Mathematical Statistics
Stockholm University

Stein-Haff identity for the exponential family

Gustav Alfelt

Research Report 2018:17

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se

Mathematical Statistics
Stockholm University
Research Report **2018:17**,
http://www.math.su.se

Stein-Haff identity for the exponential family

Gustav Alfelt*

July 2018

Abstract

In this paper, the Stein-Haff identity is established for positive-definite and symmetric random matrices belonging to the exponential family. The identity is then applied to the matrix-variate gamma distribution, and an estimator that dominates the maximum likelihood estimator in terms of Stein's loss is obtained. Finally a simulation study is conducted in order to support the theoretical results.

Keywords and phrases: Random matrices, matrix-variate gamma distribution, decision theory.
2010 Mathematics Subject Classification: Primary 62H12; Secondary 62C99.

1 Introduction

The Stein-Haff identity was first derived by Stein (1977) and Haff (1979) regarding the problem of estimating the covariance matrix of multivariate normal populations. The $p \times p$ sample covariance matrix $\mathbf{W}$ of such a population follows a Wishart distribution, and is commonly estimated using the unbiased estimator $\mathbf{W}/n$, where n is the sample size. However, the eigenvalues of the estimator $\mathbf{W}/n$ tends to spread out more over the positive real line, than the equivalent eigenvalues of the population covariance matrix $\mathbf{\Sigma}$. For example, letting $\lambda_1, \dots, \lambda_p$ be the p ordered eigenvalues of $\mathbf{\Sigma}$ and $l_1, \dots, l_p$ be the p ordered sample eigenvalues of $\mathbf{W}/n$, l_1 is an positively biased estimator of λ_1 and l_p is a negatively biased estimator of λ_p (see e.g. Van der Vaart (1961)). As such, it can often be useful to consider estimators that aim to decrease larger sample eigenvalues and increase smaller sample eigenvalues.

*E-mail: gustava@math.su.se. Address: Department of Mathematics, Stockholm University, Roslagsvägen 101, SE-10691 Stockholm, Sweden.

Additionally, the problem of an estimator of the covariance matrix of a normal population have been well studied from a decision-theoretic viewpoint[1]. In this approach estimators are evaluated with a non-negative *loss function* $L(\hat{\theta}, \theta)$ and associated *risk function* $E[L(\hat{\theta}, \theta)]$, where θ is a parameter vector and $\hat{\theta}$ is an estimator of θ and the expectation is taken under the true parameter value θ. Moreover, the estimator $\hat{\theta}_2$ is said to *dominate* the estimator $\hat{\theta}_1$ with respect to a given loss function if $E[L(\hat{\theta}_2, \theta)] \leq E[L(\hat{\theta}_1, \theta)] \forall \theta$, with strict inequality for at least one value of θ. Depending on the loss function used, several estimators of $\mathbf{\Sigma}$ that dominate $\mathbf{W}/n$ have been proposed (see e.g. James and Stein (1961), Takemura (1984), Dey and Srinivasan (1985), Sheena (1995), Kubokawa (2005), Konno (2009) and Tsukuma (2014)), the majority of which are based on functions of the sample eigenvalues.

Furthermore, a class of estimators of $\mathbf{\Sigma}$ often considered are orthogonal invariant estimators , i.e. estimators $\hat{\mathbf{\Sigma}}$ that can be written as $\hat{\mathbf{\Sigma}} = \mathbf{H}\Phi(\mathbf{l})\mathbf{H}'$, $\Phi(\mathbf{l}) = \text{diag}(\phi_1(\mathbf{l}), \ldots, \phi_p(\mathbf{l}))$, $\phi_i(\mathbf{l}) > 0, i = 1, \ldots, p.$, where $\mathbf{l}$ is the vector of ordered sample eigenvalues of $\mathbf{W}$, and $\mathbf{H}$ is the orthogonal matrix of the eigenvalue decomposition $\mathbf{W} = \mathbf{HLH}$ with $\mathbf{L} = \text{diag}(\mathbf{l})$. The Stein-Haff identity, which expresses $E[\text{tr}(\mathbf{H}\Phi(\mathbf{l})\mathbf{H}'\mathbf{\Sigma}^{-1})]$ in terms of the function $\Phi(\mathbf{l})$, is a flexible tool that readily applies to evaluate various risk function of orthogonal estimators $\hat{\mathbf{\Sigma}}$. One such risk function is the one associated with Stein's loss[2] $E[L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma})] = E[\text{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}] - E[\log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}|] - p$ where the identity is directly applicable to the first term. Further, the identity can also be used in order to derive various moments of the Wishart distribution, as presented in for example Haff (1979).

Apart from the derivation by Stein (1977) and Haff (1979) in the case of the non-singular Wishart matrix, equivalent identities have also been presented in the case of a singular Wishart matrix (see Kubokawa and Srivastava (2008)), in the case of a complex Wishart matrix (see Konno (2009)) and in the case of elliptically countoured distributions (see Kubokawa and Srivastava (1999), Konno (2007) and Bodnar and Gupta (2009)).

In this paper, we generalize the Stein-Haff identity to the case of positive-definite and symmetric random matrices of the exponential family, given certain conditions on the density function of the considered distribution. For such a random matrix $\mathbf{S}$, the result expresses $E\left[\text{tr}\left(\mathbf{H\Phi(l)H}'\theta_2\right)\right]$, where $\mathbf{H}$ and $\mathbf{l}$ are the components of the eigenvalue decomposition $\mathbf{S} = \mathbf{HLH}'$ with $\mathbf{L} = \text{diag}(\mathbf{l})$ and θ_2 a matrix parameter, in terms of the function $\Phi(\mathbf{l})$ and the various components of the matrix distributions density function, a formula

[1] For a general discussion on the decision-theoretic framework, see for example Ferguson (1967).

[2] A commonly used loss function first considered in James and Stein (1961).

readily applicable to both estimation problems and derivation of moments. The identity is then applied to the matrix-variate gamma distribution, where it is used to evaluate estimators for samples of the matrix-variate gamma distribution with common scale matrix and different shape parameters. For the class of simple estimators consisting of a constant times the sum of the observed matrices, the estimator that minimizes Stein's loss is derived, and turns out to be the maximum likelihood estimator. Moreover, the identity is used to derive a condition for orthogonoally invariant estimators to dominate the maximum likelihood estimator, together with an example of such an estimator. Finally a small simulation study is conducted in order to support the dominance results.

The rest of the paper is organized as follows. Section 2 consists of the main contribution of this paper, the generalization of the Stein-Haff identity to matrices of the exponential family. Section 3 applies the identity to the matrix-variate gamma distribution and together with a simulation study to support the theoretical results. Section 4 concludes. Lemmas with proofs used throughout the paper can be found in the Appendix.

