Available online at www.sciencedirect.com

ScienceDirect

Journal of Multivariate Analysis 98 (2007) 932–944

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Maximum likelihood estimation of Wishart mean matrices under Löwner order restrictions

Ming-Tien Tsai*

Academia Sinica, Institute of Statistical Science, Taipei 11529, Taiwan, ROC

Received 12 July 2005
Available online 7 November 2006

Abstract

For Wishart density functions, there remains a long-time question unsolved. That is whether there exists the closed-form MLEs of mean matrices over the partially Löwner ordering sets. In this note, we provide an affirmative answer by demonstrating a unified procedure on exactly how the closed-form MLEs are obtained for the simple ordering case. Under the Kullback–Leibler loss function, a property of obtained MLEs is further studied. Some applications of the obtained closed-form MLEs, including the comparison between our ML estimates and Calvin and Dykstra's [Maximum likelihood estimation of a set of covariance matrices under Löwner order restrictions with applications to balanced multivariate variance components models, Ann. Statist. 19 (1991) 850–869.] which obtained by iterative algorithm, are also made.

AMS 2000 subject classification: 62H12; 62F30

Keywords: Exterior differential forms; Matrix factorizations; Kullback–Leibler loss function; Simple ordering set; Wishart density function

1. Introduction

Let $\mathbf{A}_i, i = 1, \ldots, k$, be k independent $p \times p$ matrices which are Wishart distributed with n_i $(\geqslant p)$ degrees of freedom and expectation $n_i\boldsymbol{\Sigma}_i$ being positive definite, denoted by $\mathbf{A}_i \sim W_p(n_i, \boldsymbol{\Sigma}_i), i = 1, \ldots, k$. Let

$$\mathbf{G}_i = n_i^{-1}\mathbf{A}_i, \quad i = 1, \ldots, k, \tag{1.1}$$

* Fax: +886 2 2783 1523.
E-mail address: mttsai@stat.sinica.edu.tw.

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.09.008

then the log-likelihood function of $\mathbf{G}_i$ $(i = 1, \ldots, k)$ can be expressed as

$$l(\mathbf{G}, \mathbf{\Sigma}) = \frac{1}{2}\sum_{i=1}^{k} n_i\{\ln|\mathbf{\Sigma}_i^{-1}\mathbf{G}_i| - \mathrm{tr}(\mathbf{\Sigma}_i^{-1}\mathbf{G}_i)\} + c, \tag{1.2}$$

where c is a constant in the sense that it is a function of $\mathbf{G} = (\mathbf{G}_1, \ldots, \mathbf{G}_k)'$ but it does not depend on $\mathbf{\Sigma} = (\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_k)'$ and $\mathrm{tr}(\mathbf{B})$ denotes the trace of matrix $\mathbf{B}$. One of the advantages of the way the objective function presented is that the information of the degrees of freedom associated with the individual mean square matrices is incorporated.

Let $\preccurlyeq$ be a partial order on the index set $\{1, \ldots, k\}$. The vector $\mathbf{\Sigma}$ is said to be isotonic with respect to $\preccurlyeq$ if it is order preserving in the Löwner sense. This means that if $j \preccurlyeq i$, then $\mathbf{\Sigma}_i - \mathbf{\Sigma}_j$ is positive semi-definite (p.s.d.), which is written as $\mathbf{\Sigma}_i \succcurlyeq \mathbf{\Sigma}_j$ throughout this paper. Define

$$\begin{aligned}\mathcal{K} = \{\mathbf{\Sigma} = (\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_k)' :\ & \mathbf{\Sigma} \text{ is partial order in the sense that there} \\ & \text{exists a permutation } (i_1, \ldots, i_k) \text{ of } (1, \ldots, k) \text{ such that} \\ & \mathbf{\Sigma}_{i_r} \succcurlyeq \mathbf{\Sigma}_{i_s} \text{ for every } r \geqslant s \text{ if } \mathbf{\Sigma}_{i_r} \text{ and } \mathbf{\Sigma}_{i_s} \text{ are comparable}\}.\end{aligned} \tag{1.3}$$

Obviously, $\mathcal{K}$ is a closed and convex cone. The main goal of this note is to find out the maximum likelihood estimator (MLE) of $\mathbf{\Sigma}$ which lies in $\mathcal{K}$. Note that if (i) $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_2\}$, (ii) $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_i, i = 2, \ldots, k\}$ and (iii) $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k\}$, the results for the problems can be easily applied to find the corresponding MLEs of unknown covariance matrices under the multinormal set up for (i) the completely balanced multivariate one-way random effects model, (ii) the completely balanced multivariate multi-way random effects models without interactions and (iii) the completely balanced multivariate multi-way random effects nested models, respectively.

The maximum likelihood estimation problems for multivariate random effects models and their related areas have been extensively studied for a long time. In the literature, beginning with Anderson [3], Morris and Olkin [11], Klotz and Putter [9], and Amemyia and Fuller [1] had studied these problems as well. Anderson et al. [2] first successfully obtained the MLEs of covariance matrices for the completely balanced multivariate one-way random effect model in which only two matrices are involved, namely for the case (i) above, $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_2\}$.

Calvin and Dykstra [5] pointed out that "it is not obvious what should be done when more than two covariance matrices are involved", and claimed that "this is a difficult optimization problem which cannot be solved in closed-form". Hence, they used the Fenchel duality techniques to develop a numerical iterative MLEs algorithm for balanced data when the models are with isotonic covariance structure. Two of the most well-known partially ordering sets in the literature are detailedly studied in their paper: one is the simple tree ordering set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_i, i = 2, \ldots, k\}$ and the other is the simple ordering set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k\}$. The closed-form MLEs of mean matrices over the simple tree ordering set had been obtained by Tsai [16], however, it remains unsolved for the simple ordering case which appears technically more difficult.

The approach of Tsai [16] is to simultaneously decompose the mean matrices into feasible components according to the structure of set $\mathcal{K}$, then make use of the property of concavity for the log-likelihood function of those matrix components to solve the problem. In this note, his approach is further incorporated to find the closed-form MLEs for the partially Löwner ordering set $\mathcal{K}$ defined in (1.3). To proceed with this approach, it is sufficient to consider only the simple ordering case because other cases can be parallelly handled as well. In Section 2, we demonstrate the unified procedure by working out the closed-form MLEs of mean matrices over the simple ordering set. Under the Kullback–Leibler loss function, a property of obtained MLEs is studied in the same section. Some applications and remarks of the results are also made in the final section.

