Теорія ймовірностей
та математична статистика
Вип. 2(99)/2018, с. 7–18

Teoriya Ĭmovirnosteĭ
ta Matematychna Statystyka
No. 2(99)/2018, pp. 7–18

UDC 519.21

STEIN–HAFF IDENTITY FOR THE EXPONENTIAL FAMILY

G. ALFELT

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.

Key words 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 [15] and [5] regarding the problem of estimating the covariance matrix of multivariate normal populations. Consider a sample of n i.i.d. $p \times 1$ vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ (where $n > p$) following a multivariate normal distribution with mean vector μ and covariance matrix $\mathbf{\Sigma}$. Then the $p \times p$ matrix $\mathbf{W} = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ follows a Wishart distribution with $k = n - 1$ degrees of freedom and covariance matrix $\mathbf{\Sigma}$. As such, the covariance matrix of such a population is commonly estimated using the unbiased estimator $\mathbf{W}/k$. However, the eigenvalues of the estimator $\mathbf{W}/k$ tend to spread out more over the positive real line, than the corresponding 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 $\mathbf{W}/k$, l_1 is a positively biased estimator of λ_1 and l_p is a negatively biased estimator of λ_p (see e. g. [18]). As such, it can often be useful to consider estimators that aim to decrease larger sample eigenvalues and increase smaller sample eigenvalues.

Additionally, the problem of estimating 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* $\mathsf{E}[L(\hat{\theta}, \theta)]$, 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 $\mathsf{E}[L(\hat{\theta}_2, \theta)] \leq \mathsf{E}[L(\hat{\theta}_1, \theta)]$ $\forall\theta$, with strict inequality for at least one value of θ. Depending on the loss function used, several estimators of $\mathbf{\Sigma}$ that dominate $\mathbf{W}/k$ have been proposed (see e.g. [2, 7, 9, 10, 14, 16] and [17]), the majority of which are based on functions of the sample eigenvalues.

Furthermore, a class of estimators of $\mathbf{\Sigma}$ often considered is orthogonal invariant estimators , i. e. estimators $\hat{\mathbf{\Sigma}}$ that can be written as

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

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 $\mathbf{L} = \operatorname{diag}(\mathbf{l})$. The Stein–Haff identity, which expresses $\mathsf{E}[\operatorname{tr}(\mathbf{H}\Phi(\mathbf{l})\mathbf{H}'\mathbf{\Sigma}^{-1})]$ in terms of the function $\Phi(\mathbf{l})$, is a flexible tool that

[1] For a general discussion on the decision-theoretic framework, see for example [3].

readily applies to the evaluation of various risk functions of orthogonal estimators $\hat{\mathbf{\Sigma}}$. One such risk function is the one associated with Stein's loss [2]

$$\mathsf{E}[L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma})] = \mathsf{E}[\operatorname{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}] - \mathsf{E}[\log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}|] - p,$$

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, [5].

Apart from the derivation by [15] and [5] in the case of the non-singular Wishart matrix, equivalent identities have also been presented in the case of a singular Wishart matrix (see [12]), in the case of a complex Wishart matrix (see [9]) and in the case of elliptically contoured distributions (see [11], [8] and [1]).

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 the Stein–Haff identity as a formula readily applicable to both estimation problems and derivation of moments. This formula 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. Further, the result is used to derive a condition for orthogonally 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 provides 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 $\mathbf{S}$ be a real, positive-definite and symmetric $p \times p$ 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 $a(\theta)$ and $h(\mathbf{S})$ are known continuous functions, 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:

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

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

where $u(\mathbf{l})$ and $\mathbf{v}(\mathbf{l})$ are known differentiable functions. 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 $\operatorname{vech}(\mathbf{S})$ and one part dependent only on $\mathbf{l}$. Further, let θ be decomposed as $\theta = (\theta_1', \theta_2')'$, where θ_1 is a vector of the same length as the vector $\mathbf{v}(\mathbf{l})$ and θ_2 is a $p(p+1)/2 \times 1$ vector. Now, let $\theta_2 = -\mathbf{D}_p'\mathbf{D}_p \operatorname{vech}(\mathbf{\Omega})$ where $\mathbf{\Omega}$ is a $p \times p$ matrix and $\mathbf{D}_p$ is the duplication matrix [3], such that

$$\theta_2' \operatorname{vech}(\mathbf{S}) = -\operatorname{vech}(\mathbf{\Omega})'\mathbf{D}_p'\mathbf{D}_p \operatorname{vech}(\mathbf{S}) = -\operatorname{vec}(\mathbf{\Omega})' \operatorname{vec}(\mathbf{S}) = -\operatorname{tr}(\mathbf{\Omega}\mathbf{S}). \tag{4}$$

---

[2] A commonly used loss function first considered in [7].

