfrac{n}{2}\right)^{nr/2} \prod_{i=1}^r l_i^{(n+m-1)/2-r} \prod_{i<j}^r (l_i - l_j)}{\Gamma_r[n/2]\Gamma_r[m/2] \prod_{i=1}^r \lambda_i^{n/2}} {}_0F_0^m\left(-\frac{n}{2}L_r, \mathbf{\Sigma}^-\right),$$

*where* $L_r = \operatorname{diag}(l_1, \dots l_r)$, $l_1 > \cdots > l_r > 0$ *and*

$${}_0F_0^m\left(-\frac{n}{2}L_r, \mathbf{\Sigma}^-\right) = \sum_{k=1}^{\infty} \sum_{\kappa} \frac{C_\kappa\left(-\frac{n}{2}L_r\right) C_\kappa\left(\mathbf{\Sigma}^-\right)}{k! C_\kappa(I)}$$

*see Muirhead (1982, p. 259).*

*Proof.* The demonstration is followed from Díaz-García *et al.* (1997) or Díaz-García and González-Farías (2005b).
■

Next we propose the version for the singular case of the sphericity test

$$H_0 : \mathbf{\Sigma} = \lambda \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{6}$$

Observe that if this hypothesis is accepted we conclude that all the nonzero principal components have the same variance and they equally contribute to the total variance, therefore it is not possible to reduce the dimension of the problem. If the hypothesis is rejected it is possible, for example, that the $r-1$ smallest eigenvalues are equal, being able to reduce the dimension of the problem to the first principal component, and so on. In the same way we will have been interested in proving the following null hypothesis sequentially

$$H_{0,k} : \lambda_{k+1} = \cdots = \lambda_m$$

for $k = 0, 1, \dots, m-2$.

Given the importance of this topic and others in the statistics, next we propose the sphericity test for the singular case, but before it consider the following observation.

**Remark 3** *Note that in the singular case, when there is sphericity,* $\mathbf{\Sigma}$ *not* *necessarily has the form (6). In general it will have the form* $\mathbf{\Sigma} = \lambda\mathcal{I}$*, where*

$$\mathcal{I} = M' \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} M,$$

*and* $M \in \mathcal{O}(m)$ *is a permutation matrix (i.e in* $M'BM$*,* $M'$ *permutes the rows of* $B$ *and* $M$ *permutes the columns of* $B$*). Nevertheless, without loss of generality we can assume that* $\mathbf{\Sigma}$ *has the form (6), thus we have to consider the necessary permutations of rows and columns of* $\mathbf{\Sigma} = \lambda\mathcal{I}$ *to write it in the form (6), but now the action is over the matrix* $S$ *(or the matrix* $A$*), in such way that if* $S_1$ *is the original sample covariance matrix and now* $S$ *is defined as* $S = MS_1M'$*, then*

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

*with* $S_{11} > 0$ *and the rank* $(S_{11})=$ *rank* $(S_1)=$ *rank* $(S)$*. Note this does not change the likelihood function, because:*

(1) *If* $ch_i(A)$ *denotes the eigenvalores of matrix* $A$*,*

$$ch_i(\lambda\mathcal{I}) = ch_i(\lambda M\mathcal{I}M') = ch_i\left(\lambda \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right).$$

(2) *Also,*

$$\begin{aligned}
\operatorname{tr}\left(\lambda^{-1}\mathcal{I}^{-}S_1\right) &= \operatorname{tr}\left(\lambda^{-1}\mathcal{I}^{-}M'MS_1M'M\right) = \operatorname{tr}\left(\lambda^{-1}M\mathcal{I}^{-}M'MS_1M'\right) \\
&= \operatorname{tr}\left(\lambda^{-1}\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} S\right) \\
&= \operatorname{tr}\left(\lambda^{-1}(I_r \vdots \mathbf{0})S\begin{pmatrix} I_r \\ \mathbf{0} \end{pmatrix}\right) = \operatorname{tr}\lambda^{-1}S_{11} = \lambda^{-1}\sum_{i=1}^{r} l_i
\end{aligned}$$