2. Main results

For each partially Löwner ordering set $\mathcal{K}$ defined in (1.3), it is dissected to extract the inside information. The key idea is that we simultaneously decompose the mean matrices $\mathbf{\Sigma}_i$, $i = 1, \ldots, k$, into feasible components according to the structure of $\mathcal{K}$. Let $\mathbf{\Lambda}_i = \text{ch}(\mathbf{\Sigma}_i \mathbf{\Sigma}_{i-1}^{-1})$, $i = 2, \ldots, k$, where ch$(\mathbf{B})$ denotes the ordered diagonal matrix of eigenvalues of $\mathbf{B}$. Then by Theorem A9.9 of Muirhead [12], the pair $(\mathbf{\Sigma}_i, \mathbf{\Sigma}_{i+1})$ can be written as $\mathbf{\Sigma}_i = \mathbf{\Gamma}_i \mathbf{\Gamma}_i'$ and $\mathbf{\Sigma}_{i+1} = \mathbf{\Gamma}_i \mathbf{\Lambda}_{i+1} \mathbf{\Gamma}_i'$, where $\mathbf{\Gamma}_i \in \mathcal{N}(p)$, the group of $p \times p$ nonsingular matrices, $i = 1, \ldots, k-1$. Note that $\mathbf{\Gamma}_{i+1}\mathbf{\Gamma}_{i+1}' = \mathbf{\Gamma}_i \mathbf{\Lambda}_{i+1} \mathbf{\Gamma}_i'$, and thus by Theorem A9.5 of Muirhead [12], we have $\mathbf{\Gamma}_{i+1} = \mathbf{\Gamma}_i \mathbf{\Lambda}_{i+1}^{1/2} \mathbf{Q}_{i+1}$, $\forall i = 1, \ldots, k-1$, where $\mathbf{Q}_2 = \mathbf{I}$ and $\mathbf{Q}_j \in \mathcal{O}(p)$, the group of $p \times p$ orthogonal matrices, $j = 3, \ldots, k$.

Therefore, under the simple ordering set, we can simultaneously make the following decompositions:

$$\mathbf{\Sigma}_i = \mathbf{\Gamma}_1 \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{1/2} \mathbf{Q}_j \right) \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{1/2} \mathbf{Q}_j \right)' \mathbf{\Gamma}_1', \quad i = 1, \ldots, k, \tag{2.1}$$

where $\mathbf{\Lambda}_1 = \mathbf{Q}_1 = \mathbf{Q}_2 = \mathbf{I}$. The dimension of each $\mathbf{\Sigma}_i$ is $p(p+1)/2$, $i = 1, \ldots, k$; therefore, the total dimension of parameters is $kp(p+1)/2$. The new parameters in (2.1) are $\mathbf{\Gamma}_1, \mathbf{Q}_3, \ldots, \mathbf{Q}_k$, and $\mathbf{\Lambda}_2, \mathbf{\Lambda}_3, \ldots, \mathbf{\Lambda}_k$. The dimension of $\mathbf{\Gamma}_1$ is p^2; the dimension of each $\mathbf{Q}_i$ is $p(p-1)/2$; the dimension of each $\mathbf{\Lambda}_i$ is p; therefore, the total dimension of the new parameters is $p^2 + p(p-1)(k-2)/2 + p(k-1) = kp(p+1)/2$. Thus, we have one-to-one correspondence between $\{\mathbf{\Sigma}_i, i = 1, \ldots, k\}$ and $\{(\mathbf{\Gamma}_1, \mathbf{\Lambda}_i, \mathbf{Q}_3, \ldots, \mathbf{Q}_k), i = 2, \ldots, k\}$. Note that the technique of parameterizations in (2.1) for $\mathbf{\Sigma}_i$, $i = 1, \ldots, k$, is different from that of the simple tree ordering case (see Tsai [16]). Similarly, for the sample counterparts

$$\mathbf{G}_i = \mathbf{W}_1 \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2} \mathbf{V}_j \right) \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2} \mathbf{V}_j \right)' \mathbf{W}_1', \tag{2.2}$$

$i = 1, \ldots, k$, where $\mathbf{F}_1 = \mathbf{V}_1 = \mathbf{V}_2 = \mathbf{I}$, $\mathbf{F}_i = \text{ch}(\mathbf{G}_i \mathbf{G}_{i-1}^{-1})$, $i = 2, \ldots, k$, $\mathbf{W}_1 \in \mathcal{N}(p)$ (with probability one), and $\mathbf{V}_j \in \mathcal{O}(p)$ (with probability one), $j = 3, \ldots, k$.

For the sake of manipulations, we adopt the idea of Anderson et al. [2] to make the following transformation:

$$\mathbf{H} = \mathbf{\Gamma}_1^{-1} \mathbf{W}_1. \tag{2.3}$$

Then note that $\mathbf{H} \in \mathcal{N}(p)$ with probability one, $\mathbf{\Gamma}_1$ and $\mathbf{H}$ are one-to-one correspondence with probability one.

The maximization for this problem can be similarly proceeded as that for the simple tree ordering case, and hence we present only the necessary steps and the details are omitted. Let $\mathbf{Q} = (\mathbf{Q}_3, \ldots, \mathbf{Q}_k)'$ and $\mathbf{\Lambda} = (\mathbf{\Lambda}_2, \ldots, \mathbf{\Lambda}_k)'$, thus by (1.2) and (2.1)–(2.3) we have

$$\sup_{\mathcal{K}} \sum_{i=1}^{k} n_i \{\ln |\mathbf{\Sigma}_i^{-1} \mathbf{G}_i| - \text{tr}(\mathbf{\Sigma}_i^{-1} \mathbf{G}_i)\} = \sup_{\left\{ \substack{\mathbf{\Lambda}_i \geqslant \mathbf{I} \\ i=2,\ldots,k}, \substack{\mathbf{Q}_j \in \mathcal{O}(p) \\ j=3,\ldots,k}, \mathbf{H} \in \mathcal{N}(p) \right\}} \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q}), \tag{2.4}$$

where the log-likelihood function

$$\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q}) = \sum_{i=1}^{k} n_i \left\{ \ln |\mathbf{HH}'| + \ln \left| \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1} \mathbf{F}_j \right) \right| - \operatorname{tr} \left[\left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1/2} \mathbf{Q}_j \right) \right.\right.$$
$$\left.\left. \times \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1/2} \mathbf{Q}_j \right)' \mathbf{H} \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2} \mathbf{V}_j \right) \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2} \mathbf{V}_j \right)' \mathbf{H}' \right] \right\}. \tag{2.5}$$

Write $\mathbf{H} = ((h_{ij}))$ and let $\mathcal{E} = \operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_p)$ with $\varepsilon_i = 1$ or -1, according to h_{ii} is nonnegative or negative, $\forall i = 1, \ldots, p$, then $\ln |\mathcal{E}\mathbf{HH}'\mathcal{E}| = \ln |\mathbf{HH}'|$, $\mathcal{E}\mathbf{Q}_3 \in \mathcal{O}(p)$ and $\mathcal{E}\mathbf{A}\mathcal{E}$ is still positive definite if $\mathbf{A}$ is positive definite. If $\mathbf{H}$ is replaced by $\mathcal{E}\mathbf{H}$, what has to be changed for the log-likelihood function $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ in (2.5) is that $\mathbf{Q}_3$ be replaced by the orthogonal matrix $\mathbf{Q}_3^*$ $(= \mathcal{E}\mathbf{Q}_3)$. Moreover, $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ is symmetric in $\mathbf{H}$, and hence without loss of generality we assume that $\mathbf{H} \in \mathcal{N}^*(p)$, where $\mathcal{N}^*(p) = \{\mathbf{H} = ((h_{ij})) : \mathbf{H} \in \mathcal{N}(p),\ |\mathbf{H}| > 0$ and $h_{ii} \geqslant 0, i = 1, \ldots, p\}$. Further, Tsai [16] used the notion of exterior differential forms to show the following lemma:

Lemma 1. *Let* $\mathbf{A}, \mathbf{B}$ *be positive definite matrices, and* $\mathcal{N}^*(p) = \{\mathbf{H} = ((h_{ij})) : \mathbf{H} \in \mathcal{N}(p),$ $|\mathbf{H}| > 0,$ *and* $h_{ii} \geqslant 0, i = 1, \ldots, p\}$, *where* $\mathcal{N}(p)$ *is the group of* $p \times p$ *nonsingular matrices. If* $\mathbf{H} \in \mathcal{N}^*(p)$, *then* (i) $\operatorname{tr}(\mathbf{AHBH}')$ *is strictly convex in* $\mathbf{H}$, *and* (ii) $|\mathbf{HH}'|$ *is strictly logconcave in* $\mathbf{HH}'$.

With similar proof as those in Lemma 1, it can be shown that the log-likelihood function $l(\mathbf{G}_i, \mathbf{\Sigma}_i, 1 \leqslant i \leqslant k)$ defined in (1.2) is continuous and strictly concave in $\mathbf{\Sigma}_i^{-1}$, $i = 1, \ldots, k$, on the space of positive definite matrices, and thus the MLE of each $\mathbf{\Sigma}_i$, $i = 1, \ldots, k$, is unique over the convex cone $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k\}$. In passing, we may note that it is one-to-one correspondence between $\mathbf{\Sigma}$ and $(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$. The log-likelihood function $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ attains its maximum value at $(\widehat{\mathbf{H}}, \widehat{\mathbf{\Lambda}}, \widehat{\mathbf{Q}})$ which satisfies the partial differential equations of $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ with respect to $\mathbf{H}$, $\mathbf{\Lambda}_i$, $i = 2, \ldots, k$, and $\mathbf{Q}_j$, $j = 3, \ldots, k$, respectively.

Note that when $k = 2$, the convex cone in (1.3) reduces to the simplest case $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_2\}$. For this simplest case, Anderson et al. [2] showed that the maximum value of $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ over the set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_2\}$ can occur only when $\mathbf{H} \in \mathcal{D}(p)$, where $\mathcal{D}(p)$ denotes the group of diagonal matrices with positive elements. It is believed that the more restricted ordering set (i.e., the more restricted cone $\mathcal{K}$) is, the more restricted solution of $\mathbf{H}$ is. Tsai [16, p. 296] further showed that the maximum value of the log-likelihood function ($\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$) over the simple tree ordering set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_i, i = 2, \ldots, k\}$ occurs only when $\mathbf{H} \in \mathcal{D}(p)$. With similar arguments as those in Tsai [16] for the simple tree ordering case, we can also show that the maximum value of $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ over the simple ordering set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k\}$ occurs only when $\mathbf{H} \in \mathcal{D}(p)$. We will adopt another simpler method to claim that $\mathbf{H} \in \mathcal{D}(p)$ is one of the sufficient conditions of the estimation equations with respect to $\mathbf{Q}$ in Proposition 1.

Note that, as $\mathbf{H}$ approaches the boundaries while $\mathbf{Q}$ and $\mathbf{\Lambda}$ are fixed, $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q}) \to -\infty$. Thus, in order to maximize $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ over the set $\{\mathbf{H} : \mathbf{H} \in \mathcal{N}^*(p)\}$, one needs to examine the first derivative equation of $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ with respect to $\mathbf{H}$. Recall that $d \operatorname{tr}(\mathbf{HAH}'\mathbf{B}) = \operatorname{tr}[(\mathbf{AH}'\mathbf{B} + \mathbf{A}'\mathbf{H}'\mathbf{B}')(d\mathbf{H})]$ and $d|\mathbf{H}| = |\mathbf{H}| \operatorname{tr}[(\mathbf{H}^{-1})(d\mathbf{H})]$, and hence the partial differential of $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$

with respect to **H** gives

$$d\ell_0(\mathbf{H},\mathbf{\Lambda},\mathbf{Q}) = 2\,\mathrm{tr}\left\{\left[n\mathbf{H}^{-1} - \sum_{i=1}^{k} n_i \left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)\left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)'\mathbf{H}' \times \left(\prod_{j=1}^{i}\mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)\left(\prod_{j=1}^{i}\mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)'\right](d\mathbf{H})\right\} \tag{2.6}$$

for the exterior product $(d\mathbf{H})$, where $n = \sum_{i=1}^{k} n_i$. Thus, after some straightforward manipulations, $d\ell_0(\mathbf{H},\mathbf{\Lambda},\mathbf{Q}) = 0$, for the exterior product $(d\mathbf{H}) \neq 0$, leads to

$$\sum_{i=1}^{k} n_i \left(\prod_{j=1}^{i}\mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)'\mathbf{H}\left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)\left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)'\mathbf{H}'\left(\prod_{j=1}^{i}\mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right) = n\mathbf{I}. \tag{2.7}$$

This is the corresponding estimation equation with respect to the parameter **H** (i.e., $\partial\ell_0(\mathbf{H},\mathbf{\Lambda},\mathbf{Q})/\partial\mathbf{H} = \mathbf{0}$). Other estimation equations with respect to $\mathbf{\Lambda}_i$, (i.e., $\partial\ell_0(\mathbf{H},\mathbf{\Lambda},\mathbf{Q})/\partial\mathbf{\Lambda}_i = \mathbf{0}$), $i = 2,\ldots,k$, and $\mathbf{Q}_j$ (i.e., $\partial\ell_0(\mathbf{H},\mathbf{\Lambda},\mathbf{Q})/\partial\mathbf{Q}_j = \mathbf{0}$), $j = 3,\ldots,k$, respectively, can be similarly obtained.