[3] Defined as in e. g. [6], s. t. for a symmetric $p \times p$ matrix $\mathbf{A}$ we have $\mathbf{D}_p \operatorname{vech}(\mathbf{A}) = \operatorname{vec}(\mathbf{A})$.

Equation (1) is then the minimal representation of matrix distributions with density functions of the form

$$f(\mathbf{S}) = a(\theta)u(\mathbf{l})e^{\theta_1{}'\mathbf{v}(\mathbf{l})-\mathrm{tr}(\boldsymbol{\Omega}\mathbf{S})}, \tag{5}$$

such as the matrix-variate gamma distribution, which is further discussed in Section 3. In the rest of this presentation, $\boldsymbol{\Omega}$ will sometimes be denoted $\boldsymbol{\Omega}(\theta)$ in order to emphasize its dependency on the canonical parameter. Finally, note that in the case of a real, symmetric matrix $\mathbf{A}$, the common matrix-to-scalar operators $\mathrm{tr}(\mathbf{A})$ and $|\mathbf{A}|$ depend only on the eigenvalues of $\mathbf{A}$. Hence, conditions (2) and (3) still allow for a wide range of density functions.

As an example of an exponential distribution of the form (1) conforming to (2) and (3), consider the case of a multivariate normal sample presented in Section 1. Thus, $\mathbf{W}$ follows a Wishart distribution with $k \in \mathbb{N}$ degrees of freedom, where $k \geq p$, and positive-definite covariance matrix $\boldsymbol{\Sigma}$. The density of $\mathbf{W}$ can then be expressed in the form (1) with

$$\theta = -\frac{1}{2}\mathbf{D}_p'\mathbf{D}_p \operatorname{vech}\big(\boldsymbol{\Sigma}^{-1}\big), \quad a(\theta) = |\boldsymbol{\Omega}(\theta)|^{k/2}/\Gamma_p(k/2), \quad h(\mathbf{W}) = u(\mathbf{l}) = \prod_{i=1}^{p} l_i^{(k-p-1)/2},$$

where $\Gamma_p(a)$ is the multivariate gamma function[4], $\boldsymbol{\Omega}(\theta) = -\operatorname{vech}^{-1}((\mathbf{D}_p'\mathbf{D}_p)^{-1}\theta) = = \boldsymbol{\Sigma}^{-1}/2$, $l_1, \ldots, l_p$ are the eigenvalues of $\mathbf{W}$ and $\mathbf{t}(\mathbf{W}) = \operatorname{vech}(\mathbf{W})$. Here, $\operatorname{vech}^{-1}(\cdot)$ denotes the inverse of the vech-operator.

Further, as discussed in Section 1, for problems concerning estimation of the parameters of a random matrix distribution, it is often required to compute the expected value of a function of the observed random matrices. For example, such is the case when working with loss and risk functions in the decision-theoretic framework. Furthermore, these 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 $\mathbf{H} \in \mathcal{O}_p$ and $\mathbf{L} = \operatorname{diag}(\mathbf{l})$. From Theorem 3.2.17 in [13], 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

$$\frac{\pi^{p^2/2}}{\Gamma_p(p/2)} \prod_{i<j}(l_i - l_j) \int_{\mathcal{O}_p} f(\mathbf{HLH}')d\mathbf{H}. \tag{6}$$

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 $\mathsf{E}[|g(\mathbf{H}, \mathbf{L})|] < \infty$,

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

[4] Defined as $\Gamma_p(a) = \int_{\mathbf{A}>0} \exp(\mathrm{tr}(-\mathbf{A}))|\mathbf{A}|^{a-(m+1)/2}d\mathbf{A}$.

