

# Article A New Class of Bayes Minimax Estimators of the Mean Matrix of a Matrix Variate Normal Distribution

Shokofeh Zinodiny<sup>1</sup> and Saralees Nadarajah<sup>2,\*</sup>

- <sup>1</sup> Department of Mathematics, Amirkabir University of Technology, Tehran 15916-34311, Iran; shokofeh\_zinodiny@yahoo.com or shokofeh@aut.ac.ir
- <sup>2</sup> Department of Mathematics, University of Manchester, Manchester M13 9PL, UK
- \* Correspondence: mbbsssn2@manchester.ac.uk

**Abstract:** Bayes minimax estimation is important because it provides a robust approach to statistical estimation that considers the worst-case scenario while incorporating prior knowledge. In this paper, Bayes minimax estimation of the mean matrix of a matrix variate normal distribution is considered under the quadratic loss function. A large class of (proper and generalized) Bayes minimax estimators of the mean matrix is presented. Two examples are given to illustrate the class of estimators, showing, among other things, that the class includes classes of estimators presented by Tsukuma.

Keywords: Bayes estimation; matrix variate normal distribution; mean matrix; minimax estimation

**MSC:** 62E99

## 1. Introduction

Let  $\mathbf{X} = (x_{i,j})$  be a  $p \times m$  matrix random variable with a matrix variate normal distribution with mean matrix  $\mathbf{\Theta} = (\theta_{i,j})$  and covariance matrix  $\mathbf{I}_p \otimes \mathbf{I}_m$ , where  $\mathbf{I}_k$  is the  $k \times k$  identity matrix and  $\otimes$  denotes the Kronecker product.

The matrix variate normal distribution finds applications across various fields, including multivariate statistical analysis, machine learning, and signal processing. In multivariate statistical analysis, it serves as a fundamental tool for modeling covariance structures in datasets where the observations are matrices, such as in longitudinal studies or multivariate time series analysis. In machine learning, it is utilized for modeling complex dependencies among high-dimensional data, particularly in tasks involving matrix-valued inputs or outputs, such as in recommender systems or tensor factorization. Moreover, in signal processing, the matrix variate normal distribution is employed for modeling the joint distribution of multiple correlated signals or images, enabling efficient estimation and inference in applications such as array processing or medical imaging.

Some recent applications of the matrix variate normal distribution include analysis of multiple vector autoregressions [1]; brain connectivity alternation detection [2]; capacity for severely fading MIMO channels [3]; integrated principal components analysis [4]; determination of the relationship between incidence and mortality of asthma with PM2.5, ozone, and household air pollution [5]; autism spectrum disorder identification [6]; and identification of depression disorder using multi-view high-order brain function networks [7], to mention just a few.

Bayesian minimax estimation is a statistical approach that combines Bayesian inference with minimax decision theory. In traditional Bayesian inference, we use prior knowledge and observed data to update our beliefs about the parameters of interest. Minimax decision theory, on the other hand, focuses on minimizing the maximum possible loss (risk) that can occur under different parameter values.

In Bayesian minimax estimation, we seek an estimator that minimizes the maximum possible posterior expected loss, where the expectation is taken with respect to the posterior



**Citation:** Zinodiny, S.; Nadarajah, S. A New Class of Bayes Minimax Estimators of the Mean Matrix of a Matrix Variate Normal Distribution. *Mathematics* **2024**, *12*, 1098. https:// doi.org/10.3390/math12071098

Academic Editor: Manuel Alberto M. Ferreira

Received: 4 January 2024 Revised: 24 February 2024 Accepted: 4 April 2024 Published: 5 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). distribution of the parameter given the observed data. This approach is particularly useful when there is uncertainty about the true parameter value and when it is important to protect against worst-case scenarios.

There has not been much work on Bayesian estimation of the parameters of the matrix variate normal distribution. Ref. [8] extended the so-called Stein effect and proposed an empirical Bayes estimator, outperforming the maximum likelihood estimator,  $\mathbf{X}$ , for the case m > p + 1. Since then, many classes of minimax estimators better than the maximum likelihood estimator have been found. Ref. [9] derived a large class of unbiased risk estimators, including a class of minimax estimators obtained by [8]. Using the result of Stein, ref. [10] extended the results of [11] to the multivariate case. For the case of  $\Sigma \otimes I_m$ , where  $\Sigma$  is an unknown positive definite matrix and p > m + 1, ref. [12] introduced a class of minimax estimators containing those of [8]. Ref. [13] derived a large class of minimax estimators using the Stein identity and the Haff identity [14] for the case m > p + 1. For the case of  $\Sigma = I_m$ , Ref. [15] found orthogonally invariant hierarchical priors, resulting in Bayes estimators that are admissible and minimax. For the case of an unknown covariance matrix, Ref. [16] obtained a generalized Bayes class of minimax estimators of the mean matrix for m > p + 1, p > m + 1. Ref. [17] obtained Bayes minimax estimators of the mean for the case of common unknown variance. Ref. [18] obtained Bayes minimax estimators of the normal mean matrix for the case of common unknown variances.

For the problem of estimating the mean matrix of an elliptically contoured distribution, Ref. [19] derived generalized Bayes minimax estimators for the mean matrix; ref. [20] also obtained a class of minimax estimators for the mean matrix, which was used to find a class of proper Bayes minimax estimators of  $\Theta$ .

In this paper, we derive a large class of (proper and generalized) Bayes minimax estimators of  $\Theta$  containing estimators of [15] as a special case. In fact, we extend the results of [21] to the multivariate case. The main result, giving a large class of (proper and generalized) Bayes minimax estimators, is developed in Section 2. Section 3 considers two examples of classes of (proper and generalized) Bayes estimators. In particular, example 1 demonstrates a result from [15]. Some concluding remarks are given in Section 4.

Throughout this paper, let  $|\mathbf{A}|$ , tr( $\mathbf{A}$ ) and  $\mathbf{A}'$  denote, respectively, the determinant, trace and transpose of a matrix  $\mathbf{A}$ . Also for  $\mathbf{A}$  and  $\mathbf{B}$ , let  $\mathbf{B} < \mathbf{A}$  mean that  $\mathbf{A} - \mathbf{B}$  is positive definite.

## 2. A Class of Bayes Minimax Estimators of the Mean Matrix

Let  $N_{p \times m}(\Theta, \mathbf{I}_p \otimes \mathbf{I}_m)$  denote the matrix variate normal distribution with mean matrix  $\Theta$  and covariance matrix  $\mathbf{I}_p \otimes \mathbf{I}_m$ . Assume that  $\mathbf{X} \sim N_{p \times m}(\Theta, \mathbf{I}_p \otimes \mathbf{I}_m)$ . Assume also that

$$\Theta|\Lambda \sim N_{p \times m} \left( \mathbf{0}_{p \times m}, \Lambda^{-1} (\mathbf{I}_p - \Lambda) \otimes \mathbf{I}_m \right), \ \Lambda \sim |\Lambda|^{\frac{a}{2} - 1} g\left(\frac{tr(\Lambda)}{p}\right), \ \mathbf{0}_{p \times p} < \Lambda < \mathbf{I}_p, \ a > -m,$$
(1)

where  $\mathbf{\Lambda} = (\lambda_{i,j})$  is a  $p \times p$  matrix distributed as g, and g is a differentiable positive function on (0, 1). In addition, assume g is such that

$$\pi(\mathbf{\Theta}) = (2\pi)^{-\frac{pm}{2}} \int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} |\mathbf{I}_{p} - \mathbf{\Lambda}|^{-\frac{m}{2}} e^{-\frac{tr\left((\mathbf{I}_{p} - \mathbf{\Lambda})^{-1}\mathbf{\Lambda}\mathbf{\Theta}\mathbf{\Theta}'\right)}{2}} d\mathbf{\Lambda},$$
(2)

where  $d\mathbf{\Lambda} = \bigwedge_{i < j}^{p} d\lambda_{i,j}$ . Note that (2) will be proper if *g* is integrable on its domain.