*where* $l_1, \dots, l_r$ *are the nonzero eigenvalores of* $S$*.*

**Theorem 4** *Let* $X_1, \dots, X_N$ *be independent* $\mathcal{N}_m^r(\boldsymbol{\mu}, \mathbf{\Sigma})$ *random vectors. The likelihood ratio test of size* $\alpha$ *of*

$$H_0 : \mathbf{\Sigma} = \lambda \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

*with $\lambda$ is unspecified, rejects $H_0$ if*

$$V = \frac{\displaystyle\prod_{i=1}^{r} a_i}{\left(\displaystyle\frac{1}{r}\sum_{i=1}^{r} a_i\right)^r} \leq c_\alpha$$

*where $c_\alpha$ is chosen so that the size of the test is $\alpha$ and $a_i$ are the nonzero eigenvalues of $A$.*

*Proof.* The likelihood function is given by

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-Nr} \prod_{i=1}^{r} \lambda_i^{-N/2} \operatorname{etr}\left\{-\tfrac{1}{2}\boldsymbol{\Sigma}^{-}A - \frac{N}{2}(\bar{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-}(\bar{X}-\boldsymbol{\mu})]\right\}.$$

Taking into account the Remark 3 and denoting

$$\mathbb{I}_r = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} I_r \\ \mathbf{0} \end{pmatrix} (I_r \mathbf{0}),$$

the likelihood ratio is given by

$$\Lambda = \frac{\displaystyle\sup_{\boldsymbol{\mu}\in\Re^m,\ \lambda>0} L\left(\boldsymbol{\mu}, \lambda\mathbb{I}_r\right)}{\displaystyle\sup_{\boldsymbol{\mu}\in\Re^m,\ \boldsymbol{\Sigma}\geq 0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}. \tag{7}$$

Under the alternative hypothesis $\left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right) = \left(\bar{X}, \frac{1}{N}A\right)$, denoting the nonzero eigenvalues of $A$ by $a_i$, $i = 1, \ldots, r$ and eliminating the constant, we get

$$\begin{aligned}\sup_{\boldsymbol{\mu}\in\Re^m,\ \boldsymbol{\Sigma}\geq 0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= N^{Nr/2} \prod_{i=1}^{r} a_i^{-N/2} \operatorname{etr}\left\{-\frac{N}{2}A^{-}A\right\} \\ &= N^{Nr/2} \prod_{i=1}^{r} a_i^{-N/2} \exp\left\{-\frac{Nr}{2}\right\} \end{aligned} \tag{8}$$

but the matrix $A^{-}A$ represents a projection, then $\operatorname{tr}(A^{-}A) =$ rank of $(A^{-}A) = r$.

Now, eliminating the constant

$$\sup_{\boldsymbol{\mu}\in\Re^m,\ \lambda>0} L\left(\boldsymbol{\mu}, \lambda\mathbb{I}_r\right) = \sup_{\boldsymbol{\mu}\in\Re^m,\ \lambda>0} \lambda^{-Nr/2} \operatorname{etr}\left\{-\frac{1}{2\lambda}\mathbb{I}_r A - \frac{N}{2\lambda}(\bar{X}-\boldsymbol{\mu})'\mathbb{I}_r(\bar{X}-\boldsymbol{\mu})]\right\}$$

$$
\begin{aligned}
&= \sup_{\lambda>0} \lambda^{-Nr/2} \operatorname{etr}\left\{ -\frac{1}{2\lambda} \begin{pmatrix} I_r \\ \mathbf{0} \end{pmatrix} (I_r \mathbf{0}) A \right\} \\
&= \sup_{\lambda>0} \lambda^{-Nr/2} \operatorname{etr}\left\{ -\frac{1}{2\lambda} A_{11} \right\} \qquad (9)
\end{aligned}
$$

where

$$A_{11} = (I_r \vdots \mathbf{0}) A \begin{pmatrix} I_r \\ \mathbf{0} \end{pmatrix},$$

taking natural logarithm. And noticing that $\operatorname{tr} A_{11} = \sum_{i=1}^{r} a_i$, then we get

$$\mathcal{L}(\boldsymbol{\mu}, \lambda \mathbb{I}_r) = \log L(\boldsymbol{\mu}, \lambda \mathbb{I}_r) = -\frac{rN}{2} \log \lambda - \frac{1}{2\lambda} \sum_{i=1}^{r} a_i,$$

where the $a_i$ denote the nonzero eigenvalues of the matrix $A$, see Remak 3.