$$= \frac{\pi^{p^2/2}}{\Gamma_p(p/2)} a(\theta) \int_{\mathcal{L}_p} \prod_{i<j} (l_i - l_j) u(\mathbf{l}) \exp\left(\theta_1' \mathbf{v}(\mathbf{l})\right) \times$$

$$\times \int_{\mathcal{O}_p} g(\mathbf{H}, \mathbf{L}) \exp\left(-\operatorname{tr}\left(\mathbf{\Omega} \mathbf{H} \mathbf{L} \mathbf{H}'\right)\right) d\mathbf{H} d\mathbf{l}, \tag{7}$$

where the second equality comes from inserting (1) and the third equality is due to $\theta_2 = -\mathbf{D}_p' \mathbf{D}_p \operatorname{vech}(\mathbf{\Omega})$ and the identity (4). Now, let $\mathbf{A} = \mathbf{H}' \mathbf{\Omega} \mathbf{H}$ and denote the elements of $\mathbf{A}$ as $a_{ij}(\mathbf{H})$. Then

$$\begin{aligned} \operatorname{tr}(\mathbf{\Omega} \mathbf{H} \mathbf{L} \mathbf{H}') &= \operatorname{tr}(\mathbf{L} \mathbf{A}) = \\ &= \sum_{i=1}^{p} l_i a_{ii}(\mathbf{H}). \end{aligned}$$

As such, (7) 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 the matrix-variate exponential family. The proof is a generalization of the derivations in [14].

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

(i) $\mathsf{E}[|\operatorname{tr}(\mathbf{H} \mathbf{\Phi}(\mathbf{l}) \mathbf{H}' \mathbf{\Omega})|] < \infty$;
(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

$$\mathsf{E}[\operatorname{tr}(\mathbf{H} \mathbf{\Phi}(\mathbf{l}) \mathbf{H}' \mathbf{\Omega})] = \sum_{i=1}^{p} \mathsf{E}\left[\frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + \frac{\partial u(\mathbf{l})}{\partial l_i} \frac{\phi_i(\mathbf{l})}{u(\mathbf{l})} + \theta_1' \phi_i(\mathbf{l}) \frac{\partial \mathbf{v}(\mathbf{l})}{\partial l_i} + \sum_{i<j} \frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j}\right], \tag{8}$$

where $u(\mathbf{l})$ *and* $\mathbf{v}(\mathbf{l})$ *are defined in (2) and (3).*

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

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

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 condition (iii), I can now be written as

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

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

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

and thus

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

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, 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 (8) to obtain

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

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 in the form (1) together with the identity. Section 3.2 applies the identity in order to derive a condition under which estimators dominate the maximum likelihood estimator and provides an example of such an estimator, while Section 3.3 verifies the results through a simulation study.

3.1. **Stein–Haff identity for the matrix-variate gamma distribution.** Let a $p \times p$ matrix $\mathbf{S}$ follow a matrix-variate gamma distribution with shape $\alpha > (p-1)/2$ and symmetric scale matrix $\boldsymbol{\Sigma} > 0$, denoted by $\mathbf{S} \sim MG_p(\alpha, \boldsymbol{\Sigma})$. As such, in accordance with e.g. Definition 3.6.1 in [4], the p.d.f. of $\mathbf{S}$ is

$$f(\mathbf{S}) = \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}|\mathbf{S}|^{\alpha-(p+1)/2}\exp\big(\operatorname{tr}\big(-\boldsymbol{\Sigma}^{-1}\mathbf{S}\big)\big). \tag{9}$$

This matrix distribution belongs to the exponential family and thus the above p.d.f. can be written in the form (1). In the application following in this section, α will be considered to be known, and as such, we set $\theta = -\mathbf{D}_p'\mathbf{D}_p \operatorname{vech}\big(\boldsymbol{\Sigma}^{-1}\big)$, $\mathbf{t}(\mathbf{S}) = \operatorname{vech}(\mathbf{S})$, $a(\theta) = |\boldsymbol{\Omega}(\theta)|^{\alpha}/\Gamma_p(\alpha)$, with $\boldsymbol{\Omega}(\theta) = -\operatorname{vech}^{-1}\big((\mathbf{D}_p'\mathbf{D}_p)^{-1}\theta\big) = \boldsymbol{\Sigma}^{-1}$, and $h(\mathbf{S}) = u(\mathbf{l}) = \prod_{i=1}^{p} l_i^{\alpha-(p+1)/2}$. Thus, this density conforms to conditions (2) and (3).