The purpose of this section is to construct generalized (and proper) Bayes minimax estimators of  $\Theta$  under the loss function

$$L(\boldsymbol{\delta};\boldsymbol{\Theta}) = tr((\boldsymbol{\delta}-\boldsymbol{\Theta})(\boldsymbol{\delta}-\boldsymbol{\Theta})').$$
(3)

The following lemmas give sufficient conditions on g and a such that the generalized (or proper) Bayes estimators with respect to (2) are minimax.

Let  $\mathbf{O}_p$  be the set of orthogonal matrices of order p. Let  $\mathbf{V}_{m,p} = {\mathbf{V} \in \mathbf{R}^{m \times p}, \mathbf{V}'\mathbf{V} = \mathbf{I}_p}$ , where  $m \ge p$ . Write  $\mathbf{X}$  as  $\mathbf{ULV}'$ , where  $\mathbf{U} \in \mathbf{O}_p$ ,  $\mathbf{V} \in \mathbf{V}_{m,p}$  and  $\mathbf{L} = \text{diag}(l_1, l_2, \dots, l_p)$  with  $l_1 > l_2 > \dots > l_p > 0$ .

**Lemma 1.** For i = 1, ..., p, write  $\phi_i = \phi_i(\mathbf{F})$  and  $\mathbf{F} = \text{diag}(f_1, ..., f_p) = \mathbf{L}^2$ . The risk of a shrinkage equivariant estimator  $\delta = \mathbf{UL}(\mathbf{I}_p - \mathbf{\Phi}(\mathbf{F}))\mathbf{V}'$  is

$$R(\boldsymbol{\delta};\boldsymbol{\Theta}) = mp + E\left[\sum_{i=1}^{p} \left\{ f_i \phi_i^2 - 2(m-p+1)\phi_i - 4f_i \frac{\partial \phi_i}{\partial f_i} - 4\sum_{j>i} \frac{f_i \phi_i - f_j \phi_j}{f_i - f_j} \right\}\right]$$
(4)

provided each expectation exists.

#### Proof. See [9].

If  $\mathbf{\Phi}(\mathbf{F}) = \mathbf{F}^{-1} \mathbf{\Psi}(\mathbf{F})$ , where  $\mathbf{\Psi}(\mathbf{F}) = \text{diag}(\psi_1(\mathbf{F}), \dots, \psi_p(\mathbf{F}))$ , then by replacing  $\phi_i$  by  $\frac{\psi_i}{f_i}$ , (4) can be written as

$$R(\boldsymbol{\delta};\boldsymbol{\Theta}) = mp + E\left[\sum_{i=1}^{p} \left\{\frac{\psi_i^2}{f_i} - 2(m-p-1)\frac{\psi_i}{f_i} - 4f_i\left(\frac{\frac{\partial\psi_i}{\partial f_i}}{f_i}\right) - 4\sum_{j>i}\frac{\psi_i - \psi_j}{f_i - f_j}\right\}\right].$$
 (5)

Using (5), we obtain Corollary 1.  $\Box$ 

Corollary 1. Suppose

$$\boldsymbol{\delta} = \left(\mathbf{I}_p - \mathbf{U}\mathbf{F}^{-1}\boldsymbol{\Psi}(\mathbf{F})\mathbf{U}'\right)\mathbf{X},\tag{6}$$

where  $\mathbf{F}^{-1} = \text{diag}(f_1^{-1}, f_2^{-1}, \dots, f_p^{-1})$ . Then,  $\delta$  is minimax under (3) if

- *I.* For any *i*,  $\psi_i$  is non-decreasing with respect to  $f_i$ ;
- II.  $0 \leq \psi_p \leq \psi_{p-1} \leq \cdots \leq \psi_1 \leq 2(m-p-1).$

We give conditions on *g* and *a* for obtaining generalized (proper) Bayes estimators of the form (6) such that the resulting estimators satisfy the conditions of Corollary 1, and hence are minimax. Note that the conditional distribution of  $\Theta$  given **X**,  $\Lambda$  is  $N_{p \times m}((\mathbf{I}_p - \Lambda)\mathbf{X}, (\mathbf{I}_p - \Lambda) \otimes \mathbf{I}_m)$ . Therefore, the generalized Bayes estimator of  $\Theta$  with respect to (2) under (3) is (see [15])

$$\delta_{\pi}(\mathbf{X}) = E[\mathbf{\Theta}|\mathbf{X}] = E[E[\mathbf{\Theta}|\mathbf{X}, \mathbf{\Lambda}]|\mathbf{X}] = (\mathbf{I}_p - E[\mathbf{\Lambda}|\mathbf{X}])\mathbf{X}.$$
(7)

 $E[\Lambda|\mathbf{X}]$  denotes expectation with respect to the posterior distribution of  $\Lambda$ , that is

$$p(\mathbf{\Lambda}|\mathbf{X}) \propto g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{X}')}{2}} I(\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_p), \tag{8}$$

so the resulting estimator  $\delta_{\pi}(\mathbf{X})$  can be written as  $\delta_{\pi}(\mathbf{X}) = (\mathbf{I}_p - E[\mathbf{\Lambda}|\mathbf{X}])\mathbf{X}$ , where

$$E[\mathbf{\Lambda}|\mathbf{X}] = \frac{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} \mathbf{\Lambda} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{X}')}{2}} d\mathbf{\Lambda}}{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{X}')}{2}} d\mathbf{\Lambda}}.$$
(9)

Now, using X = ULV' and letting  $\Lambda \rightarrow U\Lambda U'$ ,

$$E[\mathbf{\Lambda}|\mathbf{X}] = \mathbf{U} \frac{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} \mathbf{\Lambda} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{F})}{2}d\mathbf{\Lambda}}}{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{F})}{2}d\mathbf{\Lambda}}} \mathbf{U}'.$$
(10)

So, we have  $\delta_{\pi} = (\mathbf{I}_p - \mathbf{U} \mathbf{\Phi}(\mathbf{F}) \mathbf{U}') \mathbf{X}$ , where

$$\mathbf{\Phi}(\mathbf{F}) = \frac{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} \mathbf{\Lambda} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{F})}{2}d\mathbf{\Lambda}}}{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{F})}{2}d\mathbf{\Lambda}}}.$$
(11)

The estimation problem discussed in this paper is invariant with respect to  $X \rightarrow PXQ$ and  $\Theta \rightarrow P\Theta Q$  for any  $P \in O_p$  and  $Q \in O_m$ . Also, (2) is orthogonally invariant, namely

. . . . . .