2 Stein-Haff identity for the exponential family

Let the $p \times p$ matrix $\mathbf{S}$ be a real, positive-definite, symmetric random matrix belonging to the exponential family. As such, the density function of $\mathbf{S}$ can be factorized as

$$f(\mathbf{S}) = a(\theta)h(\mathbf{S})e^{(\theta'\mathbf{t}(\mathbf{S}))}, \tag{1}$$

where θ is the canonical parameter and $\mathbf{t}(\mathbf{S})$ is the canonical statistic. Further, let $\mathbf{l}$ denote the $p \times 1$ vector of ordered eigenvalues of $\mathbf{S}$ and impose the following conditions:

$$h(\mathbf{S}) = u(\mathbf{l}) \tag{2}$$

$$\mathbf{t}(\mathbf{S}) = (\mathbf{v}(\mathbf{l})', \mathrm{vec}(\mathbf{S})')', \tag{3}$$

where $\mathrm{vec}(\cdot)$ is the vectorization operator[3]. As such, the above conditions require that $h(\mathbf{S})$ is dependent only on the eigenvalues $\mathbf{l}$ and that the canonical statistic can be decomposed into one part consisting of $\mathrm{vec}(\mathbf{S})$ and one part dependent only on $\mathbf{l}$. For notational convenience, let θ be decomposed as $\theta = (\theta_1', -\mathrm{vec}(\theta_2)')'$, where θ_1 is a vector of the same length as the vector $\mathbf{v}(\mathbf{l})$ and θ_2 is a $p \times p$ matrix. Note that in the case of a real, symmetric matrix $\mathbf{A}$,

[3]The operator that stacks all columns of a matrix into a vector.

the common matrix-to-scalar operators $\mathrm{tr}(\mathbf{A})$ and $|\mathbf{A}|$ depends only on the eigenvalues of $\mathbf{A}$. Hence conditions (2) and (3) still allows for a wide range of density funcitons.

As an example of an exponential distribution of the form (1) conforming to (2) and (3), consider the problem presented in Section 1. Thus, suppose $\mathbf{Z} \sim N_{n\times p}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Sigma})$ such that $\mathbf{S} = \mathbf{Z}'\mathbf{Z}$ follows a Wishart distribution with $n \in \mathbb{Z}$ degrees of freedom and postive-definite covariance matrix $\mathbf{\Sigma}$. Then the density of $\mathbf{S}$ can be expressed in the form (1) with $\theta = \mathrm{vec}(2\mathbf{\Sigma})^{-1}$, $a(\theta) = |\theta|^{n/2}\Gamma_p(n/2), h(\mathbf{S}) = u(\mathbf{l}) = \prod_{i=1}^{p} l_i^{(n-p-1)/2}$ and $\mathbf{t}(\mathbf{S}) = \mathrm{vec}(\mathbf{S})$.

Further, as discussed in Section 1, for problems concerning estimation of the parameters of a random matrix, it is often required to compute the expected value of a function of the observed random matrices. Such is for example the case when working with loss and risk functions in the decision-theoretic framework. Furthermore, the functions are often readily expressed in terms of the observed random matrices associated eigenvalues and eigenvectors. As such, we now derive the expectation of such functions with regard to distributions of the form (1). To this end, let $\mathcal{O}_p$ denote the set of $p \times p$ orthogonal matrices and let $\mathbf{S} = \mathbf{HLH}'$ be the eigendecomposition of $\mathbf{S}$, where $H \in \mathcal{O}_p$ and $\mathbf{L} = \mathrm{diag}(\mathbf{l})$. From Theorem 3.2.17 in Muirhead (1982), note that for a $p \times p$ positive-definite random matrix $\mathbf{S}$ with density function $f(\mathbf{S})$, the joint density of the p eigenvalues $l_1, \ldots, l_p$, where $l_1 > \ldots > l_p > 0$, is given by

$$\frac{\pi^{p^2/2}}{\Gamma_p(p/2)} \prod_{i<j}^{p} (l_i - l_j) \int_{\mathcal{O}_p} f(\mathbf{HLH}')(d\mathbf{H}). \tag{4}$$

Thus, letting $\mathcal{L}_p = \{\mathbf{l} | l_1 > l_2 > \ldots > l_p > 0\}$, we have for any scalar function $g(\mathbf{H}, \mathbf{L})$ with $E[g(\mathbf{H}, \mathbf{L})] < \infty$,

$$\begin{aligned} E[g(\mathbf{H}, \mathbf{L})] &= \frac{\pi^{p^2/2}}{\Gamma_p(p/2)} \int_{\mathcal{L}_p} \prod_{i<j} (l_i - l_j) \int_{\mathcal{O}_p} g(\mathbf{H}, \mathbf{L}) f(\mathbf{HLH}') d\mathbf{H} d\mathbf{L} \\ &= \frac{\pi^{p^2/2}}{\Gamma_p(p/2)} a(\theta) \int_{\mathcal{L}_p} \prod_{i<j} (l_i - l_j) u(\mathbf{l}) \exp(\theta_1' \mathbf{v}(\mathbf{l})) \\ &\quad \int_{\mathcal{O}_p} g(\mathbf{H}, \mathbf{L}) \exp(-\mathrm{vec}(\theta_2)' \mathrm{vec}(\mathbf{HLH}')) d\mathbf{H} d\mathbf{L} \\ &= \frac{\pi^{p^2/2}}{\Gamma_p(p/2)} a(\theta) \int_{\mathcal{L}_p} \prod_{i<j} (l_i - l_j) u(\mathbf{l}) \exp(\theta_1' \mathbf{v}(\mathbf{l})) \\ &\quad \int_{\mathcal{O}_p} g(\mathbf{H}, \mathbf{L}) \exp(-\mathrm{tr}(\theta_2 \mathbf{HLH}')) d\mathbf{H} d\mathbf{L}, \end{aligned} \tag{5}$$

where the second equality comes from inserting (1) and the third equality comes from the identity $\mathrm{vec}(\mathbf{A})'\mathrm{vec}(\mathbf{B}) = \mathrm{tr}(\mathbf{A}'\mathbf{B})$ together with the symmetry of $\mathbf{S}$. Now let $\mathbf{A} = \mathbf{H}'\theta_2\mathbf{H}$ and denote the elements of $\mathbf{A}$ as $a_{ij}(\mathbf{H})$. Then

$$\begin{aligned} \mathrm{tr}(\theta_2\mathbf{HLH}') &= \mathrm{tr}(\mathbf{LA}) \\ &= \sum_{i=1}^{p} l_i a_{ii}(\mathbf{H}). \end{aligned}$$

As such, (5) becomes

$$\frac{\pi^{p^2/2}}{\Gamma_p(p/2)} a(\theta) \int_{\mathcal{L}_p} \prod_{i<j} (l_i - l_j) u(\mathbf{l}) \exp\left(\theta_1' \mathbf{v}(\mathbf{l})\right) \int_{\mathcal{O}_p} g(\mathbf{H}, \mathbf{L}) \exp\left(-\sum_{i=1}^{p} l_i a_{ii}(\mathbf{H})\right) d\mathbf{H} d\mathbf{L}.$$

Further denote

$$\begin{aligned} c &= \frac{\pi^{p^2/2}}{\Gamma_p(p/2)} a(\theta) \\ b(\mathbf{l}) &= \prod_{i<j} (l_i - l_j) u(\mathbf{l}) \exp\left(\theta_1' \mathbf{v}(\mathbf{l})\right), \\ w(\mathbf{l}) &= \int_{\mathcal{O}_p} \exp\left(-\sum_{i=1}^{p} l_i a_{ii}(\mathbf{H})\right) d\mathbf{H}. \end{aligned}$$

and define, for $i = 1, \ldots, p$,

$$\begin{aligned} l_0 &= \infty \\ l_{p+1} &= 0 \\ \mathbf{l}_{(i)} &= (l_1, \ldots, l_{i-1}, l_{i+1}, \ldots, l_p) \\ \mathcal{L}_{(i)} &= \{\mathbf{l}_{(i)} | l_1 > \ldots > l_{i-1} > l_{i+1} > \ldots > l_p\}. \end{aligned}$$

We can now formulate the Stein-Haff identity for $\mathbf{S}$. The proof is a generalization of the derivations in Sheena (1995).