By applying (8), we can derive the Stein–Haff identity for the matrix-variate gamma distribution as

$$\begin{aligned}\mathsf{E}\big[\operatorname{tr}\big(\mathbf{H}\boldsymbol{\Phi}(\mathbf{l})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\big)\big] &= \mathsf{E}[\operatorname{tr}(\mathbf{H}\boldsymbol{\Phi}(\mathbf{l})\mathbf{H}'\boldsymbol{\Omega})] = \\ &= \sum_{i=1}^{p}\mathsf{E}\left[\frac{\partial\phi_i(\mathbf{l})}{\partial l_i} + \left(\alpha - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j}\right].\end{aligned} \tag{10}$$

3.2. **Estimation of the scale matrix $\boldsymbol{\Sigma}$.** Now, consider a sample of independent matrices $\mathbf{S}_1, \ldots, \mathbf{S}_n$, where $\mathbf{S}_k \sim MG_p(\alpha_k, \boldsymbol{\Sigma}), k = 1, \ldots, n$ and $\alpha_k > (p-1)/2$ are known, while $\boldsymbol{\Sigma} > 0$ is unknown [5]. Further, suppose we are interested in an orthogonally invariant estimator for $\boldsymbol{\Sigma}$, such that the estimator can be written as

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

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

$$L(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p, \tag{11}$$

which has the associated risk function

$$\mathsf{E}[L(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})] = \mathsf{E}[\operatorname{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1})] - \mathsf{E}[\log|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}|] - p. \tag{12}$$

Now, let $\mathbf{V} = \sum_{k=1}^{n}\mathbf{S}_k$. By Lemma A2, $\mathbf{V} \sim MG_p(q, \boldsymbol{\Sigma})$, where $q = \sum_{k=1}^{n}\alpha_k$. In accordance with Lemma A3, the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is $\hat{\boldsymbol{\Sigma}}_{MLE} = \mathbf{V}/q$. Moreover, one can show that for the class of estimators of the form $\hat{\boldsymbol{\Sigma}} = d\mathbf{V}$, the risk function (12) is minimized by $\hat{\boldsymbol{\Sigma}}_{MLE} = \mathbf{V}/q$.

---

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

Equation (10) now allows us to obtain a condition under which estimators dominate the maximum likelihood estimator $\mathbf{V}/q$.

Theorem 2. *Let* $\mathbf{S}_k \sim MG_p(\alpha_k, \boldsymbol{\Sigma}), k = 1, \dots, n$, *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*

$$\hat{\boldsymbol{\Sigma}}_D = \mathbf{H}\boldsymbol{\Phi}(\mathbf{l})\mathbf{H}', \quad \textit{with} \quad \boldsymbol{\Phi}(\mathbf{l}) = \mathrm{diag}(\phi_1(\mathbf{l}), \dots, \phi_p(\mathbf{l})), \quad \phi_i(\mathbf{l}) > 0,\ i = 1, \dots, p,$$

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 (11), if and only if*

$$\sum_{i=1}^p \mathsf{E}\left[\frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + \left(q - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j} - \log\frac{\phi_i(\mathbf{l})}{l_i}\right] \le p + p\log q, \quad (13)$$

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

$$\mathsf{E}[L(\hat{\boldsymbol{\Sigma}}_D, \boldsymbol{\Sigma})] \le \mathsf{E}[L(\hat{\boldsymbol{\Sigma}}_{MLE}, \boldsymbol{\Sigma})], \quad (14)$$

for all values of $\boldsymbol{\Sigma}$, with strict inequality for at least one value of $\boldsymbol{\Sigma}$. By (10) we have, since $\sum_{k=1}^n \mathbf{S}_k \sim MG_p(q, \boldsymbol{\Sigma})$, that

$$\begin{aligned}
\mathsf{E}[L(\hat{\boldsymbol{\Sigma}}_D, \boldsymbol{\Sigma})] &= \mathsf{E}[\mathrm{tr}(\mathbf{H}\boldsymbol{\Phi}(\mathbf{l})\mathbf{H}'\boldsymbol{\Sigma}^{-1})] - \mathsf{E}[\log|\mathbf{H}\boldsymbol{\Phi}(\mathbf{l})\mathbf{H}'\boldsymbol{\Sigma}^{-1}|] - p = \\
&= \sum_{i=1}^p \mathsf{E}\left[\frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + \left(q - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j}\right] - \\
&\quad - \mathsf{E}\left[\log\prod_{i=1}^p \phi_i(\mathbf{l})\right] + \log|\boldsymbol{\Sigma}| - p = \\
&= \sum_{i=1}^p \mathsf{E}\left[\frac{\partial \phi_i(\mathbf{l})}{\partial l_i} + \left(q - \frac{p+1}{2}\right)\frac{\phi_i(\mathbf{l})}{l_i} + \sum_{i<j}\frac{\phi_i(\mathbf{l}) - \phi_j(\mathbf{l})}{l_i - l_j} - \log\phi_i(\mathbf{l})\right] + \\
&\quad + \log|\boldsymbol{\Sigma}| - p. \qquad (15)
\end{aligned}$$