$$\pi(\mathbf{\Theta}) = \pi(\mathbf{P}\mathbf{\Theta}\mathbf{Q}). \tag{12}$$

For every  $\mathbf{P} \in \mathbf{O}_p$  and  $\mathbf{Q} \in \mathbf{O}_m$ , according to Lemma 1 in [15],  $\mathbf{\Phi}(\mathbf{F})$  is a diagonal matrix, say  $\mathbf{\Phi}(\mathbf{F}) = \text{diag}(\phi_1(\mathbf{F}), \dots, \phi_p(\mathbf{F}))$ . Also,  $\delta_{\pi} = (\mathbf{I}_p - \mathbf{U}\mathbf{\Phi}(\mathbf{F})\mathbf{U}')\mathbf{X}$ , where  $\mathbf{\Phi} = \mathbf{F}^{-1}\mathbf{\Psi}(\mathbf{F})$  and the resulting  $\delta_{\pi}$  is of the form (4) with  $\mathbf{\Psi}(\mathbf{F}) = \text{diag}(\psi_1, \psi_2, \dots, \psi_p)$  and

$$\psi_{i}(\mathbf{F}) = f_{i} \frac{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} \lambda_{i,i} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{F})}{2}} d\mathbf{\Lambda}}{\int_{\mathbf{0}_{p \times p} < \mathbf{\Lambda} < \mathbf{I}_{p}} g\left(\frac{tr(\mathbf{\Lambda})}{p}\right) |\mathbf{\Lambda}|^{\frac{a+m}{2}-1} e^{-\frac{tr(\mathbf{\Lambda}\mathbf{F})}{2}} d\mathbf{\Lambda}}.$$
(13)

Now, let  $\lambda_k = \lambda_{k,k}$  for k = 1, ..., p and  $\lambda_{k,l} = \gamma_{k,l} \sqrt{\lambda_{k,k} \lambda_{l,l}}$  for k < l. The Jacobian of this transformation is

$$J((\lambda_{1,1},\ldots,\lambda_{p,p},\lambda_{1,2},\ldots,\lambda_{p-1,p}) \to (\lambda_{1},\ldots,\lambda_{p},\gamma_{1,2},\ldots,\gamma_{p-1,p})) = \prod_{k=1}^{p} \lambda_{k}^{\frac{p-1}{2}}.$$
 (14)

It holds that  $|\mathbf{\Lambda}| = |\mathbf{\Gamma}| \prod_{k=1}^{p} \lambda_k$ , where  $\mathbf{\Gamma} = (\gamma_{k,l})$  is a  $p \times p$  positive definite matrix with  $\gamma_{k,k} = 1$ . Denoting  $d\mathbf{\Gamma} = \bigwedge_{k < l} d\gamma_{k,l}$  and  $d\mathbf{\lambda} = \bigwedge_{k=1}^{p} d\lambda_k$ , we can write  $\psi_i$  as

$$\psi_{i}(\mathbf{F}) = f_{i} \frac{\int_{\mathbf{0}_{p \times p} < \mathbf{\Gamma} < \mathbf{I}_{p}} \int_{0}^{1} \cdots \int_{0}^{1} \lambda_{i} g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right) |\mathbf{\Gamma}|^{\frac{a+m}{2}-1} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum_{k=1}^{p} \lambda_{k} f_{k}}{2}} d\lambda d\mathbf{\Gamma}$$

$$\int_{\mathbf{0}_{p \times p} < \mathbf{\Gamma} < \mathbf{I}_{p}} \int_{0}^{1} \cdots \int_{0}^{1} g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right) |\mathbf{\Gamma}|^{\frac{a+m}{2}-1} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum_{k=1}^{p} \lambda_{k} f_{k}}{2}} d\lambda d\mathbf{\Gamma}$$
(15)

Note that  $\int_{\mathbf{0}_{p \times p} < \mathbf{\Gamma} < \mathbf{I}_p} |\mathbf{\Gamma}|^{\left(\frac{a+m}{2}\right)-1} d\mathbf{\Gamma}$  is finite for a > -m (see, for example, Theorem 1.4.5 on page 22 of [22]). Then, we can write

$$\psi_{i}(\mathbf{F}) = f_{i} \frac{\int_{0}^{1} \cdots \int_{0}^{1} \lambda_{i} g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right) \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{p} \lambda_{k} f_{k}}{2}} d\lambda}{\int_{0}^{1} \cdots \int_{0}^{1} g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right) \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{p} \lambda_{k} f_{k}}{2}} d\lambda}.$$
(16)

Lemma 2 shows that  $0 \le \psi_p \le \psi_{p-1} \le \cdots \le \psi_1$ .

**Lemma 2.** If g(t) is a decreasing function in t, then  $0 \le \psi_p \le \psi_{p-1} \le \cdots \le \psi_1$ .

**Proof.** We show that  $\psi_i - \psi_j \ge 0$  for j > i. The proof is similar to the proof of part (iv) of Lemma 3.1 in [20]. By using the transformation  $y_k = \frac{\lambda_k}{f_k}$ , k = 1, ..., p with the Jacobian  $I(\lambda_k, \dots, \lambda_k) \mapsto (u_k, \dots, u_k) = \prod_{k=1}^{p} f^{-1}$  (16) can be written as

$$J((\lambda_1, \dots, \lambda_p) \mapsto (y_1, \dots, y_p)) = \prod_{k=1}^{n} f_k^{-1}$$
, (16) can be written as

$$\psi_{i} = \frac{\int_{0}^{f_{1}} \cdots \int_{0}^{f_{p}} y_{i}g\left(p^{-1}\sum_{k=1}^{p} \frac{y_{k}}{f_{k}}\right) \left(\prod_{k=1}^{p} y_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{p} y_{k}}{2}} dy_{p} \cdots dy_{1}}{\int_{0}^{f_{1}} \cdots \int_{0}^{f_{p}} g\left(p^{-1}\sum_{k=1}^{p} \frac{y_{k}}{f_{k}}\right) \left(\prod_{k=1}^{p} y_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{p} y_{k}}{2}} dy_{p} \cdots dy_{1}}$$
(17)

for i = 1, ..., p. For j > i, we can write

$$\psi_{i} - \psi_{j} = \frac{\int_{0}^{f_{1}} \cdots \int_{0}^{f_{p}} (y_{i} - y_{j}) g(p^{-1} tr(YF^{-1})) |Y|^{\frac{a+p+m-3}{2}} e^{-\frac{tr(Y)}{2}} dY}{\int_{0}^{f_{1}} \cdots \int_{0}^{f_{p}} g(p^{-1} tr(YF^{-1})) |Y|^{\frac{a+p+m-3}{2}} e^{-\frac{tr(Y)}{2}} dY},$$
(18)

where  $Y = \text{diag}(y_1, y_2, \dots, y_p)$ .  $\Box$ 

In order to prove  $\psi_i - \psi_j \ge 0$  for every j > i, without any loss of generality, it is enough to show that the following function is non-negative

$$L(F) = \int_0^{f_1} \cdots \int_0^{f_p} (y_1 - y_2) g\left(p^{-1} tr\left(YF^{-1}\right)\right) |Y|^{\frac{a+p+m-3}{2}} e^{-\frac{tr(Y)}{2}} dY.$$
 (19)

Now, let  $\mathbf{O}_{1,2}$  denote the  $p \times p$  permutation matrix which interchanges the first and second rows by letting  $\mathbf{Y} \to \mathbf{O}_{1,2}\mathbf{Y}$ . Let  $W = \mathbf{O}_{1,2}\mathbf{Y}\mathbf{O}_{1,2}$ . The Jacobian of the transformation is  $J(Y \mapsto W) = 1$  because  $O_{1,2} = O'_{1,2} = O^{-1}_{1,2}$  and  $W = \text{diag}(w_1, w_2, \dots, w_p)$ . We can rewrite (19) as

$$L(F) = \int_0^{f_1} \cdots \int_0^{f_p} (w_2 - w_1) g\left( p^{-1} \left( \frac{w_1}{f_2} + \frac{w_2}{f_1} + \sum_{k \neq 1, 2}^p \frac{w_k}{f_k} \right) \right) |W|^{\frac{a+p+m-3}{2}} e^{-\frac{tr(W)}{2}} dW.$$
 (20)