Theorem 1. *Let* $\mathbf{S}$ *be a real, positive-definite, symmetric* $p \times p$ *random matrix from the exponential family for which conditions (2) and (3) hold. Further let* $\mathbf{S} = \mathbf{HLH}'$ *be the eigendecomposition of* $\mathbf{S}$ *and let* $\Phi(\mathbf{l}) = diag(\phi_1(\mathbf{l}), \ldots, \phi_p(\mathbf{l}))$. *Moreover, assume that*

(i) All the following expectations are finite;

(ii) $\phi_i(\mathbf{l}) b(\mathbf{l}), i = 1, \ldots, p$ *is absolutely continuous with respect to* l_i*;*

(iii) $\phi_i(\mathbf{l}), i = 1, \ldots, p$ *satisfies*

$$\lim_{l_i \to l_{i+1}} \phi_i(\mathbf{l}) b(\mathbf{l}) w(\mathbf{l}) = 0 \quad \text{and} \quad \lim_{l_i \to l_{i-1}} \phi_i(\mathbf{l}) b(\mathbf{l}) w(\mathbf{l}) = 0 \quad \forall \mathbf{l} \in \mathcal{L}_p.$$

Then the following identity holds

$$E\left[tr\left(\mathbf{H}\mathbf{\Phi}(\mathbf{l})\mathbf{H}'\theta_2\right)\right]=\sum_{i=1}^{p}E\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i}+\frac{\partial u(\mathbf{l})}{\partial l_i}\frac{\phi_i(\mathbf{l})}{u(\mathbf{l})}+\theta_1'\phi_i(\mathbf{l})\frac{\partial\mathbf{v}(\mathbf{l})}{\partial l_i}+\sum_{i<j}\frac{\phi_i(\mathbf{l})-\phi_j(\mathbf{l})}{l_i-l_j}\right]. \quad (6)$$

Proof. Let $I=E\left[\mathrm{tr}\left(\mathbf{H}\mathbf{\Phi}(\mathbf{l})\mathbf{H}'\theta_2\right)\right]$. We then have

$$\begin{aligned}
I &= E\left[\mathrm{tr}\left(\mathbf{\Phi}(\mathbf{l})\mathbf{A}\right)\right]\\
&= \sum_{i=1}^{p}E\left[\phi_i(\mathbf{l})a_{ii}(\mathbf{H})\right]\\
&= \sum_{i=1}^{p}c\int_{\mathcal{L}_p}\phi_i(\mathbf{l})b(\mathbf{l})\int_{\mathcal{O}_p}a_{ii}(\mathbf{H})\exp\left(-\sum_{i=1}^{p}l_ia_{ii}(\mathbf{H})\right)d\mathbf{H}d\mathbf{L}\\
&= -\sum_{i=1}^{p}c\int_{\mathcal{L}_{(i)}}\int_{l_{i+1}}^{l_{i-1}}\phi_i(\mathbf{l})b(\mathbf{l})\frac{\partial}{\partial l_i}\left[\int_{\mathcal{O}_p}\exp\left(-\sum_{i=1}^{p}l_ia_{ii}(\mathbf{H})\right)d\mathbf{H}\right]dl_id\mathbf{l}_{(i)}\\
&= -\sum_{i=1}^{p}c\int_{\mathcal{L}_{(i)}}\int_{l_{i+1}}^{l_{i-1}}\phi_i(\mathbf{l})b(\mathbf{l})\frac{\partial w(\mathbf{l})}{\partial l_i}dl_id\mathbf{l}_{(i)}.
\end{aligned}$$

By condition (ii) we can apply integration by parts and write

$$\begin{aligned}
\int_{l_{i+1}}^{l_{i-1}}\phi_i(\mathbf{l})b(\mathbf{l})\frac{\partial w(\mathbf{l})}{\partial l_i}dl_i = {}& \lim_{l_i\to l_{i-1}}\phi_i(\mathbf{l})b(\mathbf{l})w(\mathbf{l})-\lim_{l_i\to l_{i+1}}\phi_i(\mathbf{l})b(\mathbf{l})w(\mathbf{l}).\\
&-\int_{l_{i+1}}^{l_{i-1}}\frac{\partial\phi_i(\mathbf{l})b(\mathbf{l})}{\partial l_i}w(\mathbf{l})dl_i.
\end{aligned}$$

Due to conditiion (iii), I can now be written

$$\begin{aligned}
I &= \sum_{i=1}^{p}\int_{\mathcal{L}_{(i)}}\int_{l_{i+1}}^{l_{i-1}}c\frac{\partial\phi_i(\mathbf{l})b(\mathbf{l})}{\partial l_i}w(\mathbf{l})dl_id\mathbf{l}_{(i)}\\
&= \sum_{i=1}^{p}E\left[\frac{1}{b(\mathbf{l})}\frac{\partial\phi_i(\mathbf{l})b(\mathbf{l})}{\partial l_i}\right]\\
&= \sum_{i=1}^{p}E\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i}+\phi_i(\mathbf{l})\frac{\partial b(\mathbf{l})}{\partial l_i}\frac{1}{b(\mathbf{l})}\right]\\
&= \sum_{i=1}^{p}E\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i}+\phi_i(\mathbf{l})\frac{\partial\log(b(\mathbf{l}))}{\partial l_i}\right].
\end{aligned}$$

Since $\log b(\mathbf{l})=\sum_{i<j}\log(l_i-l_j)+\log u(\mathbf{l})+\theta_1'\mathbf{v}(\mathbf{l})$ we have that

$$\frac{\partial\log b(\mathbf{l})}{\partial l_i}=\frac{\partial u(\mathbf{l})}{\partial l_i}\frac{1}{u(\mathbf{l})}+\theta_1'\frac{\partial\mathbf{v}(\mathbf{l})}{\partial l_i}+\sum_{i\neq j}\frac{1}{l_i-l_j}$$

and thus

$$\begin{aligned} I &= \sum_{i=1}^{p} E\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i} + \frac{\partial u(\mathbf{l})}{\partial l_i}\frac{\phi_i(\mathbf{l})}{u(\mathbf{l})} + \phi_i(\mathbf{l})\theta_1'(\mathbf{l})\frac{\partial \mathbf{v}(\mathbf{l})}{\partial l_i} + \sum_{i\neq j}\frac{\phi_i(\mathbf{l})}{l_i - l_j}\right] \\ &= \sum_{i=1}^{p} E\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i} + \frac{\partial u(\mathbf{l})}{\partial l_i}\frac{\phi_i(\mathbf{l})}{u(\mathbf{l})} + \phi_i(\mathbf{l})\theta_1'(\mathbf{l})\frac{\partial \mathbf{v}(\mathbf{l})}{\partial l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j}\right]. \end{aligned}$$