To directly solve those matrix estimation equations seems to be intractable, one way to overcome the difficulty is to tactfully impose the compatible conditions so that the log-likelihood function can be further simplified. Let

$$\mathbf{Q}_i'\mathbf{H}\left(\prod_{j=1}^{i}\mathbf{\Lambda}_j^{-1/2}\mathbf{F}_j^{1/2}\right) = \mathbf{H}\left(\prod_{j=1}^{i}\mathbf{\Lambda}_j^{-1/2}\mathbf{F}_j^{1/2}\right)\mathbf{V}_i', \quad i = 3,\ldots,k. \tag{2.8}$$

Moreover, we assume that $\mathbf{H} \in \mathcal{D}(p)$, which is the group of diagonal matrices with positive elements. Note that the dimensions of parameter space will not be reduced when the system of equations (2.8) and $\mathbf{H} \in \mathcal{D}(p)$ are imposed. The system of equations (2.8) and $\mathbf{H} \in \mathcal{D}(p)$ are called the compatible conditions if the solutions of (2.4) are not changed when they are imposed. The main advantage of imposing those compatible conditions is that the optimization procedure will become much easier. To see that, first note that by the system of equations (2.8) and $\mathbf{H} \in \mathcal{D}(p)$ we have

$$\begin{aligned}
&\left(\prod_{j=1}^{3}\mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)'\mathbf{H}\left(\prod_{j=1}^{3}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)\left(\prod_{j=1}^{3}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)'\mathbf{H}'\left(\prod_{j=1}^{3}\mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)\\
&= (\mathbf{Q}_3'\mathbf{H}\mathbf{\Lambda}_2^{-1/2}\mathbf{F}_2^{1/2}\mathbf{\Lambda}_3^{-1/2}\mathbf{F}_3^{1/2}\mathbf{V}_3)(\mathbf{Q}_3'\mathbf{H}\mathbf{\Lambda}_2^{-1/2}\mathbf{F}_2^{1/2}\mathbf{\Lambda}_3^{-1/2}\mathbf{F}_3^{1/2}\mathbf{V}_3)'\\
&= (\mathbf{H}\mathbf{\Lambda}_2^{-1/2}\mathbf{F}_2^{1/2}\mathbf{\Lambda}_3^{-1/2}\mathbf{F}_3^{1/2}\mathbf{V}_3'\mathbf{V}_3)(\mathbf{H}\mathbf{\Lambda}_2^{-1/2}\mathbf{F}_2^{1/2}\mathbf{\Lambda}_3^{-1/2}\mathbf{F}_3^{1/2}\mathbf{V}_3'\mathbf{V}_3)'\\
&= \mathbf{H}^2\prod_{j=1}^{3}(\mathbf{\Lambda}_j^{-1}\mathbf{F}_j).
\end{aligned} \tag{2.9}$$

Assume that $(\prod_{j=1}^{l-1} \mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j)'\mathbf{H}(\prod_{j=1}^{l-1} \mathbf{F}_j^{1/2}\mathbf{V}_j) = \mathbf{H}(\prod_{j=1}^{l-1} \mathbf{\Lambda}_j^{-1/2}\mathbf{F}_j^{1/2})$ holds for any l, $4 \leqslant l \leqslant i$, then

$$\begin{aligned}
&\left(\prod_{j=1}^{l} \mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)' \mathbf{H} \left(\prod_{j=1}^{l} \mathbf{F}_j^{1/2}\mathbf{V}_j\right) \\
&= \mathbf{Q}_l'\mathbf{\Lambda}_l^{-1/2} \left(\prod_{j=1}^{l-1} \mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)' \mathbf{H} \left(\prod_{j=1}^{l-1} \mathbf{F}_j^{1/2}\mathbf{V}_j\right) \mathbf{F}_l^{1/2}\mathbf{V}_l \\
&= \mathbf{Q}_l'\mathbf{H} \left(\prod_{j=1}^{l} \mathbf{\Lambda}_j^{-1/2}\mathbf{F}_j^{1/2}\right) \mathbf{V}_l \\
&= \mathbf{H} \left(\prod_{j=1}^{l} \mathbf{\Lambda}_j^{-1/2}\mathbf{F}_j^{1/2}\right). \qquad (2.10)
\end{aligned}$$

Thus, by mathematical induction, we may conclude that

$$\begin{aligned}
&\left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right)' \mathbf{H} \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2}\mathbf{V}_j\right) \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2}\mathbf{V}_j\right)' \mathbf{H}' \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1/2}\mathbf{Q}_j\right) \\
&= \mathbf{H}^2 \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1}\mathbf{F}_j\right), \quad i = 3, \ldots, k. \qquad (2.11)
\end{aligned}$$

Therefore, the matrix estimation equation (2.7) reduces to

$$\mathbf{H}^2 \sum_{i=1}^{k} n_i \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1}\mathbf{F}_j\right) = n\mathbf{I}. \qquad (2.12)$$

On the other hand, by virtue of (2.11), the log-likelihood function (2.5) can be further simplified to

$$\begin{aligned}
\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q}) &= \sum_{i=1}^{k} n_i \left\{ \ln|\mathbf{H}^2| + \ln\left|\left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1}\mathbf{F}_j\right)\right| - \operatorname{tr}\left[\mathbf{H}^2 \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1}\mathbf{F}_j\right)\right]\right\} \\
&= \ell_1(\mathbf{H}, \mathbf{\Lambda}), \quad \text{say}, \qquad (2.13)
\end{aligned}$$

with the solutions should satisfy the conditions that $\mathbf{H} \in \mathcal{D}(p)$ and the system of equations (2.8) holds. Set the partial derivative of $\ell_1(\mathbf{H}, \mathbf{\Lambda})$ with respect to $\mathbf{H}$ to be zero, then we get the same matrix estimation equation as in (2.12).

By the system of equations (2.8), we may note that $\mathbf{Q}_j$ is the function of $(\mathbf{H}, \mathbf{\Lambda})$, $j = 3, \ldots, k$. Thus, we have

$$\frac{\partial \ell_1(\mathbf{H}, \mathbf{\Lambda})}{\partial \mathbf{H}} = \frac{\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})}{\partial \mathbf{H}} + \sum_{j=3}^{k} \frac{\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})}{\partial \mathbf{Q}_j} \frac{\partial \mathbf{Q}_j}{\partial \mathbf{H}} \tag{2.14}$$

and

$$\frac{\partial \ell_1(\mathbf{H}, \mathbf{\Lambda})}{\partial \mathbf{\Lambda}_i} = \frac{\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})}{\partial \mathbf{\Lambda}_i} + \sum_{j=3}^{k} \frac{\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})}{\partial \mathbf{Q}_j} \frac{\partial \mathbf{Q}_j}{\partial \mathbf{\Lambda}_i}, \quad i = 2, \ldots, k. \tag{2.15}$$

Furthermore, as mentioned earlier that under the conditions that $\mathbf{H} \in \mathcal{D}(p)$ and the system of equations (2.8) holds, the log-likelihood function $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ in (2.5) can be reduced to the form in (2.13) which is free of $\mathbf{Q}$, and hence it is obvious to see that $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{Q}_j = \mathbf{0}$, $j = 3, \ldots, k$. Thus, we may conclude that $\mathbf{H} \in \mathcal{D}(p)$ and the system of equations (2.8) holds are the sufficient conditions of estimation equations $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{Q}_j = \mathbf{0}$, $j = 3, \ldots, k$. Therefore, by (2.14) and (2.15), we have the following.