Note that we can replace  $w_k$ 's with  $y_k$ 's in (20); its value does not change, meaning

$$L(F) = \int_0^{f_1} \cdots \int_0^{f_p} (y_2 - y_1) g\left( p^{-1} \left( \frac{y_1}{f_2} + \frac{y_2}{f_1} + \sum_{k \neq 1, 2}^p \frac{y_k}{f_k} \right) \right) |W|^{\frac{a+p+m-3}{2}} e^{-\frac{tr(Y)}{2}} dY.$$
(21)

Combining (19) and (21) yields

$$2L(F) = \int_0^{f_1} \cdots \int_0^{f_p} (y_1 - y_2) \left( g\left( p^{-1} \sum_{k=1}^p \frac{y_k}{f_k} \right) - g\left( p^{-1} \left( \frac{y_1}{f_2} + \frac{y_2}{f_1} + \sum_{k \neq 1, 2}^p \frac{y_k}{f_k} \right) \right) \right) |Y|^{\frac{a+p+m-3}{2}} e^{-\frac{tr(Y)}{2}} dY.$$

$$(22)$$

Note that we have two cases. One case is  $y_1 \ge y_2$  and the other case is  $y_1 < y_2$ . If  $y_1 \ge y_2$ , since  $f_1 > f_2$ , then

$$y_1\left(\frac{1}{f_1} - \frac{1}{f_2}\right) \le y_2\left(\frac{1}{f_1} - \frac{1}{f_2}\right)$$
 (23)

which implies

$$p^{-1}\left(\frac{y_1}{f_1} + \frac{y_2}{f_2} + \sum_{k \neq 1,2}^p \frac{y_k}{f_k}\right) = p^{-1}\left(\sum_{k=1}^p \frac{y_k}{f_k}\right) \le p^{-1}\left(\frac{y_1}{f_2} + \frac{y_2}{f_1} + \sum_{k \neq 1,2}^p \frac{y_k}{f_k}\right).$$
(24)

Since  $g(\cdot)$  is a decreasing function,

$$g\left(p^{-1}\sum_{k=1}^{p}\frac{\lambda_{k}}{f_{k}}\right) \geq g\left(p^{-1}\left(\frac{\lambda_{1}}{f_{2}} + \frac{\lambda_{2}}{f_{1}} + \sum_{k\neq 1,2}^{p}\frac{\lambda_{k}}{f_{k}}\right)\right),\tag{25}$$

so we have

$$(y_1 - y_2)\left(g\left(p^{-1}\sum_{k=1}^p \frac{\lambda_k}{f_k}\right) - g\left(p^{-1}\left(\frac{\lambda_1}{f_2} + \frac{\lambda_2}{f_1} + \sum_{k\neq 1,2}^p \frac{\lambda_k}{f_k}\right)\right)\right) \ge 0.$$

$$(26)$$

Hence,  $L(F) \ge 0$  for the case  $y_1 \ge y_2$ . For the case  $y_1 \le y_2$ , we can similarly show  $L(F) \ge 0$ . It can be proven similarly that  $\psi_i - \psi_j \ge 0$  for every j > i and hence  $\psi_p \le \psi_{p-1} \le \cdots \le \psi_1$ . Clearly,  $\psi_p \ge 0$ , and hence  $0 \le \psi_p \le \psi_{p-1} \le \cdots \le \psi_1$ .

We need Lemma 3 to continue.

**Lemma 3.** Let  $\zeta$  denote a probability density function with respect to a  $\sigma$ -finite measure v on  $\mathbf{R}^p$ . For any two points,  $\lambda = (\lambda_1, \ldots, \lambda_p)'$  and  $\mu = (\mu_1, \ldots, \mu_p)'$ , define  $\lambda \wedge \mu = (\min(\lambda_1, \mu_1), \ldots, \min(\lambda_p, \mu_p))'$  and  $\lambda \vee \mu = (\max(\lambda_1, \mu_1), \ldots, \max(\lambda_p, \mu_p))'$ . Suppose  $\zeta$  satisfies

$$\zeta(\lambda)\zeta(\mu) \le \zeta(\lambda \land \mu)\zeta(\lambda \lor \mu).$$
<sup>(27)</sup>

If functions f and g are non-decreasing in each argument and if f, g and fg are integrable with respect to  $\zeta$ , then

$$\int f(\lambda)g(\lambda)\zeta(\lambda)dv(\lambda) \ge \int f(\lambda)\zeta(\lambda)dv(\lambda) \int g(\lambda)\zeta(\lambda)dv(\lambda).$$
(28)

**Proof.** See [23]. □

Lemma 4 gives conditions for  $\psi_i(f_1, f_2, \dots, f_{i-1}, f_i, f_{i+1}, \dots, f_p)$  to be non-decreasing in  $f_i, i = 1, \dots, p$  for fixed  $f_1, f_2, \dots, f_{i-1}, f_{i+1}, \dots, f_p$ .

Lemma 4. Suppose g satisfies

I. 
$$\lim_{\lambda_i \to 0} \lambda_i^{\frac{a+m+p-1}{2}} g\left(p^{-1} \sum_{k=1}^p \lambda_k\right) = 0 \text{ for } i = 1, \dots, p;$$
$$\lambda_i g'\left(p^{-1} \sum_{k=1}^p \lambda_k\right)$$

II. For 
$$i = 1, ..., p$$
,  $\frac{\frac{1}{k} \left( p - \frac{p}{k-1} \times k \right)}{g\left( p^{-1} \sum_{k=1}^{p} \lambda_k \right)}$  is non-increasing in  $\lambda_j, j = 1, ..., p$ ;

III. For  $\lambda = (\lambda_1, \lambda_2, ..., \lambda_p)$  and  $\lambda' = (\lambda'_1, \lambda'_2, ..., \lambda'_p)$ , where  $0 < \lambda_i, \lambda'_i < 1, i = 1, ..., p$ ,  $g(\cdot)$  satisfies

$$g\left(p^{-1}\sum_{k=1}^{p}\lambda_{k}\right)g\left(p^{-1}\sum_{k=1}^{p}\lambda_{k}'\right) \leq g\left(p^{-1}\sum_{k=1}^{p}\left(\lambda_{k}\vee\lambda_{k}'\right)\right)g\left(p^{-1}\sum_{k=1}^{p}\left(\lambda_{k}\wedge\lambda_{k}'\right)\right).$$
 (29)

*Then, for any i,*  $\psi_i$  *is non-decreasing with respect to*  $f_i$ *.* 

**Proof.** We can write (16) as

$$\psi_i = f_i \frac{\Phi_1(\mathbf{F})}{\Phi_0(\mathbf{F})},\tag{30}$$

where

$$\Phi_k(\mathbf{F}) = \int_0^1 \cdots \int_0^1 \lambda_i^k g\left(p^{-1} \sum_{k=1}^p \lambda_k\right) \left(\prod_{k=1}^p \lambda_k^{\frac{a+p+m-3}{2}}\right) e^{-\frac{p}{\sum_{k=1}^p \lambda_k f_k}} d\lambda$$
(31)

for k = 0, 1. We have

$$\frac{\partial \psi_i}{\partial f_i} = \frac{\Phi_1(\mathbf{F})}{\Phi_0(\mathbf{F})} + \frac{f_i}{\Phi_0^2(\mathbf{F})} \left( \frac{\partial \Phi_1(\mathbf{F})}{\partial f_i} \Phi_0(\mathbf{F}) - \frac{\partial \Phi_0(\mathbf{F})}{\partial f_i} \Phi_1(\mathbf{F}) \right)$$
(32)

and

$$\frac{\partial \Phi_k(\mathbf{F})}{\partial f_i} = \frac{1}{f_i} \left\{ C(\mathbf{F}) - \left( k + \frac{a+m+p-1}{2} \right) \Phi_k - B_k(\mathbf{F}) \right\},\tag{33}$$

where

$$C(\mathbf{F}) = \int_0^1 \cdots \int_0^1 g\left(p^{-1}\left(1 + \sum_{k \neq i} \lambda_k\right)\right) \left(\prod_{k=1}^p \lambda_k^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum_{k \neq i} \lambda_k f_k}{2}} d\lambda_{-i}$$
(34)

and

$$B_{k}(\mathbf{F}) = \int_{0}^{1} \cdots \int_{0}^{1} \lambda_{i}^{k+1} \frac{\partial g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right)}{\partial \lambda_{i}} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{p} \lambda_{k} f_{k}}{2}} d\lambda.$$
(35)