Further, note that by letting $\boldsymbol{\Phi}^*(\mathbf{l}) = \mathrm{diag}(dl_1, \dots, dl_p)$, we have by equation (10) that

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

And thus, by setting $d = 1/q$, we have $\mathsf{E}\left[\mathrm{tr}\left(\frac{\mathbf{V}}{q}\boldsymbol{\Sigma}^{-1}\right)\right] = p$ and can write

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

$$= -\mathsf{E}\left[\log \prod_{i=1}^{p} l_i\right] + \log|\boldsymbol{\Sigma}| + p\log q =$$

$$= -\sum_{i}^{p} \mathsf{E}[\log l_i] + \log|\boldsymbol{\Sigma}| + p\log q. \tag{16}$$

Inserting (15) and (16) into (14) gives the desired result. □

Finally, Theorem 2 can be applied in order to derive an estimator that dominates $\hat{\boldsymbol{\Sigma}}_{MLE}$. Here, we will consider orthogonally invariant estimators where $\boldsymbol{\Phi}(\mathbf{l}) = \operatorname{diag}(\phi_1(\mathbf{l}), \dots, \phi_p(\mathbf{l}))$ is of the form $\phi_i(\mathbf{l}) = d_i l_i, i = 1, \dots, p$, where d_i is a constant.

Corollary 1. *Let* $\mathbf{S}_k \sim MG_p(\alpha_k, \boldsymbol{\Sigma}), k = 1, \dots, n$, *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* $\boldsymbol{\Phi}(\mathbf{l}) = \operatorname{diag}(d_1 l_1, \dots, d_p l_p)$ *and*

$$d_i = \frac{1}{q + (p+1)/2 - i}, \quad i = 1, \dots, p, \tag{17}$$

be an estimator of $\boldsymbol{\Sigma}$. *Then* $\hat{\boldsymbol{\Sigma}}_1$ *dominates* $\hat{\boldsymbol{\Sigma}}_{MLE}$ *with regard to Stein's loss function (11).*

Proof. First, note that by definition $l_1 > \dots > l_p$ and further that $d_1 < \dots < d_p$. By (13) in Theorem 2, we have that if

$$\sum_{i=1}^{p} \mathsf{E}\left[d_i + \left(q - \frac{p+1}{2}\right)d_i + \sum_{i<j} \frac{d_i l_i - d_j l_j}{l_i - l_j} - \log d_i\right] < p + p\log q, \tag{18}$$

$\hat{\boldsymbol{\Sigma}}_1$ will dominate $\hat{\boldsymbol{\Sigma}}_{MLE}$. Now (18) can be written as

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

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

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

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

Let $m_i = \sum_{i<j} \mathsf{E}\left[\frac{l_j}{l_i - l_j}\right](d_i - d_j)$ and note that $m_i < 0, i = 1, \dots, p$ 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

$$p + p\log q > \sum_{i=1}^{p}[1 + m_i + \log(q + (p+1)/2 - i)],$$

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

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

$$q^p > \prod_{i}^{p}(q + (p+1)/2 - i). \tag{19}$$

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

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

If p is even, we can write

$$\prod_i^p (q + (p+1)/2 - i) = \prod_i^{p/2} (q + a_i) \prod_i^{p/2} (q - a_i) < q^p \tag{21}$$

where the inequality is in accordance with (20). On the contrary, if p is odd, we can write

$$\prod_i^p (q + (p+1)/2 - i) = (q) \prod_i^{(p-1)/2} (q + a_i) \prod_i^{(p-1)/2} (q - a_i) \leq q^p, \tag{22}$$

where the inequality again is due to (20). Combining (21) and (22) shows (19), which completes the proof. □