Proposition 1. *Let* $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ *and* $\ell_1(\mathbf{H}, \mathbf{\Lambda})$ *be defined as in* (2.5) *and* (2.13), *respectively. Also let* $\mathcal{D}(p)$ *denote the group of diagonal matrices with positive elements. If the systems of equations* (2.8), $\partial \ell_1(\mathbf{H}, \mathbf{\Lambda})/\partial \mathbf{H} = \mathbf{0}$, $\partial \ell_1(\mathbf{H}, \mathbf{\Lambda})/\partial \mathbf{\Lambda}_i = \mathbf{0}$, $i = 2, \ldots, k$, *and the condition that* $\mathbf{H} \in \mathcal{D}(p)$ *hold, then the systems of equations* $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{H} = \mathbf{0}$, $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{\Lambda}_i = \mathbf{0}$, $i = 2, \ldots, k$, *and* $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{Q}_j = \mathbf{0}$, $j = 3, \ldots, k$, *hold.*

Due to the unique solution $(\widehat{\mathbf{H}}, \widehat{\mathbf{\Lambda}}, \widehat{\mathbf{Q}})$ of the systems of estimation equations $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{H} = \mathbf{0}$, $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{\Lambda}_i = \mathbf{0}$, $i = 2, \ldots, k$, and $\partial \ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})/\partial \mathbf{Q}_j = \mathbf{0}$, $j = 3, \ldots, k$, we may conclude that the original problem of maximizing the log-likelihood function $\ell_0(\mathbf{H}, \mathbf{\Lambda}, \mathbf{Q})$ under the set $\{ {}_{i=2,\ldots,k}^{\mathbf{\Lambda}_i \geqslant \mathbf{I}}, {}_{j=3,\ldots,k}^{\mathbf{Q}_j \in \mathcal{O}(p)}, \mathbf{H} \in \mathcal{N}(p)\}$ can be performed through the problem of maximizing the log-likelihood function $\ell_1(\mathbf{H}, \mathbf{\Lambda})$ under the conditions that $\mathbf{H} \in \mathcal{D}(p)$, $\mathbf{\Lambda}_i \geqslant \mathbf{I}$, $i = 2, \ldots, k$, and $\mathbf{Q}_j \in \mathcal{O}(p)$, $j = 3, \ldots, k$, such that the system of equations (2.8) holds.

Now, we start to proceed the maximization problem under the new setup with the help of the system of equations (2.8) and $\mathbf{H} \in \mathcal{D}(p)$. Substitute (2.12) into (2.13), thus $\ell_1(\mathbf{H}, \mathbf{\Lambda})$ becomes

$$n \ln n - np + \sum_{i=2}^{k} n_i \ln \left| \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1} \mathbf{F}_j \right) \right| - n \ln \left| \sum_{i=1}^{k} n_i \left(\prod_{j=1}^{i} \mathbf{\Lambda}_j^{-1} \mathbf{F}_j \right) \right| = \ell_2(\mathbf{\Lambda}), \quad \text{say.} \tag{2.16}$$

Rewrite $\ell_2(\mathbf{\Lambda})$ as

$$\begin{aligned} & n \ln n - np + \sum_{i=1}^{k-2} \left(\sum_{j=1}^{i} n_j \right) \ln |\mathbf{\Lambda}_{i+1} \mathbf{F}_{i+1}^{-1}| + n_k \ln |\mathbf{\Lambda}_k^{-1} \mathbf{F}_k| \\ & - n \ln \left| \sum_{i=1}^{k-2} n_i \left(\prod_{j=i+1}^{k-1} \mathbf{\Lambda}_j \mathbf{F}_j^{-1} \right) + n_{k-1} \mathbf{I} + n_k \mathbf{\Lambda}_k^{-1} \mathbf{F}_k \right| = \ell_3(\mathbf{\Lambda}), \quad \text{say.} \end{aligned} \tag{2.17}$$

We then maximize $\ell_3(\mathbf{\Lambda})$ subject to the conditions that $\mathbf{\Lambda}_i \succcurlyeq \mathbf{I}$, $i = 2, \ldots, k$. Let $\mathbf{T} = (\mathbf{T}_2, \ldots, \mathbf{T}_k)'$, where $\mathbf{T}_i = \mathrm{diag}(t_{i1}, \ldots, t_{ip})$ with $\mathbf{T}_i \succcurlyeq \mathbf{0}$ and $\mathbf{T}_i(\mathbf{\Lambda}_i - \mathbf{I}) = \mathbf{0}$, $i = 2, \ldots, k$. Define the Lagrangian function

$$\ell_4(\mathbf{\Lambda}, \mathbf{T}) = -\ell_3(\mathbf{\Lambda}) - \sum_{i=2}^{k} \mathrm{tr}[\mathbf{T}_i(\mathbf{\Lambda}_i - \mathbf{I})]. \tag{2.18}$$

Then set the partial derivatives of $\ell_4(\mathbf{\Lambda}, \mathbf{T})$ with respect to $\mathbf{\Lambda}_i$ and $\mathbf{T}_i$, $i = 2, \ldots, k$, to be zero, we obtain that

$$\mathbf{T}_k\mathbf{\Lambda}_k = n_k\mathbf{I} - nn_k\mathbf{\Lambda}_k^{-1}\mathbf{F}_k \left[\sum_{i=1}^{k-2} n_i \left(\prod_{j=i+1}^{k-1} \mathbf{\Lambda}_j\mathbf{F}_j^{-1} \right) + n_{k-1}\mathbf{I} + n_k\mathbf{\Lambda}_k^{-1}\mathbf{F}_k \right]^{-1} \tag{2.19}$$

and

$$\mathbf{T}_i\mathbf{\Lambda}_i = -\left(\sum_{j=1}^{i-1} n_j \right)\mathbf{I} + n\sum_{j=1}^{i-1} n_j \left(\prod_{m=j+1}^{k-1} \mathbf{\Lambda}_m\mathbf{F}_m^{-1} \right) \left[\sum_{i=1}^{k-2} n_i \left(\prod_{j=i+1}^{k-1} \mathbf{\Lambda}_j\mathbf{F}_j^{-1} \right) \right.$$
$$\left. + n_{k-1}\mathbf{I} + n_k\mathbf{\Lambda}_k^{-1}\mathbf{F}_k \right]^{-1}, \quad i = 2, \ldots, k-1. \tag{2.20}$$

A Kuhn–Tucker–Lagrange (KTL) point is any point $(\mathbf{\Lambda}^*, \mathbf{T}^*)$ which satisfies the following conditions: (i) $\mathbf{T}_i \succcurlyeq \mathbf{0}$, (ii) $\mathbf{\Lambda}_i \succcurlyeq \mathbf{I}$, (iii) $\mathrm{tr}[\mathbf{T}_i(\mathbf{\Lambda}_i - \mathbf{I})] = 0$, and (iv) the system of equations (2.19) and (2.20). Let $\mathbf{A} = ((a_{ij}))$ and $\mathbf{B} = ((b_{ij}))$ be any two matrices, and denote $\max\{\mathbf{A}, \mathbf{B}\} = ((\max(a_{ij}, b_{ij})))$. Then after some algebraic manipulations, by KTL point formula theorem (Hadley [6]) the MLE of $\mathbf{\Lambda}_k$ is

$$\widehat{\mathbf{\Lambda}}_k = \max\{\mathbf{F}_k, \mathbf{I}\}. \tag{2.21}$$

Substitute (2.21) back into (2.17) and repeat the above processes until $k = 2$, we finally obtain the following recurrence formula:

$$\widehat{\mathbf{\Lambda}}_{k-i} = \max\left\{ \left(\sum_{j=k-i}^{k} n_j \right)^{-1} \mathbf{F}_{k-i} \left[n_{k-i}\mathbf{I} + \sum_{j=k-i+1}^{k} n_j \left(\prod_{m=k-i+1}^{j} \widehat{\mathbf{\Lambda}}_m^{-1}\mathbf{F}_m \right) \right], \mathbf{I} \right\},$$
$$i = 1, \ldots, k-2. \tag{2.22}$$