By replacing (33) with (32), we have

$$\frac{\partial \psi_i}{\partial f_i} = \frac{1}{\Phi_0^2(\mathbf{F})} \{ (\Phi_0 - \Phi_1) C(\mathbf{F}) + (B_0 \Phi_1 - B_1 \Phi_0) \}.$$
(36)

Since  $C(\mathbf{F}) \geq 0$  and also  $\Phi_0 - \Phi_1 \geq 0$ , we have  $(\Phi_0 - \Phi_1)C(\mathbf{F}) \geq 0$ . In order to prove  $\frac{\partial \psi_i}{\partial f_i} \geq 0$ , it suffices to show that  $\frac{B_0(\mathbf{F})}{\Phi_0} \frac{\Phi_1}{\Phi_0} \geq \frac{B_1(\mathbf{F})}{\Phi_0}$ . That is,

$$\int_{0}^{1} \cdots \int_{0}^{1} -\frac{\lambda_{i}^{2} g'\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right)}{g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right)} \xi(\lambda) d\lambda$$

$$\geq \int_{0}^{1} \cdots \int_{0}^{1} -\frac{\lambda_{i} g'\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right)}{g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right)} \xi(\lambda) d\lambda \int_{0}^{1} \cdots \int_{0}^{1} \lambda_{i} \xi(\lambda) d\lambda, \qquad (37)$$

where

$$\xi(\boldsymbol{\lambda}) = \frac{g\left(p^{-1}\sum_{k=1}^{p}\lambda_{k}\right)\left(\prod_{k=1}^{p}\lambda_{k}^{\frac{a+p+m-3}{2}}\right)e^{-\frac{\sum\limits_{k=1}^{p}\lambda_{k}f_{k}}{2}}}{\int_{0}^{1}\cdots\int_{0}^{1}g\left(p^{-1}\sum_{k=1}^{p}\lambda_{k}\right)\left(\prod\limits_{k=1}^{p}\lambda_{k}^{\frac{a+p+m-3}{2}}\right)e^{-\frac{\sum\limits_{k=1}^{p}\lambda_{k}f_{k}}{2}}d\boldsymbol{\lambda}}.$$
(38)

Using condition (III), it is easy to show that

$$\xi(\lambda)\xi(\lambda') \leq \xi(\lambda \lor \lambda')\xi(\lambda \land \lambda') \tag{39}$$

for  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$  and  $\lambda' = (\lambda'_1, \lambda'_2, \dots, \lambda'_p)$ . Because of condition (II), Lemma 3 can be applied to prove (37), so  $\frac{\partial \psi_i}{\partial f_i} \ge 0$  for all  $i = 1, \dots, p$ .  $\Box$ 

Lemma 5 gives conditions for determining an upper bound on  $\psi_1$ .

g

**Lemma 5.** Assume that  $g(1^-) = \lim_{t \uparrow 1} g(t) < \infty$  and

$$\lim_{t \to 0} \frac{g(t)}{e^{\alpha t}} = c.$$
(40)

*For some*  $\alpha \leq 0$  *and some* c > 0*,* 

$$\lim_{f_p \to \infty} \lim_{f_{p-1} \to \infty} \cdots \lim_{f_1 \to \infty} \psi_1(\mathbf{F}) = a + m + p - 1.$$
(41)

**Proof.** Note that  $\frac{g(t)}{e^{at}}$  is continuous in (0, 1) and has limits at the points 0 and 1. So, this function is bounded on its domain, meaning there exists a k > 0 such that

$$(t) \le k e^{\alpha t}.\tag{42}$$

p

From (16),

$$\psi_{1}(\mathbf{F}) = f_{1} \frac{\int_{0}^{1} \cdots \int_{0}^{1} \lambda_{1} g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right) \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{k} \lambda_{k} f_{k}}{2}} d\lambda}{\int_{0}^{1} \cdots \int_{0}^{1} g\left(p^{-1} \sum_{k=1}^{p} \lambda_{k}\right) \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{k} \lambda_{k} f_{k}}{2}} d\lambda}.$$
(43)

Letting  $\lambda_k \rightarrow \frac{\lambda_k}{f_k}$ , we obtain

$$\psi_1(\mathbf{F}) = \frac{\int_0^\infty \cdots \int_0^\infty M_1(\mathbf{F}, \boldsymbol{\lambda}) d\boldsymbol{\lambda}}{\int_0^\infty \cdots \int_0^\infty M_0(\mathbf{F}, \boldsymbol{\lambda}) d\boldsymbol{\lambda}},$$
(44)

where for i = 0, 1,

$$M_i(\mathbf{F}, \boldsymbol{\lambda}) = \lambda_1^i g\left(p^{-1} \sum_{k=1}^p \frac{\lambda_k}{f_k}\right) \left(\prod_{k=1}^p \left(\lambda_k^{\frac{a+p+m-3}{2}}\right) I(0 < \lambda_k < f_k)\right) e^{-\frac{\sum_{k=1}^p \lambda_k}{2}}.$$
(45)

We now bound the integrand of  $M_i$  in order to apply the Lebesgue-dominated convergence theorem. First, using (42), we have

$$\int_{0}^{\infty} \int_{0}^{\infty} \cdots \int_{0}^{\infty} M_{i}(\mathbf{F}, \boldsymbol{\lambda}) d\boldsymbol{\lambda}$$

$$= \int_{0}^{f_{1}} \int_{0}^{f_{2}} \cdots \int_{0}^{f_{p}} \lambda_{1}^{i} g\left(\frac{1}{p}\left(\sum_{k=1}^{p} \frac{\lambda_{k}}{f_{k}}\right)\right) \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{p}{\sum_{k=1}^{\lambda_{k}} \lambda_{k}}} d\boldsymbol{\lambda}$$

$$\leq k \int_{0}^{f_{1}} \int_{0}^{f_{2}} \cdots \int_{0}^{f_{p}} \lambda_{1}^{i} e^{\frac{p}{p}\left(\sum_{k=1}^{p} \frac{\lambda_{k}}{f_{k}}\right)} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{p}{\sum_{k=1}^{\lambda_{k}} \lambda_{k}}} d\boldsymbol{\lambda}.$$
(46)