As an example, consider $p = 3$, such that the constants of the estimator $\hat{\mathbf{\Sigma}}_1$ become $d_1 = 1/(q+1), d_2 = 1/q, d_3 = 1/(q-1)$. 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 Corollary 1 is closely related to the estimator derived by [15] and [2] 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 Corollary 1, dominates $\hat{\mathbf{\Sigma}}_{MLE}$ in terms of Stein's loss, we conduct a brief Monte Carlo simulation study. As such, we first define the difference in estimation loss r as

$$r = L(\hat{\mathbf{\Sigma}}_{MLE}, \mathbf{\Sigma}) - L(\hat{\mathbf{\Sigma}}_1, \mathbf{\Sigma}), \tag{23}$$

such that $\mathsf{E}[r] > 0$ 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 dimensions $p = \{2, 4, 10\}$ 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 expected since $\mathsf{E}[L(\hat{\mathbf{\Sigma}}_{MLE}, \mathbf{\Sigma})] > \mathsf{E}[L(\hat{\mathbf{\Sigma}}_1, \mathbf{\Sigma})]$. 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. Finally, in all the considered cases, the loss difference is smaller when the off-diagonal elements of $\mathbf{\Sigma}$ are non-zero compared to the cases when they are zero. This suggests that the risk improvement is greater for the identity matrix, similar to, for example, the conclusions of [2] in the case of a normal population.

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

α/p	$\boldsymbol{\Sigma} = \mathbf{I}_p$			$\boldsymbol{\Sigma} = \mathbf{J}_p$		
	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	$9.9 \cdot 10^{-7}$	$9.5 \cdot 10^{-6}$	$2.0 \cdot 10^{-4}$

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-variate 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 increase with dimension.

Topics for future research include 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 results regarding the matrix-variate gamma distribution needed for the derivations in Section 3.2, most of which are directly related to results on the Wishart distribution.

Lemma A1. *If $\mathbf{S} \sim MG_p(\alpha, \boldsymbol{\Sigma})$, then the characteristic function of $\mathbf{S}$ is*

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

where $\boldsymbol{\Theta}$ is a symmetric $p \times p$ matrix, $\mathbf{T} = (t_{ij}), i, j = 1, \ldots, p$ 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 (9), we have

$$\begin{aligned}
\mathsf{E}[\exp(\operatorname{tr}(i\mathbf{T}\mathbf{S}))] &= \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)} \int_{\mathbf{S}>0} |\mathbf{S}|^{\alpha-(p+1)/2} \exp(\operatorname{tr}(-\boldsymbol{\Sigma}^{-1}\mathbf{S})) \exp(\operatorname{tr}(i\mathbf{T}\mathbf{S})) d\mathbf{S} = \\
&= \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)} \int_{\mathbf{S}>0} |\mathbf{S}|^{\alpha-(p+1)/2} \exp(\operatorname{tr}(i\mathbf{T}\mathbf{S} - \boldsymbol{\Sigma}^{-1}\mathbf{S})) d\mathbf{S} = \\
&= \frac{|\boldsymbol{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)} \int_{\mathbf{S}>0} |\mathbf{S}|^{\alpha-(p+1)/2} \exp(-\operatorname{tr}((\boldsymbol{\Sigma}^{-1} - i\mathbf{T})\mathbf{S})) d\mathbf{S}. \qquad (24)
\end{aligned}$$

By setting $\mathbf{B}^{-1} = \boldsymbol{\Sigma}^{-1} - i\mathbf{T}$, we can write (24) as

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

$$= |\mathbf{\Sigma}|^{-\alpha}|(\mathbf{I}_p - i\mathbf{T}\mathbf{\Sigma})\mathbf{\Sigma}^{-1}|^{-\alpha} =$$
$$= |\mathbf{I}_p - i\mathbf{T}\mathbf{\Sigma}|^{-\alpha},$$

where the first equality is due to $\int_{\mathbf{S}>0} \exp(-\operatorname{tr}(\mathbf{B}^{-1}\mathbf{S}))|\mathbf{S}|^{\alpha-(p+1)/2} d\mathbf{S} = \Gamma_p(\alpha)|\mathbf{B}|^{\alpha}$, which can be seen from the fact that (9) is a p.d.f. and thus integrates to one over $\mathbf{S} > 0$. □

Lemma A2. *Let $\mathbf{S}_1, \ldots, \mathbf{S}_k$ be independent and $\mathbf{S}_k \sim MG_p(\alpha_k, \mathbf{\Sigma}), k = 1, \ldots, n$. Then*