Thus, deriving and equaling to zero we get

$$\widehat{\lambda} = \frac{1}{Nr} \sum_{i=1}^{r} a_i.$$

Then

$$L\left(\bar{X}, \widehat{\lambda}\mathbb{I}_r\right) = \left( \frac{1}{Nr} \sum_{i=1}^{r} a_i \right)^{-Nr/2} \exp\left\{ -\frac{Nr}{2} \right\}. \qquad (10)$$

Using (8) and (10) in (7), we obtain

$$V = \Lambda^{2/N} = \frac{\displaystyle\prod_{i=1}^{r} a_i}{\left( \displaystyle\frac{1}{r} \sum_{i=1}^{r} a_i \right)^{r}}.$$

Then $H_0$ is rejected for small values of $V$. ■

Alternatively, observe that the statistics $V$ is given by

$$V = \frac{\displaystyle\prod_{i=1}^{r} l_i}{\left( \displaystyle\frac{1}{r} \sum_{i=1}^{r} l_i \right)^{r}},$$

where $l_1, \ldots, l_r$ are the nonzero eigenvalues of the matrix of covariances $S$.

Similarly to the nonsingular case, we can say that a test of asymptotic size alpha rejects $H_0$ if

$$-\left(n - \frac{2r^2+r+2}{6r}\right)\log V > c\left(\alpha; \tfrac{1}{2}(r+2)(r-1)\right),$$

where $c(\alpha; b)$ denotes the upper $100\alpha\%$ point of the $\chi^2$ distribution with $b$ degrees freedom, see Muirhead (1982, p. 406).

**Theorem 5** *Given a sample of size $N$ from $\mathcal{N}_m^r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, the likelihood ratio statistic for testing the null hypothesis*

$$H_{0,k} : \lambda_{k+1} = \cdots \lambda_r (= \lambda, \textit{unknoun}),\ k = 0, 1, \dots, (r-2)$$

*is $V_k = \Lambda^{2/N}$, where*

$$V_k = \frac{\displaystyle\prod_{i=k+1}^{r} l_i}{\left(\displaystyle\frac{1}{r-k}\sum_{i=k+1}^{r} l_i\right)^{r-k}}$$

*Proof.* The ratio of likelihood is given by

$$\Lambda = \frac{\displaystyle\sup_{H_{0,k}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\displaystyle\sup_{\boldsymbol{\mu}\in\Re^m,\ \boldsymbol{\Sigma}\geq 0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}. \tag{11}$$

Where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are not restricted. Then

$$\sup_{\boldsymbol{\mu}\in\Re^m,\ \boldsymbol{\Sigma}\geq 0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\frac{N}{n}\right)^{Nr/2} \prod_{i=1}^{r} l_i^{-N/2} \exp\left\{-\frac{Nr}{2}\right\}, \tag{12}$$

see the equation (8).

Using the same technique that in proof of Theorem 1, under the null hypothesis $H_{0,k}$ and ignoring the constant, we obtain

$$h(\lambda_i) = -\frac{N}{2}\sum_{i=1}^{r}\log\lambda_i - \frac{n}{2}\sum_{i=1}^{r}\frac{l_i}{\lambda_i}, \tag{13}$$

see equation (5).