□

Apart from being useful in evaluating estimators, as shown in the subsequent section, Theorem 1 can also be applied in order to derive various moments of $\mathbf{S}$. For example, by noting that $\mathbf{S}^{-1} = (\mathbf{HLH}')^{-1} = \mathbf{HL}^{-1}\mathbf{H}'$ we can insert $\phi_i(\mathbf{l}) = 1/l_i, i = 1, \ldots, p$ in (6) to obtain

$$E[\text{tr}(\mathbf{S}^{-1}\theta_2)] = \sum_{i=1}^{p} E\left[-\frac{1}{l_i^2} + \frac{\partial u(\mathbf{l})}{\partial l_i}\frac{1}{u(\mathbf{l})l_i} + \frac{\theta_1'}{l_i}\phi_i(\mathbf{l})\frac{\partial \mathbf{v}(\mathbf{l})}{\partial l_i} - \sum_{i<j}\frac{1}{l_i l_j}\right].$$

3 Application to the matrix-variate gamma distribution

In this section, the identity derived in Section 2 is applied to the matrix-variate gamma distribution, a generalization of the gamma distribution to positive-definite matrices. Section 3.1 presents the distribution on the form (1) and with the identity, Section 3.2 applies the identity in order to derive a condition for which to beat the maximum likelihood estimator together with an example of such an estimator and Section 3.3 verifies the results through a simulation study.

3.1 Stein-Haff identity for the matrix-varaite gamma distribution

Let the $p \times p$ matrix $\mathbf{S}$ follow a matrix-variate gamma distribution with shape $\alpha > (p-1)/2$ and symmetric scale matrix $\mathbf{\Sigma} > 0$, denoted $\mathbf{S} \sim MG_p(\alpha, \mathbf{\Sigma})$. As such, the p.d.f of $\mathbf{S}$ is

$$f(\mathbf{S}) = \frac{|\mathbf{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}|\mathbf{S}|^{\alpha-(p+1)/2}\exp(\text{tr}\left(-\mathbf{\Sigma}^{-1}\mathbf{S}\right)). \tag{7}$$

This matrix distribution belongs to the exponential family, and the above p.d.f can be written on the form (1) by setting $\theta_1 = \alpha$, $\theta_2 = \Sigma^{-1}$, $\mathbf{t}(\mathbf{S}) =$

$(\sum_{i=1}^p \log l_i, \text{vec}(\mathbf{S})')'$, $a(\theta) = \frac{|\theta_2|^{\theta_1}}{\Gamma_p(\theta_1)}$ and $h(\mathbf{S}) = u(\mathbf{l}) = \prod_{i=1}^p l_i^{-(p+1)/2}$ and thus it also conforms to conditions (2) and (3).

By applying (6) we have that

$$E\left[\text{tr}\left(\mathbf{H}\mathbf{\Phi}(\mathbf{l})\mathbf{H}'\theta_2\right)\right] = \sum_{i=1}^{p} E\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i} + \left(\theta_1 - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j}\right]$$

or, in the parametrization $(\alpha, \mathbf{\Sigma})$,

$$E\left[\text{tr}\left(\mathbf{H}\mathbf{\Phi}(\mathbf{l})\mathbf{H}'\mathbf{\Sigma}^{-1}\right)\right] = \sum_{i=1}^{p} E\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i} + \left(\alpha - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j}\right]. \tag{8}$$

3.2 Estimation of the scale matrix $\mathbf{\Sigma}$

Now consider a sample of independent matrices $\mathbf{S}_1, \ldots, \mathbf{S}_n$, where $\mathbf{S}_k \sim MG_p(\alpha_k, \mathbf{\Sigma}), k = 1, \ldots, n$ and $\alpha_k > (p-1)/2$ are known while $\mathbf{\Sigma} > 0$ is unknown[4]. Further, suppose we are interested in an orthogonally invariant estimator for $\mathbf{\Sigma}$, such that the estimator can be written as

$$\hat{\mathbf{\Sigma}} = \mathbf{H}\mathbf{\Phi}(\mathbf{l})\mathbf{H}', \quad \Phi(\mathbf{l}) = \text{diag}(\phi_1(\mathbf{l}), \ldots, \phi_p(\mathbf{l})), \quad \phi_i(\mathbf{l}) > 0, i = 1, \ldots, p.$$

Moreover, assume that we want to minizime the risk for this estimator in terms of Stein's loss function

$$L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = \text{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p, \tag{9}$$

which has the associated risk function

$$E[L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma})] = E[\text{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1})] - E[\log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}|] - p. \tag{10}$$

Consider first estimators of the form $\hat{\mathbf{\Sigma}} = d\mathbf{V}$, where d is a constant and $\mathbf{V} = \sum_{k=1}^n \mathbf{S}_k$. By Lemma A3, $\mathbf{V} \sim MG_p(q, \mathbf{\Sigma})$, where $q = \sum_{k=1}^n \alpha_k$. Further, letting $\mathbf{V} = \mathbf{H}\mathbf{L}\mathbf{H}'$, such estimators can be written as $\hat{\mathbf{\Sigma}} = \mathbf{H}\mathbf{\Phi}(\mathbf{l})\mathbf{H}'$ where $\phi_i(\mathbf{l}) = dl_i$. As such, the first term in the risk function (10) becomes,

[4]Comparable to the case of sample covariance matrices for a multivariate normal distribution with a common unknown covariance matrix $\mathbf{\Sigma}$.

using (8),

$$
\begin{aligned}
E[\mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1})] &= E\left[\mathrm{tr}\left(\mathbf{H}\mathbf{\Phi}(\mathrm{l})\mathbf{H}'\mathbf{\Sigma}^{-1}\right)\right] \\
&= \sum_{i=1}^{p} E\left[d + d\left(q - \frac{p+1}{2}\right) + d\sum_{i<j} 1\right] \\
&= \frac{dp(p-1)}{2} + \sum_{i=1}^{p} d + d\left(q - \frac{p+1}{2}\right) \\
&= \frac{dp(p-1)}{2} + dp + dpq - \frac{dp(p+1)}{2} \\
&= dpq
\end{aligned}
$$

Let $\mathbf{M} \sim MG_p(q, 2\mathbf{I}_p)$ and $t_{ii} \sim \Gamma(q-(i+1)/2, 2)$. The second term of (10) can then be written as

$$
\begin{aligned}
E[|\log \hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}|] &= E[\log|d\mathbf{V}\mathbf{\Sigma}^{-1}|] \\
&= E[\log|\mathbf{\Sigma}^{-1/2}\mathbf{V}\mathbf{\Sigma}^{-1/2}|] + p\log d \\
&= E[\log|2^{-1}\mathbf{M}|] + p\log d \\
&= E[\log|\mathbf{M}|] - p\log 2 + p\log d| \\
&= E[\log|\mathbf{T}\mathbf{T}'|] - p\log 2 + p\log d \\
&= E\left[\log\prod_{i=1}^{p} t_{ii}^2\right] - p\log 2 + p\log d \\
&= \sum_{i=1}^{p} E[\log t_{ii}^2] - p\log 2 + p\log d \\
&= \sum_{i=1}^{p} \psi(q-(i+1)/2) + p\log d,
\end{aligned}
$$

where the third equality is due to Lemma A4, the fifth equality is due to Lemma A5 and $\psi(\cdot)$ is the digamma function. As such, we can write (10) as

$$
E[L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma})] = dpq - \sum_{i}^{p} \psi(q-(i+1)/2) - p\log d - p. \tag{11}
$$

Deriving (11) w.r.t. d and setting it equal to zero we obtain that the risk function for estimators of the form $d\mathbf{V}$ has its minimum at $d = q^{-1} = 1/(\sum_{k=1}^{n} \alpha_k)$, which in accordance with Lemma A6 also is the maximum likelihood estimate of $\mathbf{\Sigma}$ in the considered case. In addition, note that in this case the risk function is constant with respect to $\mathbf{\Sigma}$.