Since  $\alpha \leq 0$ , we have

$$\int_{0}^{f_{1}} \int_{0}^{f_{2}} \cdots \int_{0}^{f_{p}} \lambda_{1}^{i} e^{\frac{\alpha}{p} \left(\sum_{k=1}^{p} \frac{\lambda_{k}}{f_{k}}\right)} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum_{k=1}^{p} \lambda_{k}}{2}} d\lambda$$

$$\leq \int_{0}^{f_{1}} \int_{0}^{f_{2}} \cdots \int_{0}^{f_{p}} \lambda_{1}^{i} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum_{k=1}^{p} \lambda_{k}}{2}} d\lambda.$$

$$(47)$$

v

Since a > -m, we have

$$\int_{0}^{f_{1}} \int_{0}^{f_{2}} \cdots \int_{0}^{f_{p}} \lambda_{1}^{i} \left( \prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}} \right) e^{-\frac{\sum \lambda_{k}}{2}} d\lambda$$
  
=  $2^{\frac{p(a+m+p-1)}{2}+i} \Gamma^{p-1} \left( \frac{a+p+m-1}{2} \right) \Gamma \left( \frac{a+p+m+2i-1}{2} \right).$  (48)

Thus, (46) is finite, and the Lebesgue-dominated convergence theorem can be used. Hence,

$$\lim_{f_{p}\to\infty}\lim_{f_{p-1}\to\infty}\cdots\lim_{f_{1}\to\infty}\int_{0}^{\infty}\int_{0}^{\infty}\cdots\int_{0}^{\infty}M_{i}(\mathbf{F},\boldsymbol{\lambda})d\boldsymbol{\lambda}$$

$$=\int_{0}^{\infty}\int_{0}^{\infty}\cdots\int_{0}^{\infty}\lim_{f_{p}\to\infty}\lim_{f_{p-1}\to\infty}\cdots\lim_{f_{1}\to\infty}M_{i}(\mathbf{F},\boldsymbol{\lambda})d\boldsymbol{\lambda}$$

$$=c\int_{0}^{\infty}\int_{0}^{\infty}\cdots\int_{0}^{\infty}\lambda_{1}^{i}\left(\prod_{k=1}^{p}\lambda_{k}^{\frac{a+p+m-3}{2}}\right)e^{-\frac{\sum_{k=1}^{p}\lambda_{k}}{2}}d\boldsymbol{\lambda}$$

$$=c2^{\frac{p(a+m+p-1)}{2}+i}\Gamma^{p-1}\left(\frac{a+p+m-1}{2}\right)\Gamma\left(\frac{a+p+m+2i-1}{2}\right).$$
(49)

Finally, using (44) and the above limits, we have

$$\lim_{f_p \to \infty} \lim_{f_{p-1} \to \infty} \cdots \lim_{f_1 \to \infty} \psi_1(\mathbf{F}) = a + m + p - 1,$$
(50)

the desired result.

The results of Lemmas 2, 4 and 5 are our main result.

**Theorem 1.** (a) If the conditions of Lemmas 2, 4 and 5 hold and if a < m - 3p - 1, then the generalized Bayes estimator  $\delta_{\pi}(\mathbf{X})$  with respect to (2) is minimax under the loss function (3). (b) Further, if g is integrable, then the estimator  $\delta_{\pi}$  is proper Bayes and minimax, and hence admissible under (3).

#### 3. Examples

We give two examples in this section to which our results can be applied. We also make connections to [15].

**Example 1.** Assume that g(t) = 1 for 0 < t < 1. For this function g(t) = 1, the class of prior distributions  $\pi(\Theta)$  is the form of

$$\pi(\Theta) = (2\pi)^{-\frac{mp}{2}} \int_{0_{p \times p} < \Lambda < I_p} |\Lambda|^{\frac{a+m}{2}-1} |I_p - \Lambda|^{-\frac{m}{2}} e^{-\frac{tr((I_p - \Lambda)^{-1}\Lambda \Theta \Theta')}{2}} d\Lambda.$$
(51)

This is the same class of prior distributions studied by [15]. If we do the same as in Section 2 of [15], we will obtain the class of Bayes estimators of the form  $\delta_{\pi} = (I_p - UF^{-1}\Psi(F)U')X$  with  $\Psi(F) = \text{diag}(\psi_1(F), \dots, \psi_p(F))$ , where

$$\psi_{i}(F) = f_{i} \frac{\int_{0}^{1} \cdots \int_{0}^{1} \lambda_{i} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{p} \lambda_{k} f_{k}}{2}} d\lambda}{\int_{0}^{1} \cdots \int_{0}^{1} \left(\prod\limits_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum\limits_{k=1}^{p} \lambda_{k} f_{k}}{2}} d\lambda}$$
(52)

for i = 1, ..., p. We now show the class of Bayes estimators is minimax under the loss function (3). It is sufficient to show that the conditions of Theorem 1 are satisfied, meaning that we should show that g(t) = 1 for 0 < t < 1 follows the conditions stated in Lemmas 2, 4 and 5. Note that Lemma 2 states that if function g(t) is decreasing in t, then the conditions of Lemma 2 hold. Because g(t) = 1 is a constant function, it is decreasing in t. So, the conditions of Lemma 2 hold.

For 
$$a > -\frac{m+p-1}{2}$$
,  
$$\lim_{\lambda_i \to 0} \lambda_i^{\frac{a+p+m-1}{2}} = 0$$
(53)

for i = 1, ..., p. Also  $\frac{\lambda_i g'\left(p^{-1} \sum\limits_{k=1}^p \lambda_k\right)}{g\left(p^{-1} \sum\limits_{k=1}^p \lambda_k\right)} = 0$  and so  $\frac{\lambda_i g'\left(p^{-1} \sum\limits_{k=1}^p \lambda_k\right)}{g\left(p^{-1} \sum\limits_{k=1}^p \lambda_k\right)}$  is non-decreasing in  $\lambda_j$ ,

j = 1, ..., p. For  $\lambda = (\lambda_1, \lambda_2, ..., \lambda_p)$  and  $\lambda' = (\lambda'_1, \lambda'_2, ..., \lambda'_p)$ , where  $0 < \lambda_i, \lambda'_i < 1$ , i = 1, ..., p, the inequality

$$g\left(p^{-1}\sum_{k=1}^{p}\lambda_{k}\right)g\left(p^{-1}\sum_{k=1}^{p}\lambda_{k}'\right) \leq g\left(p^{-1}\sum_{k=1}^{p}\left(\lambda_{k}\vee\lambda_{k}'\right)\right)g\left(p^{-1}\sum_{k=1}^{p}\left(\lambda_{k}\wedge\lambda_{k}'\right)\right)$$
(54)

holds because  $g\left(p^{-1}\sum_{k=1}^{p}\lambda_k\right) = 1$ . Therefore,  $g(\cdot)$  satisfies the conditions of Lemma 4. Further,  $g(1^{-1}) = \lim_{t\uparrow 1} g(t) = 1 < \infty$  and, if we select  $\alpha = 0$ , then

$$\lim_{\lambda_i \to 0} \frac{g(t)}{e^{\alpha t}} = 1.$$
(55)

Thus, the conditions of Lemma 5 hold. Now, if a < m - 3p - 1, then based on Theorem 1, the proper Bayes estimators  $\delta_{\pi}$  are minimax under loss function (3). Thus, our class of minimax estimators include [15]'s results.