By Lemma 1 and the fact that $\mathbf{H} \in \mathcal{D}(p)$, it is obvious to see that $\ell_1(\mathbf{H}, \mathbf{\Lambda})$ is continuous and strictly concave in $\mathbf{H}$ on the space $\mathcal{D}(p)$, and hence, the MLE of $\mathbf{H}$ is unique. Substitute (2.21) and (2.22) into (2.12), we then have

$$\widehat{\mathbf{H}} = n^{1/2} \left[n_1\mathbf{I} + \sum_{i=2}^{k} n_i \left(\prod_{j=2}^{i} \widehat{\mathbf{\Lambda}}_j^{-1}\mathbf{F}_j \right) \right]^{-1/2}. \tag{2.23}$$

By the system of matrix equations (2.8) and the results of (2.21)–(2.23), then

$$\widehat{\mathbf{Q}}_i' = \widehat{\mathbf{H}} \left(\prod_{j=1}^{i} \widehat{\mathbf{\Lambda}}_j^{-1/2}\mathbf{F}_j^{1/2} \right) \mathbf{V}_i' \left(\prod_{j=1}^{i} \widehat{\mathbf{\Lambda}}_j^{1/2}\mathbf{F}_j^{-1/2} \right) \widehat{\mathbf{H}}^{-1}, \quad i = 3, \ldots, k. \tag{2.24}$$

Therefore, by virtue of (2.1), (2.3) and (2.24), the MLEs of $\mathbf{\Sigma}_i$ over the simple ordering set are of the forms

$$\widehat{\mathbf{\Sigma}}_i = \mathbf{W}_1\left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)\widehat{\mathbf{H}}^{-1}\left(\prod_{j=2}^{i}\widehat{\mathbf{\Lambda}}_j\mathbf{F}_j^{-1}\right)\widehat{\mathbf{H}}^{-1}\left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)'\mathbf{W}_1', \quad i = 1, \ldots, k. \tag{2.25}$$

It is easy to note that (i) when $k = 2$,

$$\widehat{\mathbf{\Sigma}}_1 = \widehat{\mathbf{\Gamma}}_1\widehat{\mathbf{\Gamma}}_1' = \mathbf{W}_1[n^{-1}(n_1\mathbf{I} + n_2\min\{\mathbf{F}_2, \mathbf{I}\})]\mathbf{W}_1'$$

and

$$\widehat{\mathbf{\Sigma}}_2 = \widehat{\mathbf{\Gamma}}_1\widehat{\mathbf{\Lambda}}_2\widehat{\mathbf{\Gamma}}_1' = \mathbf{W}_1[n^{-1}(n_1\max\{\mathbf{F}_2, \mathbf{I}\} + n_2\mathbf{F}_2)]\mathbf{W}_1', \tag{2.26}$$

as expected they are exactly the same MLEs as obtained by Anderson et al. [2], and (ii) when $\mathbf{F}_i \succcurlyeq \mathbf{I}$, $\forall i = 2, \ldots, k$, then $\widehat{\mathbf{\Lambda}}_i = \mathbf{F}_i$, $i = 2, \ldots, k$, and hence $\widehat{\mathbf{H}} = \mathbf{I}$. Thus, the obtained MLEs of $\mathbf{\Sigma}_i$ reduce to the unrestricted ones $\mathbf{G}_i$, $i = 1, \ldots, k$.

We summarize the main results of this section in the following:

Theorem 1. *Let $\mathbf{A}_i$, $i = 1, \ldots, k$, be k independent $p \times p$ matrices which are Wishart distributed with n_i ($\geqslant p$) degrees of freedom and expectation $n_i\mathbf{\Sigma}_i$, where each $\mathbf{\Sigma}_i$ is positive definite. Let $\mathbf{G}_i = n_i^{-1}\mathbf{A}_i$ and make the decompositions $\mathbf{G}_i = \mathbf{W}_1(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j)(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j)'\mathbf{W}_1'$, $i = 1, \ldots, k$,* where $\mathbf{F}_1 = \mathbf{V}_1 = \mathbf{V}_2 = \mathbf{I}$, $\mathbf{W}_1 \in \mathcal{N}(p)$, *the group of nonsingular matrices, with probability one*, $\mathbf{V}_j \in \mathcal{O}(p)$, *the group of orthogonal matrices, with probability one*, $j = 3, \ldots, k$, *and* $\mathbf{F}_i = \mathrm{ch}(\mathbf{G}_i\mathbf{G}_1^{-1})$, $i = 2, \ldots, k$, *with* $\mathrm{ch}(\mathbf{B})$ *denoting the ordered diagonal matrix of eigenvalues of* $\mathbf{B}$. *Then the MLEs of $\mathbf{\Sigma}_i$, $i = 1, \ldots, k$, over the simple ordering set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k\}$ are of the forms*

$$\widehat{\mathbf{\Sigma}}_i = \mathbf{W}_1\left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)\widehat{\mathbf{H}}^{-1}\left(\prod_{j=2}^{i}\widehat{\mathbf{\Lambda}}_j\mathbf{F}_j^{-1}\right)\widehat{\mathbf{H}}^{-1}\left(\prod_{j=1}^{i}\mathbf{F}_j^{1/2}\mathbf{V}_j\right)'\mathbf{W}_1', \quad i = 1, \ldots, k,$$

where $\widehat{\mathbf{H}} = n^{1/2}[n_1\mathbf{I} + \sum_{i=2}^{k} n_i(\prod_{j=2}^{i}\widehat{\mathbf{\Lambda}}_j^{-1}\mathbf{F}_j)]^{-1/2}$ *with* $n = \sum_{i=1}^{k} n_k$, $\widehat{\mathbf{\Lambda}}_k = \max\{\mathbf{F}_k, \mathbf{I}\}$ *and* $\widehat{\mathbf{\Lambda}}_{k-i} = \max\{(\sum_{j=k-i}^{k} n_j)^{-1}\mathbf{F}_{k-i}[n_{k-i}\mathbf{I} + \sum_{j=k-i+1}^{k} n_j(\prod_{m=k-i+1}^{j}\widehat{\mathbf{\Lambda}}_m^{-1}\mathbf{F}_m)], \mathbf{I}\}$, $i = 1, \ldots, k-2$.