Further, the above results allows us to obtain a condition under which estimators dominate the maximum likelihood estimator $\hat{\mathbf{\Sigma}}_{MLE} = \mathbf{V}/q$.

Proposition 1. *Let* $\mathbf{S}_k \sim MG_p(\alpha_k, \boldsymbol{\Sigma}), k = 1, \ldots, n,$ *where* $\alpha_k > (p-1)/2$ *are known,* $q = \sum_i^n \alpha_k$, $\sum_{k=1}^n \mathbf{S}_k = \mathbf{HLH}'$ *and let* $\hat{\boldsymbol{\Sigma}}_D = \mathbf{H\Phi(l)H}'$, *with* $\boldsymbol{\Phi}(\mathbf{l}) = \mathrm{diag}(\phi_1(\mathbf{l}), \ldots, \phi_p(\mathbf{l}))$, $\phi_i(\mathbf{l}) > 0, i = 1, \ldots, p$, *be an orthogonal invariant estimator of* $\boldsymbol{\Sigma}$. *Then* $\hat{\boldsymbol{\Sigma}}_D$ *will dominate* $\hat{\boldsymbol{\Sigma}}_{MLE}$, *with regard to Stein's loss function (9), if and only if*

$$\sum_{i=1}^p E\left[\frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + \left(q - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j} - \log\frac{\phi_i(\mathbf{l})}{l_i}\right] \leq p + p\log q, \tag{12}$$

for all values of $\boldsymbol{\Sigma}$, *with strict inequality for at least one value of* $\boldsymbol{\Sigma}$.

Proof. We have that $\hat{\boldsymbol{\Sigma}}_D$ will dominate $\hat{\boldsymbol{\Sigma}}_{MLE}$ if and only if

$$E[L(\hat{\boldsymbol{\Sigma}}_D, \boldsymbol{\Sigma})] \leq E[L(\hat{\boldsymbol{\Sigma}}_{MLE}, \boldsymbol{\Sigma})], \tag{13}$$

for all values of $\boldsymbol{\Sigma}$, with strict inequality for at least one value of $\boldsymbol{\Sigma}$. By (8) we have, since $\sum_{k=1}^n \mathbf{S}_k \sim MG_p(q, \boldsymbol{\Sigma})$, that

$$\begin{aligned}
E[L(\hat{\boldsymbol{\Sigma}}_D, \boldsymbol{\Sigma})] &= E[\mathrm{tr}(\mathbf{H\Phi(l)H}'\boldsymbol{\Sigma}^{-1})] - E[\log|\mathbf{H\Phi(l)H}'\boldsymbol{\Sigma}^{-1}|] - p \\
&= \sum_{i=1}^p E\left[\frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + \left(q - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j}\right] \\
&\quad - E\left[\log\prod_{i=1}^p \phi_i(\mathbf{l})\right] + \log|\boldsymbol{\Sigma}| - p \\
&= \sum_{i=1}^p E\left[\frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + \left(q - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j} - \log\phi_i(\mathbf{l})\right] \\
&\quad + \log|\boldsymbol{\Sigma}| - p
\end{aligned} \tag{14}$$

Further we have

$$\begin{aligned}
E[L(\hat{\boldsymbol{\Sigma}}_{MLE}, \boldsymbol{\Sigma})] &= E\left[\mathrm{tr}\left(\frac{\mathbf{V}}{q}\boldsymbol{\Sigma}^{-1}\right)\right] - E\left[\log\left|\frac{\mathbf{V}}{q}\boldsymbol{\Sigma}^{-1}\right|\right] - p \\
&= p - E[\log|\mathbf{V}|] + \log|\boldsymbol{\Sigma}| - p\log\frac{1}{q} - p \\
&= -E\left[\log\prod_{i=1}^p l_i\right] + \log|\boldsymbol{\Sigma}| + p\log q \\
&= -\sum_i^p E[\log l_i] + \log|\boldsymbol{\Sigma}| + p\log q.
\end{aligned} \tag{15}$$

Inserting (14) and (15) into (13) gives the desired result. □

Finally, Proposition 1 can be applied in order to derive an estimator that dominates $\hat{\boldsymbol{\Sigma}}_{MLE}$. Here we will consider orthogonally invariant estimators where $\boldsymbol{\Phi}(\mathbf{l}) = \text{diag}(\phi_1(\mathbf{l}), \ldots, \phi_p(\mathbf{l}))$ is of the form $\phi_i(\mathbf{l}) = d_i l_i, i = 1, \ldots, p$, where d_i is a constant.

Proposition 2. *Let* $\mathbf{S}_k \sim MG_p(\alpha_k, \boldsymbol{\Sigma}), k = 1, \ldots, n,$ *where* $\alpha_k > (p-1)/2$ *are known,* $q = \sum_i^n \alpha_k$, $\sum_{k=1}^n \mathbf{S}_k = \mathbf{HLH}'$ *and let* $\hat{\boldsymbol{\Sigma}}_1 = \mathbf{H}\boldsymbol{\Phi}(\mathbf{l})\mathbf{H}'$ *with* $\boldsymbol{\Phi}(\mathbf{l}) = diag(d_1 l_1, \ldots, d_p l_p)$ *and*

$$d_i = \frac{1}{q + (p+1)/2 - i}, \quad i = 1, \ldots, p, \tag{16}$$

be an esitmator of $\boldsymbol{\Sigma}$. *Then* $\hat{\boldsymbol{\Sigma}}_1$ *dominates* $\hat{\boldsymbol{\Sigma}}_{MLE}$ *with regard to Stein's loss function (9).*

Proof. First, note that by definition $l_1 > \cdots > l_p$ and further that $d_1 < \cdots < d_p$. By (12) in Proposition 1 we have that if

$$\sum_{i=1}^p E\left[d_i + \left(q - \frac{p+1}{2}\right) d_i + \sum_{i<j} \frac{d_i l_i - d_j l_j}{l_i - l_j} - \log d_i\right] < p + p \log q \tag{17}$$