**Example 2.** Another class of prior distributions  $\pi(\Theta)$  can be constructed by taking  $g(\cdot)$  to be

$$g(t) = c' e^{-\beta t}, \ 0 < t < 1,$$
(56)

where c' > 0 and  $\beta > 0$ . Then,  $\pi(\Theta)$  will be

$$\pi(\Theta) = c'(2\pi)^{-\frac{mp}{2}} \int_{0_{p \times p} < \Lambda < I_p} e^{-\beta \frac{tr(\Lambda)}{p}} |\Lambda|^{\frac{a+m}{2}-1} |I_p - \Lambda|^{-\frac{m}{2}} e^{-\frac{tr((I_p - \Lambda)^{-1}\Lambda \Theta \Theta')}{2}} d\Lambda.$$
(57)

If we follow the discussion of Section 2 in order, the Bayes estimators will be of the form  $\delta_{\pi} = (I_p - UF^{-1}\Psi(F)U')X$  and  $\Psi(F) = \text{diag}(\psi_1(F), \dots, \psi_p(F))$ , with

$$\psi_{i}(F) = f_{i} \frac{\int_{0}^{1} \cdots \int_{0}^{1} \lambda_{i} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\sum_{k=1}^{p} \lambda_{k} \left(\frac{\beta}{p} + \frac{f_{k}}{2}\right)} d\lambda}{\int_{0}^{1} \cdots \int_{0}^{1} \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\sum_{k=1}^{p} \lambda_{k} \left(\frac{\beta}{p} + \frac{f_{k}}{2}\right)} d\lambda}$$
(58)

for i = 1, ..., p. We show that this class satisfies the conditions of Theorem 1. Based on Theorem 1, it is sufficient to show that the conditions of Lemmas 2, 4 and 5 are satisfied. Lemma 2 states that if g(t) is decreasing in t, then Lemma 2 holds. Since c' > 0 and  $\beta > 0$ ,  $g'(t) = -c'\beta e^{-\beta t} < 0$  for each 0 < t < 1. Therefore, the condition of Lemma 2 is satisfied.

For 
$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)$$
 and  $\boldsymbol{\lambda}' = (\lambda'_1, \lambda'_2, \dots, \lambda'_p)$ , where  $0 < \lambda_i, \lambda'_i < 1, i = 1, \dots, p$ ,

$$e^{-\frac{\beta\sum\limits_{k=1}^{\sum}\lambda_{k}}{p}}e^{-\frac{\beta\sum\limits_{k=1}^{\sum}\lambda_{k}'}{p}} \le e^{-\frac{\beta\sum\limits_{k=1}^{\sum}(\lambda_{k}\vee\lambda_{k}')}{p}}e^{-\frac{\beta\sum\limits_{k=1}^{\sum}(\lambda_{k}\wedge\lambda_{k}')}{p}}$$
(59)

and so condition III of Lemma 4 holds. Also, if a > -(m + p - 1), then

$$\lim_{\lambda_i \to 0} \lambda_i^{\frac{a+m+p-1}{2}} g\left(p^{-1} \sum_{k=1}^p \lambda_k\right) = \lim_{\lambda_i \to 0} \lambda_i^{\frac{a+m+p-1}{2}} e^{-\beta\left(p^{-1} \sum_{k=1}^p \lambda_k\right)} = 0$$
(60)

for i = 1, ..., p, so condition II of Lemma 4 holds.  $\frac{\lambda_i g' \left( p^{-1} \sum_{k=1}^p \lambda_k \right)}{g \left( p^{-1} \sum_{k=1}^p \lambda_k \right)} = -c' \beta \lambda_i$  is a decreasing

function in  $\lambda_j$  for each j = 1, ..., p, so condition I of Lemma 4 holds. We have

$$g(1^{-}) = \lim_{t\uparrow 1} g(t) = \lim_{t\uparrow 1} c'e^{-\beta t} = c'e^{-\beta} < \infty$$
(61)

and, if we choose  $\alpha = -\beta$ , we obtain

$$\lim_{t \to 0} \frac{g(t)}{e^{\alpha t}} = c' > 0.$$
(62)

Hence, the conditions of Lemma 5 hold. Also, if a < m - 3p - 1, then all the conditions of Theorem 1 hold and hence the Bayes estimator  $\delta_{\pi}$  obtained with respect to the priors  $\pi(\Theta)$  is minimax under the loss function (3).

### 4. Concluding Remarks

The problem of estimating the mean matrix  $\Theta$  of a matrix variate normal distribution with the covariance matrix  $I_p \otimes I_m$  under the loss function  $tr((\delta - \Theta)(\delta - \Theta)')$  has been investigated. This is an invariant problem with respect to the group of orthogonal transformations. We considered the following prior distribution which is invariant under the group of orthogonal transformations.

$$\pi(\Theta) = (2\pi)^{-\frac{mp}{2}} \int_{0_{p \times p} < \Lambda < I_p} g\left(\frac{tr(\Lambda)}{p}\right) |\Lambda|^{\frac{a+m}{2}-1} |I_p - \Lambda|^{-\frac{m}{2}} e^{-\frac{tr\left((I_p - \Lambda)^{-1}\Lambda\Theta\Theta'\right)}{2}} d\Lambda.$$
(63)

Using the invariant discussion, our Bayes estimators are of the form  $\delta_{\pi} = (I_p - UF^{-1}\Psi(F)U')X$ , where  $\Psi(F) = \text{diag}(\psi_1(F), \dots, \psi_p(F))$ , with

$$\psi_{i}(F) = f_{i} \frac{\int_{0}^{1} \cdots \int_{0}^{1} \lambda_{i} g\left(\frac{\sum \lambda_{k}}{p}\right) \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum \lambda_{k} f_{k}}{2}} d\lambda}{\int_{0}^{1} \cdots \int_{0}^{1} g\left(\frac{\sum \lambda_{k}}{p}\right) \left(\prod_{k=1}^{p} \lambda_{k}^{\frac{a+p+m-3}{2}}\right) e^{-\frac{\sum \lambda_{k} f_{k}}{2}} d\lambda}$$
(64)

for i = 1, ..., p.

In this paper, we have obtained the conditions for the continuous function  $g(\cdot)$  such that the resulting Bayes estimators are minimax under the given loss function. If we want to make a comparison between the results of [15] and our results in this paper, it should be said that the difference is only in  $g(\cdot)$ . [15] showed that if g(t) = 1, 0 < t < 1, then the resulting Bayes estimators are minimax under the given loss function. We obtained conditions on g(t) such that the resulting class of Bayes estimators is minimax. We also showed that the function g obtained by [15] applies to the conditions obtained in this paper, so the results of our paper include [15]'s results. Also, we presented another example in such a way that if  $g(t) = c'e^{-\beta t}$ , 0 < t < 1, where c',  $\beta > 0$ , then the conditions obtained in the paper are valid and the resulting Bayes estimators are minimax under (3). Hence, we have obtained a larger class of Bayes estimators which include the class of Bayes estimators obtained by [15].

The estimator proposed in this paper being minimax and admissible could lead to improved inference in various application areas of the matrix variate normal distribution, including analyses of multiple vector autoregressions; brain connectivity alternation detection; capacity for severely fading MIMO channels; integrated principal components analyses; determination of relationships between incidence and mortality of asthma and PM2.5, ozone, and household air pollution; autism spectrum disorder identification; and identification of depression disorder using multi-view high-order brain function networks. We provide two examples:

Example 1 —Suppose that there are three mines in one area and the owner of all three mines is the same. Suppose the owner wants to know how much gold, copper, zinc, aluminum, bronze, and iron can be extracted per kilogram of ore in each mine. They want the authorities to randomly extract one kilogram of ore from each mine in one day *n* certain times and determine the amount of the metals in a laboratory:

	gold	copper	zinc	aluminum	bronze	iron
mine 1	$X_{1,1}$	X <sub>1,2</sub>	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$
mine 2	$X_{2,1}$	X <sub>2,2</sub>	X <sub>2,3</sub>	X <sub>2,4</sub>	X <sub>2,5</sub>	X <sub>2,6</sub>
mine 3	$X_{3,1}$	X <sub>3,2</sub>	X <sub>3,3</sub>	X <sub>3,4</sub>	$X_{3,5}$	X3,6

They are faced with the following matrix of variables

$$\mathbf{X} = \begin{bmatrix} X_{1,1} & X_{2,1} & X_{3,1} \\ X_{1,2} & X_{2,2} & X_{3,2} \\ X_{1,3} & X_{2,3} & X_{3,3} \\ X_{1,4} & X_{2,4} & X_{3,4} \\ X_{1,5} & X_{2,5} & X_{3,5} \\ X_{1,6} & X_{2,6} & X_{3,6} \end{bmatrix}.$$

Based on previous experience, they know that the amount of each metal extracted from each mine is independent of other mines and the amount of other metals. They also know that the amount of metal extracted from each kilogram of ore has a small dispersion and that the amount of each metal from each mine has a normal distribution. Our results in this paper can be used to estimate the means of the metals extracted.