Next, we study a property of MLEs $\widehat{\mathbf{\Sigma}}_i$, $i = 1, \ldots, k$, over the simple ordering set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k\}$. Consider the Kullback–Leibler loss function

$$L(\widehat{\mathbf{\Sigma}}^*, \mathbf{\Sigma}) = \sum_{i=1}^{k} n_i\{\mathrm{tr}(\widehat{\mathbf{\Sigma}}_i^*\mathbf{\Sigma}_i^{-1}) - \ln|\widehat{\mathbf{\Sigma}}_i^*\mathbf{\Sigma}_i^{-1}| - p\}, \tag{2.27}$$

where $\widehat{\mathbf{\Sigma}}^* = (\widehat{\mathbf{\Sigma}}_1^*, \ldots, \widehat{\mathbf{\Sigma}}_k^*)'$. When $k = 1$, then (2.27) reduces to the Stein [14] loss function, and it also reduces to the loss function considered by Loh [10] when $k = 2$. The Kullback–Leibler loss function can be derived from the Kullback–Leibler distance for the joint density function of

$\mathbf{A}_i, i = 1, \ldots, k$. Let $\mathcal{E}_{\mathbf{\Sigma}\in\mathcal{K}} L(\widehat{\mathbf{\Sigma}}^*, \mathbf{\Sigma})$ be the Kullback–Leibler risk over the simple ordering set $\mathcal{K}$. Note that

$$\mathcal{E}_{\mathbf{\Sigma}\in\mathcal{K}} L(\widehat{\mathbf{\Sigma}}^*, \mathbf{\Sigma}) \geqslant \mathcal{E}_{\mathbf{\Sigma}\in\mathcal{K}} \left[\inf_{\mathbf{\Sigma}\in\mathcal{K}} L(\widehat{\mathbf{\Sigma}}^*, \mathbf{\Sigma}) \right]. \tag{2.28}$$

Then take $\widehat{\mathbf{\Sigma}}_i^* = a_i \widehat{\mathbf{\Sigma}}_i, a_i > 0, \forall i = 1, \ldots, k$, by similar arguments as those of maximization in the proof of Theorem 1, we obtain that

$$\begin{aligned} \inf_{\mathbf{\Sigma}\in\mathcal{K}} L(\widehat{\mathbf{\Sigma}}^*, \mathbf{\Sigma}) &= \sum_{i=1}^k n_i \{\mathrm{tr}(a_i \mathbf{I}) - \ln |a_i \mathbf{I}| - p\} \\ &= \sum_{i=1}^k n_i \{p a_i - p \ln a_i - p\} \\ &\geqslant 0. \end{aligned} \tag{2.29}$$

The minimum value of the right-hand side of (2.29) occurs when $a_i = 1, \forall i = 1, \ldots, k$. Therefore, we have the following.

Theorem 2. *Under the same set-up as in Theorem* 1, *the Kullback–Leibler risk of* $(a_1\widehat{\mathbf{\Sigma}}_1, \ldots, a_k\widehat{\mathbf{\Sigma}}_k)'$ *is minimized at* $a_i = 1, \forall i = 1, \ldots, k$ *over the simple ordering set* $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k\}$.

Theorem 2 generalizes the results presented in Theorem 7.8.1 of Anderson [4], which deals with the one-sample problem. When $k = 2$, Srivastava and Kubokawa [13] studied the risk dominance problems of the MLEs with respect to the unbiased estimators and to some of Stein-type improved estimators (for details see Haff [8], Loh [10] and the references therein) relative to the Kullback–Leibler loss (2.27). However, being different from the unrestricted case which was studied by Loh [10], the minimax problem of estimation over the set $\mathcal{K} = \{\mathbf{\Sigma} : \mathbf{\Sigma}_1 \preccurlyeq \mathbf{\Sigma}_2\}$ still remains open in the literature. For the risk dominance problems when $k \geqslant 3$, some further techniques, such as extending the Stein–Haff Wishart identity for one-sample and two-sample problems (Stein [15], Haff [7] and Loh [10]) to k-sample problems that involve more than three matrices under the partially ordering sets, are needed to be developed. However, this turns out to be a quite challenging problem.

3. Applications

For the directly applications of Theorem 1, first we note that it can be applied to find the MLEs of covariance matrices for the completely balanced multivariate random effects nested models. Calvin and Dykstra [5] used the completely balanced multivariate two-way random effects nested model to analyze the data set of patterns care studies, and gave the numerical ML estimates via their iterated algorithm as

$$\hat{\mathbf{\Sigma}}_{cdA} = \begin{bmatrix} 512.24 & 343.25 \\ 343.25 & 230.01 \end{bmatrix}, \quad \hat{\mathbf{\Sigma}}_{cdB} = \begin{bmatrix} 58.76 & 66.96 \\ 66.96 & 449.96 \end{bmatrix} \quad \text{and}$$

$$\hat{\mathbf{\Sigma}}_{cdE} = \begin{bmatrix} 255.57 & 47.96 \\ 47.96 & 63.49 \end{bmatrix},$$

respectively. The numerical computations for exact MLEs of $\mathbf{\Sigma}_i, i = 1, 2, 3$, in Theorem 1 are easily implemented. Incorporating the computational algorithm in Tsai [16] by using the

MATHEMATICA, it is easy to see that

$$\mathbf{F}_2 = \begin{bmatrix} 19.9798 & 0 \\ 0 & 1.42723 \end{bmatrix}, \quad \mathbf{F}_3 = \begin{bmatrix} 6.57006 & 0 \\ 0 & 0.613848 \end{bmatrix},$$

$$\mathbf{W}_1 = \begin{bmatrix} 0.713791 & -15.9706 \\ 7.50831 & -2.66744 \end{bmatrix}$$

and

$$\mathbf{V}_3 = \begin{bmatrix} -0.292128 & -0.95638 \\ 0.95638 & -0.292128 \end{bmatrix}, \quad \widehat{\mathbf{\Lambda}}_2 = \begin{bmatrix} 19.9798 & 0 \\ 0 & 1.17286 \end{bmatrix},$$

$$\widehat{\mathbf{\Lambda}}_3 = \begin{bmatrix} 6.57006 & 0 \\ 0 & 1 \end{bmatrix}, \quad \widehat{\mathbf{H}} = \mathbf{I}.$$

By Theorem 1, we then have

$$\widehat{\mathbf{\Sigma}}_1 = \begin{bmatrix} 255.57 & 47.96 \\ 47.96 & 63.49 \end{bmatrix}, \quad \widehat{\mathbf{\Sigma}}_2 = \begin{bmatrix} 373.076 & 181.891 \\ 181.891 & 963.394 \end{bmatrix} \quad \text{and}$$

$$\widehat{\mathbf{\Sigma}}_3 = \begin{bmatrix} 2422.01 & 1554.86 \\ 1554.86 & 1883.4 \end{bmatrix},$$

respectively. Thus, the ML estimates of three unknown covariance matrices (the strata effect $\mathbf{\Sigma}_A = (\mathbf{\Sigma}_3 - \mathbf{\Sigma}_2)/4$, the facility effect $\mathbf{\Sigma}_B = (\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1)/2$ and the random error $\mathbf{\Sigma}_E = \mathbf{\Sigma}_1$) are

$$\widehat{\mathbf{\Sigma}}_A = \begin{bmatrix} 512.233 & 343.242 \\ 343.242 & 230.003 \end{bmatrix}, \quad \widehat{\mathbf{\Sigma}}_B = \begin{bmatrix} 58.7531 & 66.9653 \\ 66.9653 & 449.952 \end{bmatrix} \quad \text{and}$$

$$\widehat{\mathbf{\Sigma}}_E = \begin{bmatrix} 255.57 & 47.96 \\ 47.96 & 63.49 \end{bmatrix},$$

respectively. Compute the values of the log-likelihood function, the expression in (1.2) after ignoring the constant c, the result is -27.188159 based on our estimates and it is -27.188161 based on Calvin and Dykstra's. Note that $|\widehat{\mathbf{\Sigma}}_A| = -3.13091 \times 10^{-11} \approx 0$ but $|\widehat{\mathbf{\Sigma}}_{cdA}| = -0.2401$, it is obvious that the ML estimates obtained by Calvin and Dykstra's numerical algorithm are slightly out of the restricted parameter space.