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

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

Proof. Since $\mathbf{S}_1, \ldots, \mathbf{S}_k$ are independent, the characteristic function of $\sum_{k=1}^{n} \mathbf{S}_k$ is the product of the characteristic functions of $\mathbf{S}_1, \ldots, \mathbf{S}_k$. It is as such, in accordance with Lemma A1,

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

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

Lemma A3. *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 estimate 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 the above expression by $\mathbf{\Sigma}$ and equating it to zero, we obtain

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

as desired. □

Acknowledgments

The author is grateful to the Editor, the Associated Editor and the two anonymous referees for valuable comments and suggestions that helped to improve the paper. The author would also like to thank Professor Joanna Tyrcha for insightful feedback.

References

1. T. Bodnar, A. Gupta, *An identity for multivariate elliptically contoured matrix distribution*, Statistics & Probability Letters, **79** (2009), 1327–1330.
2. D. K. Dey, C. Srinivasan, *Estimation of a covariance matrix under Stein's loss*, Ann. Statist., **13** (1985), 1581–1591.
3. T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York, 1967.
4. A. K. Gupta, D. K. Nagar, *Matrix variate distributions*, CRC Press, 2000.

5. L. R. Haff, *An identity for the Wishart distribution with applications*, J. Multivariate Anal., **9** (1979), 531–542.
6. D. Harville, *Matrix algebra from statistician's perspective*, Springer, New York, 1997.
7. W. James, C. Stein, *Estimation with quadratic loss*, *In:* Proc. Fourth Berkeley Symp. Math. Statist. Prob., **1** (1961), 361–380.
8. Y. Konno, *Improving on the sample covariance matrix for a complex elliptically contoured distribution*, Journal of Statistical Planning and Inference, **7** (2007), 2475–2486.
9. Y. Konno, *Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss*, J. Multivariate Anal., **100** (2009), 2237–2253.
10. T. Kubokawa, *A revisit to estimation of the precision matrix of the Wishart distribution*, J. Statist. Res., **39** (2005), 91–114.
11. T. Kubokawa, M. Srivastava, *Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution*, Ann. Statist., **27** (1999), 600–609.
12. T. Kubokawa, M. Srivastava, *Estimation of the precision matrix of a singular Wishart distribution and its application in high dimensional data*, J. Multivariate Anal., **99** (2008), 1906–1928.
13. R. J. Muirhead, *Aspects of multivariate statistical theory*, Wiley, New York, 1982.
14. Y. Sheena, *Unbiased estimator of risk for an orthogonal invariant estimator of a covariance matrix*, Journal of the Japan Statistical Society., **25** (1995), no. 1, 35–48.
15. C. Stein, *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** (1977), 4–65.
16. A. Takemura, *An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population*, Tsukaba J. Math., **8** (1984), 367–376.
17. H. Tsukuma, *Improvement on the best invariant estimators of the normal covariance and precision matrices via a lower triangular subgroup*, J. Japan Statist. Soc., **44** (2014), 195–218.
18. H. Van der Vaart, *On certain characteristics of the distribution of the latent roots of a symmetric random matrix under general conditions*, Ann. Math. Statist., **32** (1961), 864–873.

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

Received 21.06.2018

ТОТОЖНІСТЬ СТЕЙНА – ГАФФА ДЛЯ ЕКСПОНЕНЦІЙНОЇ СІМ'Ї

Г. АЛФЕЛТ

Анотація. У статті встановлено тотожність Стейна – Гаффа для додатно визначених і симетричних випадкових матриць, які належать до експоненційної сім'ї. Потім тотожність застосовано до матричнозначного гамма-розподілу та одержано оцінку, яка домінує оцінку максимальної вірогідності в термінах функції втрат Стейна. Нарешті, проведено моделювання для підтвердження теоретичних результатів.

ТОЖДЕСТВО СТЕЙНА – ХАФФА ДЛЯ ЭКСПОНЕНЦИАЛЬНОГО СЕМЕЙСТВА

Г. АЛФЕЛТ

Аннотация. В статье установлено тождество Стейна – Хаффа для положительно определенных и симметричных случайных матриц, принадлежащих экспоненциальному семейству. Потом тождество применено к матричнозначному гамма-распределению и получена оценка, которая доминирует оценку максимального правдоподобия в терминах функции потерь Стейна. Наконец, проведено моделирование для подтверждения теоретических результатов.