Example 2—Suppose that a researcher wants to investigate the effect of the number of study hours (3 or 4 h) on the progress of three students in four subjects: mathematics, history, art, and geography. They choose four classmates at random and ask them to spend 3 h per week studying each subject for half of a semester and 4 h per week for the other half of the same semester. They want to observe the results per student as a random matrix as follows:

Information of student <i>k</i>	Mathematics score	History score	Art score	Geography score
3 h	$X_{1,1,k}$	X <sub>1,2,k</sub>	$X_{1,3,k}$	X <sub>1,4,k</sub>
4 h	$X_{2,1,k}$	$X_{2,2,k}$	$X_{2,3,k}$	$X_{2,4,k}$

Suppose the numerical results are

$\mathbf{x}_1 = \begin{bmatrix} \\ \end{bmatrix}$	16.5	13.75	18.75	17.75 ]	× [	17.2	15.75	19.25	18.25	
	17.5	12.25	18.5	19.5	, $x_2 = $	17	14.85	19.5	18.5	'
$\mathbf{x}_3 = \begin{bmatrix} \\ \end{bmatrix}$	5.25	15.75	17.25	18.25	], $x_4 =$	[ 15.25	5 14.75	17.25	5 18.75	5 ]
	16.5	15.85	17.5	18.75		15.5	13.85	16.5	19.25	5

The researcher previously performed similar tests in other schools; it was found that the number of study hours (3 or 4) has no effect, the rates of progress in each course are independent from each other, and each variable of this random matrix has a normal distribution. If they want to estimate the mean of the matrix variate normal distribution, our results can be used.

A future study could be deriving explicit expressions for moments of  $\Theta|X$ . These may be obtained using the results of [24].

**Author Contributions:** Conceptualization, S.Z. and S.N.; methodology, S.Z. and S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** The authors would like to thank the editor and the three referees for careful reading and comments which greatly improved the paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

#### References

- 1. Wichitaksorn, N. Analyzing multiple vector autoregressions through matrix-variate normal distribution with two covariance matrices. *Commun. Stat.*—*Theory Methods* **2019**, *49*, 1801–1817. [CrossRef]
- 2. Xia, Y.; Li, L.X. Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Stat. Sin.* **2019**, *29*, 303–328. [CrossRef]

- 3. Ferreira, J.T. Upper bounds for the capacity for severely fading MIMO channels under a scale mixture assumption. *Entropy* **2021**, 23, 845. [CrossRef] [PubMed]
- 4. Tang, T.M.; Allen, G.I. Integrated principal components analysis. J. Mach. Learn. Res. 2021, 22, 1–71.
- Ahmadi, F.; Fallah, Z.; Shadmani, F.K.; Allahmoradi, M.; Salahshoor, P.; Ahmadi, S.; Mansori, K. Relationship between incidence and mortality of asthma with PM2.5, ozone, and household air pollution from 1990 to 2106 in the world: An ecological study. *Egypt. J. Chest Dis. Tuberc.* 2022, 71, 457–463.
- 6. Jiang, X.; Zhou, Y.Y.; Zhang, Y.N.; Zhang, L.M.; Qiao, L.S.; De Leone, R. Estimating high-order brain functional networks in Bayesian view for autism spectrum disorder identification. *Front. Neurosci.* **2022**, *16*, 872848. [CrossRef]
- Zhao, F.; Gao, T.Y.; Cao, Z.; Chen, X.B.; Mao, Y.Y.; Mao, N.; Ren, Y.D. Identifying depression disorder using multi-view high-order brain function network derived from electroencephalography signal. *Front. Comput. Neurosci.* 2022, 16, 1046310. [CrossRef] [PubMed]
- 8. Efron, B.; Morris, C. Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika* **1972**, *59*, 335–347. [CrossRef]
- Stein, C. Estimation of the mean of a multivariate normal distribution. In Proceedings of the Prague Symposium on Asymptotic Statistics, Prague, Czech Republic, 3–6 September 1973; pp. 345–381.
- 10. Zhang, Z. On estimation of matrix of normal mean. J. Multivar. Anal. 1986, 18, 70-82. [CrossRef]
- 11. Baranchik, A.J. A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Stat.* **1970**, *41*, 642–645. [CrossRef]
- 12. Bilodeau, M.; Kariya, T. Minimax estimators in the normal MANOVA model. J. Multivar. Anal. 1989, 28, 260–270. [CrossRef]
- 13. Konno, Y. On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivar. Anal.* **1991**, *36*, 44–55. [CrossRef]
- 14. Haff, R.L. An identity for the Wishart distribution with applications. J. Multivar. Anal. 1979, 9, 531–544. [CrossRef]
- 15. Tsukuma, H. Admissibility and minimaxity of Bayes estimators for a normal mean matrix. *J. Multivar. Anal.* **2008**, *99*, 2251–2264. [CrossRef]
- 16. Tsukuma, H. Generalized Bayes minimax estimation of the normal mean matrix with unknown covariance matrix. *J. Multivar. Anal.* **2009**, *100*, 2296–2304. [CrossRef]
- 17. Zinodiny, S.; Strawderman, W.E.; Parsian, A. Bayes minimax estimation of the multivariate normal mean vector for the case of common unknown variance. *J. Multivar. Anal.* 2011, 102, 1256–1262. [CrossRef]
- 18. Zinodiny, S.; Rezaei, S.; Arjmand, O.N.; Nadarajah, S. A new class of Bayes minimax estimators of the normal mean matrix for the case of common unknown variances. *Statistics* **2017**, *51*, 1082–1094. [CrossRef]
- 19. Tsukuma, H. Shrinkage priors for Bayesian estimation of the mean matrix in an elliptically contoured distribution. *J. Multivar. Anal.* **2010**, *101*, 1483–1492. [CrossRef]
- 20. Tsukuma, H. Proper Bayes minimax estimators of the normal mean matrix with common unknown variances. *J. Stat. Plan. Inference* **2010**, 140, 2596–2606. [CrossRef]
- 21. Faith, E.R. Minimax Bayes estimators of a multivariate normal mean. J. Multivar. Anal. 1978, 8, 372–379. [CrossRef]
- 22. Gupta, A.K.; Nagar, D. Matrix Variate Distributions; Chapman and Hall/CRC: London, UK, 1999.
- 23. Fortuin, C.M.; Kasteleyn, P.W.; Ginibre, J. Correlation inequalities on some partially ordered sets. *Commun. Math. Phys.* **1971**, *22*, 89–103. [CrossRef]
- 24. Mathai, A.M.; Provost, S.B.; Haubold, H.J. *Multivariate Statistical Analysis in the Real and Complex Domains*; Springer: New York, NY, USA, 2022.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.