Theorem 1 can also be directly applied to obtain the likelihood ratio test statistic for the problem of testing $H_0 : \mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$ against $H_1 : \mathbf{\Sigma}_1 \preccurlyeq \cdots \preccurlyeq \mathbf{\Sigma}_k$ under the setup of Theorem 1. Under the null hypothesis, the log-likelihood function is maximized with respect to $\mathbf{\Sigma}$, and the MLE of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}} = n^{-1} \sum_{i=1}^{k} n_i \mathbf{G}_i$. By virtue of (1.2) and Theorem 1, the maximum of the log-likelihood functions under H_0 and H_1 are

$$\frac{1}{2} \left\{ \sum_{i=1}^{k} n_i \ln \left| \left(\sum_{i=1}^{k} n_i \mathbf{G}_i \right)^{-1} \mathbf{G}_i \right| + n \ln n - np \right\} + c, \tag{3.1}$$

and

$$\frac{1}{2} \left\{ \sum_{i=1}^{k} n_i \ln \left| \left(\prod_{j=2}^{i} \widehat{\mathbf{\Lambda}}_j^{-1} \mathbf{F}_j \right) \right| - n \ln \left| n_1 \mathbf{I} + \sum_{i=2}^{k} n_i \left(\prod_{j=2}^{i} \widehat{\mathbf{\Lambda}}_j^{-1} \mathbf{F}_j \right) \right| + n \ln n - np \right\} + c, \tag{3.2}$$

respectively. Thus the likelihood ratio criterion for testing H_0 against H_1 is based on $L = e^{-\frac{1}{2}\lambda}$, where

$$\lambda = \sum_{i=1}^{k} n_i \left\{ \ln \left| \left(\prod_{j=2}^{i} \widehat{\mathbf{\Lambda}}_j^{-1} \mathbf{F}_j \right) \right| - \ln \left| \left(\sum_{i=1}^{k} n_i \mathbf{G}_i \right)^{-1} \mathbf{G}_i \right| \right\}$$

$$-n \ln \left| n_1 \mathbf{I} + \sum_{i=2}^{k} n_i \left(\prod_{j=2}^{i} \widehat{\mathbf{\Lambda}}_j^{-1} \mathbf{F}_j \right) \right|$$

$$= \sum_{i=1}^{k} n_i \ln \left| \left[n_1 \mathbf{I} + \sum_{i=2}^{k} n_i \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2} \mathbf{V}_j \right) \left(\prod_{j=1}^{i} \mathbf{F}_j^{1/2} \mathbf{V}_j \right)' \right] \prod_{j=1}^{i} \widehat{\mathbf{\Lambda}}_j^{-1} \right|$$

$$-n \ln \left| n_1 \mathbf{I} + \sum_{i=2}^{k} n_i \left(\prod_{j=2}^{i} \widehat{\mathbf{\Lambda}}_j^{-1} \mathbf{F}_j \right) \right|. \quad (3.3)$$

The critical region is

$$\lambda \geqslant \lambda_\alpha, \quad (3.4)$$

where λ_α is defined so that (3.4) holds with probability α when H_0 is true. To find the value of critical point λ_α, it involves $(k-2)$-fold integral over orthogonal groups with respect to the normalized Haar invariant measure on the space of orthogonal $p \times p$ matrices, and then the zonal polynomials as well as the invariant polynomials with matrix arguments might play important roles for finding out the (asymptotic) distribution theories of λ.

The techniques developed in Section 2 can be applied to obtain the closed-form (restricted) MLEs for factor analysis models. And it can also be parallelly applied to obtain the closed-form least estimators by minimizing the quadratic-type loss $\sum_{i=1}^{k} n_i \operatorname{tr}(\mathbf{\Sigma}_i^{-1}\mathbf{G}_i - \mathbf{I})^2$ over the partially Löwner ordering sets.

Acknowledgments

The author is grateful to the reviewer and the Editors for their constructive and helpful comments. He is also indebted to Shuenn-Jyi Sheu for his helpful discussions.

References

[1] Y. Amemiya, W. Fuller, Estimation for the multivariate errors-in-variables model with estimated error covariance matrix, Ann. Statist. 12 (1984) 497–509.
[2] B.M. Anderson, T.W. Anderson, I. Olkin, Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance, Ann. Statist. 14 (1986) 405–417.
[3] T.W. Anderson, Analysis of multivariate variance, unpublished manuscript, 1946.
[4] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, second ed., Wiley, New York, 1984.
[5] J.A. Calvin, R.L. Dykstra, Maximum likelihood estimation of a set of covariance matrices under Löwner order restrictions with applications to balanced multivariate variance components models, Ann. Statist. 19 (1991) 850–869.
[6] G. Hadley, Non-linear and Dynamic Programming, Addison-Wesley, Reading, MA, 1964.
[7] L.R. Haff, An identity for the Wishart distribution with applications, J. Multivariate Anal. 9 (1979) 531–544.
[8] L.R. Haff, The variational form of certain Bayes estimators, Ann. Statist. 19 (1991) 1163–1190.

[9] J.H. Klotz, J. Putter, Maximum likelihood estimation of multivariate covariance components for the balanced one-way layout, Ann. Math. Statist. 40 (1969) 1100–1105.
[10] W.L. Loh, Estimating covariance matrices, Ann. Statist. 19 (1991) 283–296.
[11] B. Morris, I. Olkin, Some estimation and testing problems for factor analysis models, unpublished manuscript, 1964.
[12] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[13] M.S. Srivastava, T. Kubokawa, Improved nonnegative estimation of multivariate components of variance, Ann. Statist. 27 (1999) 2008–2032.
[14] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, in: J. Neyman (Ed.), Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. I, University of California, Berkeley, 1956, pp. 197–206.
[15] C. Stein, Lectures on multivariate estimation theory, in: Investigation on Statistical Estimation Theory, vol. I, 1977, pp. 4–65, Zapiski Nauchych Seminarov LOMI im. V.A. Steklova ANSSSR 74, Leningrad (in Russian).
[16] M.T. Tsai, Maximum likelihood estimation of covariance matrices under simple tree ordering, J. Multivariate Anal. 89 (2004) 292–303.