$\hat{\boldsymbol{\Sigma}}_1$ will dominate $\hat{\boldsymbol{\Sigma}}_{MLE}$. Now (17) can be written as

$$\begin{aligned}
p + p\log q &> \sum_{i=1}^p \left(1 + q - \frac{p+1}{2}\right) d_i + E\left[\sum_{i<j} \frac{d_i l_i - d_j l_j}{l_i - l_j}\right] - \log d_i \\
&= \sum_{i=1}^p \left(1 + q - \frac{p+1}{2}\right) d_i + E\left[\sum_{i<j} \frac{l_j}{l_i - l_j}(d_i - d_j)\right] + \sum_{i<j} d_i - \log d_i \\
&= \sum_{i=1}^p \left(1 + q - \frac{p+1}{2}\right) d_i + \sum_{i<j} E\left[\frac{l_j}{l_i - l_j}\right](d_i - d_j) + d_i(p-i) - \log d_i \\
&= \sum_{i=1}^p \left(q + \frac{p+1}{2} - i\right) d_i + \sum_{i<j} E\left[\frac{l_j}{l_i - l_j}\right](d_i - d_j) - \log d_i
\end{aligned}$$

Let $m_i = \sum_{i<j} E\left[\frac{l_j}{l_i - l_j}\right](d_i - d_j)$ and note that $m_i < 0, i = 1, \ldots, p$ since $(l_j)/(l_i - l_j) > 0$ and $d_i - d_j < 0$. Inserting $d_i = 1/(q + (p+1)/2 - i)$ we get

$$\begin{aligned}
p + p\log q &> \sum_{i=1}^p 1 + m_i + \log(q + (p+1)/2 - i) \\
p\log q &> \sum_{i=1}^p \log(q + (p+1)/2 - i) + \sum_{i=1}^p m_i.
\end{aligned}$$

Since $\sum_{i=1}^{p} m_i < 0$, it will suffice to show that $p \log q > \sum_{i=1}^{p} \log(q + (p+1)/2 - i)$, or similarly

$$q^p > \prod_{i}^{p} (q + (p+1)/2 - i). \tag{18}$$

To this end set $a_i = (p+1)/2 - i$ and note that $a_i = -a_{p-i+1}$. Further we have that

$$\begin{aligned} (q + a_i)(q + a_{p-i+1}) &= (q + a_i)(q - a_i) \\ &< q^2. \end{aligned} \tag{19}$$

If p is even we can write

$$\prod_{i}^{p} (q + (p+1)/2 - i) = \prod_{i}^{p/2} (q + a_i) \prod_{i}^{p/2} (q - a_i) < q^p \tag{20}$$

where the inequality is in accordance with (19). In contrary if p is odd we can write

$$\prod_{i}^{p} (q + (p+1)/2 - i) = (q) \prod_{i}^{(p-1)/2} (q + a_i) \prod_{i}^{(p-1)/2} (q - a_i) \leq q^p, \tag{21}$$

where the inequality again is due to (19). Combining (20) and (21) shows (18) which completes the proof. □

As an example, consider $p = 3$, such that the constants of the estimator $\hat{\mathbf{\Sigma}}_1$ becomes $d_1 = 1/(q+1), d_2 = 1/q, d_3 = 1/(q-1)$. Similarly the MLE can be expressed in this form with $d_1 = d_2 = d_3 = 1/q$. As such, comparing with the equivalent constants in the MLE, the constant of $\hat{\mathbf{\Sigma}}_1$ associated with the largest sample eigenvalue is smaller than $1/q$ while the constant associated with the smallest eigenvalue is larger than $1/q$. Thus this estimator aims to pull sample eigenvalues towards a middle point. Further note that when $n = 1$ the estimator derived in Proposition 2 is closely related to the estimator derived by Stein (1977) and Dey and Srinivasan (1985) regarding the estimation of the covariance matrix of a normal population.

3.3 Simulation study

In order to illustrate that $\hat{\mathbf{\Sigma}}_1$, defined in Proposition 2, dominates $\hat{\mathbf{\Sigma}}_{MLE}$ in terms of Stein's loss, we conduct a small Monte Carlo simulation study. As such, we first define the difference in estimation loss r as

$$r = L(\hat{\mathbf{\Sigma}}_{MLE}, \mathbf{\Sigma}) - L(\hat{\mathbf{\Sigma}}_1, \mathbf{\Sigma}), \tag{22}$$

such that $E[r] > 0$ for all values of $\mathbf{\Sigma}$. Further define the matrix $\mathbf{J}_p = (0.5^{|i-j|})_{i,j}, i,j = 1,\ldots,p$. We now perform a simulation study according to the following algorithm:

1. For each combination of matrix dimension $p = \{2, 4, 10\}$ and parameters $\alpha = \{5, 10, 100\}$ and $\mathbf{\Sigma} = \{\mathbf{I}_p, \mathbf{J}_p\}$, draw a sample of $n = 10$ matrices $\mathbf{S} \sim MG_p(\alpha, \mathbf{\Sigma})$.

2. For each such sample, estimate $\hat{\mathbf{\Sigma}}_1$ and $\hat{\mathbf{\Sigma}}_{MLE}$, and compute r.

3. Repeat the above steps 1000 times and compute the average value of r for each combination of p, α and $\mathbf{\Sigma}$.

Table 1 summarizes the results. First, all average values of r are positive, as is expected since $E[L(\hat{\mathbf{\Sigma}}_{MLE}, \mathbf{\Sigma})] > E[L(\hat{\mathbf{\Sigma}}_1, \mathbf{\Sigma})]$. Further, for a given value of α and structure of $\mathbf{\Sigma}$, r tends to increase as the dimension p increases. Conversely, r tends to decrease as α increases. Additionally, in all the considered cases, the loss difference is smaller when the off-diagonal elements of $\mathbf{\Sigma}$ are non-zero compared to when they are zero. This suggest that the risk improvement is greater for the identity matrix, similar to for example the conclusions of Dey and Srinivasan (1985) in the case of a normal population. Finally note that since $E[L(\hat{\mathbf{\Sigma}}_{MLE}, \mathbf{\Sigma})]$ is constant with respect to $\mathbf{\Sigma}$, any differences in $E[r]$ due to changes in $\mathbf{\Sigma}$ stems from the term $E[L(\hat{\mathbf{\Sigma}}_1, \mathbf{\Sigma})]$.

	$\mathbf{\Sigma} = \mathbf{I}_p$			$\mathbf{\Sigma} = \mathbf{J}_p$		
α/p	2	4	10	2	4	10
5	0.0024	0.017	0.17	$4.1 \cdot 10^{-4}$	0.0043	0.071
10	$8.5 \cdot 10^{-4}$	0.0060	0.061	$6.4 \cdot 10^{-5}$	$8.9 \cdot 10^{-4}$	0.019
100	$2.7 \cdot 10^{-5}$	$2.0 \cdot 10^{-4}$	0.0020	$9.9 \cdot 10^{-7}$	$9.5 \cdot 10^{-6}$	$2.0 \cdot 10^{-4}$

Table 1: *The average of r, the difference in Stein's losses $L(\hat{\mathbf{\Sigma}}_{MLE}, \mathbf{\Sigma})$ and $L(\hat{\mathbf{\Sigma}}_1, \mathbf{\Sigma})$, for various values of p, α and $\mathbf{\Sigma}$.*

4 Conclusion

In this paper, we derive the Stein-Haff identity for random matrices of the exponential family, generalizing existent results. This identity is then applied to the matrix-varaite gamma distribution, where it is implemented in order to derive an estimator that dominates the MLE in terms of Stein's loss. In order to support these derivations, a simulation study is conducted, where the results suggest that the risk improvement is greater when the scale matrix is the identity matrix rather than a matrix with non-zero off-diagonal elements, and that improvement tends to increases with dimension.