But now

$$\Delta_r = \begin{pmatrix} \Delta_1 & \mathbf{0} \\ \mathbf{0} & \lambda I_{r-k} \end{pmatrix},$$

with $\Delta_1 = \text{diag}(\lambda_1, \dots, \lambda_k)$, $\lambda_1 > \cdots > \lambda_k > 0$, then (13) can be written as follows

$$h(\lambda_i, \lambda) = -\frac{N}{2}\sum_{i=1}^{k} \log \lambda_i - \frac{N(r-k)}{2} \log \lambda - \frac{n}{2}\sum_{i=1}^{k} \frac{l_i}{\lambda_i} - \frac{n}{2\lambda}\sum_{i=k+1}^{r} l_i. \qquad (14)$$

Differentiating (14) with respect to $\lambda_i$, $i = 1, \dots, k$ and $\lambda$ and equaling to zero such derivatives we get

$$\widehat{\lambda}_i = \frac{n}{N} l_i, \ i = 1, \dots, k \quad \text{and} \quad \widehat{\lambda} = \frac{n}{N}\frac{1}{(r-k)}\sum_{i=k+1}^{r} l_i.$$

Then, ignoring the constant,

$$\sup_{H_{0,k}} L\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \left(\frac{N}{n}\right)^{Nr/2} \left(\frac{1}{r-k}\sum_{i=k+1}^{r} l_i\right)^{-N(r-k)/2} \left(\prod_{i=1}^{k} l_i^{-N/2}\right) \exp\{-Nr/2\}.$$

Thus the statistics of likelihood ratio for testing $H_{0,k}$ is given by

$$V_k = \Lambda^{N/2} = \left[\frac{\displaystyle\sup_{H_{0,k}} L\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)}{\displaystyle\sup_{\boldsymbol{\mu}\in\Re^m,\ \boldsymbol{\Sigma}\geq 0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}\right] = \frac{\displaystyle\prod_{i=k+1}^{r} l_i}{\left(\displaystyle\frac{1}{r-k}\sum_{i=k+1}^{r} l_i\right)^{r-k}},$$

then $H_{0,k}$ is rejected for small values of $V_k$, and the proof is completed. ■

Finally, observe that by a similarity with the non singular case under the null hypothesis, the limiting distribution of the statistics

$$-\left(n - k - \frac{2p^2 + p + 2}{6p} + \sum_{i=1}^{k} \frac{\bar{l}_p^2}{(l_i - \bar{l}_p^2)}\right) \log V_k,$$

is $\chi^2$ with $(p+2)(p-1)/2$ degrees of freedom, where $p = r - k$ and $\bar{l}_p = \frac{1}{p}\sum_{i=k+1}^{r} l_i$, see Muirhead (1982, p. 409).

## 3 Acknowledgment

This work was supported partially by the research project 45974-F of CONACYT-México. Also, thanks to Francisco José Caro Lopera (CIMAT-México) for his

careful reading and excellent comments.

## References

J. A. Díaz-García, R. Gutiérrez-Jáimez, and K. V. Mardia, Wishart and Pseudo-Wishart distributions and some applications to shape theory, J. Multivariate Anal. 63 (1997) 73-87.

J. A. Díaz-García, and J. R. Gutiérrez, Proof of the conjectures of H. Uhlig on the singular multivariate beta and the jacobian of a certain matrix transformation, Ann. Statist. 25 (1997) 2018-2023.

J. A. Díaz-García, and G. González-Farías, QR, SV and Polar Decomposition and the Elliptically Contoured Distributions, Technical Report No. I-99-22 (PE/CIMAT) (1999), http://www.cimat.mx/biblioteca/RepTec.

J. A. Díaz-García, and G. González-Farías, Singular Random Matrix decompositions: Jacobians, J. Multivariate Anal. 93(2) (2005a) 196-212.

J. A. Díaz-García, and G. González-Farías, Singular Random Matrix decompositions: Distributions, J. Multivariate Anal. 94(1)(2005b) 109-122.

J. A. Díaz-García, and R. Gutiérrez-Jáimez, Distribution of the generalised inverse of a random matrix and its applications, J. Stat. Plann. Infer., (2005). In press. Also see, http://www.scincedirect.com.

C. G. Khatri, Some results for the singular normal multivariate regression nodels, Sankhyā A, 30 (1968) 267-280.

R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons, New York, 1982.

C. R. Rao, Linear Statistical Inference and its Applications, 2nd Editon, John Wiley & Sons, New York, 1973.

H. Uhlig, On singular Wishart and singular multivartiate beta distributions, Ann. Statistic. 22 (1994) 395-405.