Topics for future research includes deriving the Stein-Haff identity for even more general random matrices. One approach is to relax the condition of symmetry, or the requirements on the density function imposed by (2) and (3) in the case of the exponential family. Another related field of interest is how to improve estimators in the case of samples from the matrix-variate gamma distribution with unknown shape parameters.

Appendix

In this section we present several results regarding the matrix-variate gamma distribution needed for the derivations in Section 3.2, of which most are directly related to results on the Wishart distribution.

Lemma A1. *If $\alpha > (p-1)/2$ and $\mathbf{\Sigma}$ is a symmetric $p \times p$ matrix with $\mathbf{\Sigma} > 0$, where α and the elements of $\mathbf{\Sigma}$ are real valued, then*

$$\int_{S>0} \exp(-tr(\mathbf{\Sigma}^{-1}\mathbf{S}))|\mathbf{S}|^{\alpha-(p+1)/2} d\mathbf{S} = \Gamma_p(\alpha)|\mathbf{\Sigma}|^{\alpha}.$$

Proof. In the integral, make the variable change $\mathbf{S} = \mathbf{\Sigma}^{1/2}\mathbf{A}\mathbf{\Sigma}^{1/2}$. By Theorem 2.1.6 in Muirhead (1982) $(d\mathbf{S}) = |\mathbf{\Sigma}|^{(p+1)/2}(d\mathbf{A})$ and as such we have

$$\begin{aligned}\int_{S>0} \exp(-\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}))|\mathbf{S}|^{\alpha-(p+1)/2} d\mathbf{S} &= \int_{A>0} \exp(-\mathrm{tr}(\mathbf{A}))|\mathbf{A}|^{\alpha-(p+1)/2} d\mathbf{A}|\mathbf{\Sigma}|^{\alpha} \\ &= \Gamma_p(\alpha)|\mathbf{\Sigma}|^{\alpha},\end{aligned}$$

in accordance of the definition of the multivariate gamma function $\Gamma_p(\alpha)$. □

Lemma A2. *If $\mathbf{S} \sim MG_p(\alpha, \mathbf{\Sigma})$ then the characteristic function of $\mathbf{S}$ is*

$$\varphi(\mathbf{\Theta}) = E\left[\exp(tr(i\mathbf{T}\mathbf{S}))\right] = |\mathbf{I} - i\mathbf{\Sigma}\mathbf{T}|^{-\alpha},$$

where $\mathbf{\Theta}$ is a symmetric $p \times p$ matrix, $\mathbf{T} = (t_{ij}), i, j = 1, \dots, p$ and

$$t_{ij} = \begin{cases} \theta_{ij} & \text{if } i = j \\ \theta_{ij}/2 & \text{if } i \neq j \end{cases}$$

Proof. By the density of $\mathbf{S}$, as noted in (7), we have

$$\begin{aligned}
E\left[\exp(\mathrm{tr}(i\mathbf{TS}))\right] &= \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}\int_{\mathbf{S}>0}|\mathbf{S}|^{\alpha-(p+1)/2}\exp(\mathrm{tr}(-\boldsymbol{\Sigma}^{-1}\mathbf{S}))\exp(\mathrm{tr}(i\mathbf{TS}))d\mathbf{S}\\
&= \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}\int_{\mathbf{S}>0}|\mathbf{S}|^{\alpha-(p+1)/2}\exp(\mathrm{tr}(i\mathbf{TS}-\boldsymbol{\Sigma}^{-1}\mathbf{S}))d\mathbf{S}\\
&= \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}\int_{\mathbf{S}>0}|\mathbf{S}|^{\alpha-(p+1)/2}\exp(-\mathrm{tr}((\boldsymbol{\Sigma}^{-1}-i\mathbf{T})\mathbf{S}))d\mathbf{S}. \quad (23)
\end{aligned}$$

By setting $\mathbf{B}^{-1}=\boldsymbol{\Sigma}^{-1}-i\mathbf{T}$ we can by the aid of Lemma A1 write (23) as

$$\begin{aligned}
\frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}\int_{\mathbf{S}>0}|\mathbf{S}|^{\alpha-(p+1)/2}\exp(-\mathrm{tr}(\mathbf{B}^{-1}\mathbf{S}))d\mathbf{S} &= \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}\Gamma_p(\alpha)|\mathbf{B}|^{\alpha}\\
&= |\boldsymbol{\Sigma}|^{-\alpha}|\mathbf{B}^{-1}|^{-\alpha}\\
&= |\boldsymbol{\Sigma}|^{-\alpha}|\boldsymbol{\Sigma}^{-1}-i\mathbf{T}|^{-\alpha}\\
&= |\boldsymbol{\Sigma}|^{-\alpha}|(\mathbf{I}_p-i\mathbf{T}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}|^{-\alpha}\\
&= |\mathbf{I}_p-i\mathbf{T}\boldsymbol{\Sigma}|^{-\alpha}
\end{aligned}$$

□

Lemma A3. *Let $\mathbf{S}_1,\dots,\mathbf{S}_k$ be independent and $\mathbf{S}_k \sim MG_p(\alpha_k,\boldsymbol{\Sigma}), k=1,\dots,n$. Then*

$$\sum_{k=1}^{n}\mathbf{S}_k \sim MG_p(\alpha,\boldsymbol{\Sigma}),$$

where $\alpha=\sum_{k=1}^{n}\alpha_k$.

Proof. Since $\mathbf{S}_1,\dots,\mathbf{S}_k$ are independent, the characteristic function of $\sum_{k=1}^{n}\mathbf{S}_k$ is the product of the characteristic funcitons of $\mathbf{S}_1,\dots,\mathbf{S}_k$. It is as such, in accordance with Lemma A2,

$$\prod_{k=1}^{n}|\mathbf{I}-i\boldsymbol{\Sigma}\mathbf{T}|^{-\alpha_k}=|\mathbf{I}-i\boldsymbol{\Sigma}\mathbf{T}|^{-\alpha},$$

which is the characteristic funciton of $MG_p(\alpha,\boldsymbol{\Sigma})$, completing the proof. □

Lemma A4. *If $\mathbf{S}\sim MG_p(\alpha,\boldsymbol{\Sigma})$ and $\mathbf{M}$ is a $k\times p$ matrix of rank k then $\mathbf{MSM}'\sim MG_k(\alpha,\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}')$.*

Proof. By Lemma A2 we have that the characteristic function of $\mathbf{MSM}'$ is

$$\begin{aligned}
E[\exp(\mathrm{tr}(i\mathbf{TMSM}'))] &= E[\exp(\mathrm{tr}(i\mathbf{M}'\mathbf{TMS}))]\\
&= |\mathbf{I}_p-i\mathbf{M}'\mathbf{TM}\boldsymbol{\Sigma}|^{-\alpha}\\
&= |\mathbf{I}_k-i\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}'\mathbf{T}|^{-\alpha}\\
&= |\mathbf{I}_k-i\mathbf{TM}\boldsymbol{\Sigma}\mathbf{M}'|^{-\alpha}, \quad (24)
\end{aligned}$$

where the third equality is due to Sylvester's determinant identity (see e.g. Harville (1997)). The right side of (24) is the characteristic function of $MG_k(\alpha, \mathbf{M\Sigma M}')$ which proves the result. $\square$

Lemma A5. *Let* $\mathbf{S} \sim MG_p(\alpha, 2\mathbf{I}_p)$ *with* $\alpha > (p-1)/2$. *Further define* $\mathbf{S} = \mathbf{TT}'$ *where* $\mathbf{T} = (t_{ij})$ *is the lower-triangular Cholesky root of* $\mathbf{S}$. *Then the following holds:*

(i) t_{ij}, $1 \leq j \leq i \leq n$ *are mutually independent;*

(ii) $t_{ij} \sim \mathcal{N}(0,1)$ *(standard normal distribution) for* $1 \leq j < i \leq n$*;*

(iii) $t_{ii}^2 \sim \Gamma(\alpha - \frac{i-1}{2}, 2)$ *(gamma distribution with shape* $(\alpha - \frac{i-1}{2})$ *and scale* 2*) for* $i = 1, \ldots, n$.

Proof. We have

$$\begin{aligned}
\mathrm{tr}(\mathbf{S}) &= \sum_{j\leq i}^{p} t_{ij}^2 \\
|\mathbf{S}| &= \prod_{i=1}^{p} t_{ii}^2 \\
(d\mathbf{S}) &= 2^p \prod_{i=1}^{p} t_{ii}^{p+1-i} \bigwedge_{j\leq i}^{p} dt_{ij},
\end{aligned}$$

where the product $2^n \prod_{i=1}^{n} t_{ii}^{n+1-i}$ denotes the Jacobian of the transformation $\mathbf{S} \rightarrow \mathbf{T}$. Substituting the above equalitites into (7), including the volume element $d\mathbf{S}$ and with scale matrix $2\mathbf{I}_p$, we obtain the density of $\mathbf{T}$ as

$$f(\mathbf{T}) = \frac{2^p}{2^{p\alpha}\Gamma_p(\alpha)} e^{-\sum_{i\leq j}^{p} t_{ij}^2} \prod_{i=1}^{p} t_{ii}^{2\alpha-i} \bigwedge_{j\leq i}^{p} dt_{ij}. \quad (25)$$

Further note that

$$\begin{aligned}
\Gamma_p(\alpha) &= \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma(\alpha - \frac{i-1}{2}) \\
2^{p\alpha - p(p-1)/4 - p} &= 2^{\sum_{i=1}^{p}(\alpha - \frac{i-1}{2} - 1)}.
\end{aligned}$$

As such, (25) can be written

$$\begin{aligned}
f(\mathbf{T}) &= \prod_{j<i}^{p} \frac{1}{\sqrt{2\pi}} e^{-\frac{t_{ij}^2}{2}} dt_{ij} \prod_{i=1}^{p} \frac{1}{2^{\alpha-(i-1)/2-1}\Gamma(\alpha - \frac{i-1}{2})} e^{-\frac{t_{ii}^2}{2}} t_{ii}^{2\alpha-i} dt_{ii}. \\
&= \prod_{j<i}^{p} \frac{1}{\sqrt{2\pi}} e^{-\frac{t_{ij}^2}{2}} dt_{ij} \prod_{i=1}^{p} \frac{1}{2^{\alpha-(i-1)/2}\Gamma(\alpha - \frac{i-1}{2})} e^{-\frac{t_{ii}^2}{2}} (t_{ii}^2)^{\alpha - \frac{i-1}{2} - 1} dt_{ii}^2,
\end{aligned}$$

which is the joint density of the independent random variables $t_{ij} \sim \mathcal{N}(0,1)$ and $t_{ii}^2 \sim \Gamma(\alpha - \frac{i-1}{2}, 2), 1 \leq j < i \leq p$. $\square$

Lemma A6. *Consider an i.i.d. sample* $\mathbf{S}_1, \ldots, \mathbf{S}_k$, *where* $\mathbf{S}_k \sim MG_p(\alpha_k, \mathbf{\Sigma}), k = 1, \ldots, n$, $\alpha_k > (p-1)/2$ *are known and* $q = \sum_{k=1}^n \alpha_k$. *The maximum likelihood esitmate of* $\mathbf{\Sigma}$ *is then given by*

$$\hat{\mathbf{\Sigma}}_{MLE} = \frac{\sum_{k=1}^n \mathbf{S}_k}{q}$$

Proof. The log-likelihood function for the sample $\mathbf{S}_1, \ldots, \mathbf{S}_n$ is

$$l(\mathbf{S}_1, \ldots, \mathbf{S}_n) = -q \log |\mathbf{\Sigma}| - n \log \Gamma_p(\alpha) + \left(\alpha + \frac{p+1}{2}\right) \sum_{k=1}^n \log |\mathbf{S}_k| - \operatorname{tr}\left(\mathbf{\Sigma}^{-1} \sum_{k=1}^n \mathbf{S}_k\right).$$

Deriving by $\mathbf{\Sigma}$ and equating to zero we obtain

$$\begin{aligned} q\mathbf{\Sigma}^{-1} &= \mathbf{\Sigma}^{-1} \sum_{k=1}^n \mathbf{S}_k \mathbf{\Sigma}^{-1} \\ \hat{\mathbf{\Sigma}} &= \frac{\sum_{k=1}^n \mathbf{S}_k}{q}, \end{aligned}$$

as desired. $\square$

Acknowledgement

The author is grateful to Professor Joanna Tyrcha for valuable comments and suggestions.

References

Bodnar, T. and Gupta, A. (2009). An identity for multivariate elliptically contoured matrix distribution. *Statistics & Probability Letters*, 79:1327–1330.

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *Ann. Statist. 13*, 13:1581–1591.

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

Haff, L. R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, 9:531–542.

Harville, D. (1997). *Matrix Algebra from Statistician's Perspective.* Springer, New York.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In: *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1:361–380.

Konno, Y. (2007). Improving on the sample covariance matrix for a complex elliptically contoured distribution. *Journal of Statistical Planning and Inference*, 7:2475–2486.

Konno, Y. (2009). Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss. *J. Multivariate Anal.*, 100:2237–2253.

Kubokawa, T. (2005). A revisit to estimation of the precision matrix of the Wishart distribution. *J. Statist. Res.*, 39:91–114.

Kubokawa, T. and Srivastava, M. (1999). Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution. *Ann. Statist.*, 27:600–609.

Kubokawa, T. and Srivastava, M. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high dimensional data. *J. Multivariate Anal.*, 99:1906–1928.

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

Sheena, Y. (1995). Unbiased estimator of risk for an orthogonal invariant estimator of a covariance matrix. *Journal of the Japan Statistical Society.*, 25(1):35–48.

Stein, C. (1977). Lectures on the theory of estimation of many parameters. *Studies in the Statistical Theory of Estimation I (*eds. I. A. Ibraimov and M. S. Nikulin*), Proceeding of Scientific Seiminars of the Steklov Institute, Leningrad Division*, 74:4–65.

Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukaba J. Math.*, 8:367–376.

Tsukuma, H. (2014). Improvement on the best invariant estimators of the normal covariance and precision matrices via a lower triangular subgroup. *J. Japan Statist. Soc.*, 44:195–218.

Van der Vaart, H. (1961). On certain characteristics of the distribution of the latent roots of a symmetric random matrix under general conditions. *Ann. Math. Statist.*, 32:864–873.