On the exact and approximate distributions of the product of a Wishart matrix with a normal vector

Taras Bodnar, Stepan Mazur, Yarema Okhrin

Angaben zur Veröffentlichung / Publication details:

Bodnar, Taras, Stepan Mazur, and Yarema Okhrin. 2013. "On the exact and approximate distributions of the product of a Wishart matrix with a normal vector." *Journal of Multivariate Analysis* 122: 70–81. https://doi.org/10.1016/j.jmva.2013.07.007.

On the exact and approximate distributions of the product of a Wishart matrix with a normal vector

Taras Bodnar [a,*], Stepan Mazur [b], Yarema Okhrin [c]

[a] *Department of Mathematics, Humboldt University of Berlin, D-10099 Berlin, Germany*
[b] *Department of Statistics, European University Viadrina, D-15207 Frankfurt (Oder), Germany*
[c] *Department of Statistics, University of Augsburg, D-86159 Augsburg, Germany*

A B S T R A C T

In this paper we consider the distribution of the product of a Wishart random matrix and a Gaussian random vector. We derive a stochastic representation for the elements of the product. Using this result, the exact joint density for an arbitrary linear combination of the elements of the product is obtained. Furthermore, the derived stochastic representation allows us to simulate samples of arbitrary size by generating independently distributed chi-squared random variables and standard multivariate normal random vectors for each element of the sample. Additionally to the Monte Carlo approach, we suggest another approximation of the density function, which is based on the Gaussian integral and the third order Taylor expansion. We investigate, with a numerical study, the properties of the suggested approximations. A good performance is documented for both methods.

AMS 2000 subject classifications:
62H10
60E05

Keywords:
Wishart distribution
Multivariate normal distribution
Stochastic representation
Integral approximation

1. Introduction

The basic building block of classical multivariate analysis is the multivariate normal distribution. Its properties are very well understood. Unlike the normal distribution, the theory of the Wishart distribution is less established, but nevertheless, contains numerous important and useful results. Many important distributional properties of Wishart matrices, inverse Wishart matrices, and related statistics are discussed in detail by [9,1,6], and others. The characterization of the Wishart distribution is presented in [16], who extended the results of [7,8], while [13–15] considered the generalization of the Wishart distribution constructed as a quadratic form of a T-distributed random matrix (cf. [5]) whose density function is based on the beta function of the matrix argument (see, e.g., [12]).

The joint distribution functions of the multivariate normal and the Wishart distributions have not been extensively studied in the statistical literature. There are mainly results on the distribution of quadratic forms (see e.g. [17]), despite the fact that the product of the (inverse) Wishart distribution and the normal distribution appears in many applications. A classical example can be found in discriminant analysis, where the elements of the discriminant function are computed as products of the inverse sample covariance matrix multiplied by the sample mean vector. Another important example is taken from portfolio theory in finance, where the weights of the tangency portfolio are estimated by the same product using historical asset returns (see e.g., [4,2]). If frequentist methods are used in analyzing the distributional properties of the discriminant function and/or of the estimated portfolio weights, then we have to deal with the product of the inverse Wishart matrix and a normal vector. This issue was investigated recently by [2]. However, in the Bayesian framework, we obtain the inverse Wishart distribution as the posterior distribution of the sample covariance matrix. This leads to the product of the Wishart matrix and a normal vector. Obtaining the distributional properties of this product statistic is the main goal of the present paper.

* Corresponding author.
E-mail address: bodnar@math.hu-berlin.de (T. Bodnar).

In this paper we investigate the Wishart distribution in combination with a Gaussian vector. In particular, we consider the expressions which depend on **Az**, where **A** is a Wishart matrix and **z** is a Gaussian vector, which are independently distributed. First, we derive a stochastic representation and the exact density function of **LAz** for an arbitrary deterministic matrix **L**. Second, we consider two important special cases. In the first example it is assumed that the covariance matrix $\mathbf{\Sigma}$ is equal to the identity matrix and **L** is a vector in the second case. In both cases the stochastic representations and the exact densities are derived. Moreover, we suggest a further approximation of the density of **LAz** which is based on the Gaussian integral and the third order Taylor series expansion. The performance of the approximate densities is analyzed with an extensive Monte Carlo study.

The rest of the paper is structured as follows. The main results are presented in Section 2, where the stochastic representation for the product **LAz** is derived as Theorem 1. It is applied to derive the density function in Corollary 1. Several important special cases are considered in Corollaries 2 and 3. In Section 2.1 we find an approximation for the density of **LAz** that is based on the third order Taylor series approximation (Theorem 2). The results of numerical studies are given in Section 3, while Section 4 summarizes the paper. The Appendix contains the Proof of Theorem 2.

2. Main results

Let **A** be a k-dimensional Wishart matrix with n degrees of freedom and covariance matrix $\mathbf{\Sigma}$, that is, $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$. We assume that $n > k$, implying that the matrix **A** is non-singular. Furthermore, let $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\mathbf{\Sigma})$, i.e., it follows a k-dimensional multivariate normal distribution. Throughout the paper it is assumed that $\lambda > 0$ and that $\mathbf{\Sigma}$ is positive definite. Let $\overset{d}{=}$ denote equality in distribution and $\mathbf{I}_p$ stand for the identity matrix of order p. In Theorem 1 we present a stochastic representation for p linear combinations of the elements of the random vector **Az**, that is, **LAz**, where **L** is a $p \times k$ constant matrix of rank $p < k$. The distribution of the product is given in terms of a χ^2 random variable and of two standard multivariate normal random vectors which are independently distributed. Stochastic representation is a very powerful tool in multivariate statistics. It plays an important role in the theory of elliptically contoured distributions (cf. [10]) and is widely used in Monte Carlo simulations. In particular, the simulation of the values of the product is considerably simplified if we use the stochastic representation and not the original definition based on the multivariate normal samples.

Theorem 1. *Let* $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$, $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\mathbf{\Sigma})$ *with* $\lambda > 0$ *and* $\mathbf{\Sigma}$ *is positive definite. Assume that* **A** *and* **z** *are independent. Let* **L** *be a* $p \times k$ *constant matrix of rank* $p < k$ *and let* $\mathbf{S}_1 = (\mathbf{L\Sigma L}^T)^{-1/2}\mathbf{L\Sigma}^{1/2}$, $\mathbf{S}_2 = (\mathbf{I}_p - \mathbf{S}_1^T\mathbf{S}_1)^{1/2}$. *Then the stochastic representation of* **LAz** *is given by*

$$\mathbf{LAz} \overset{d}{=} \xi(\mathbf{L\Sigma L}^T)^{1/2}\mathbf{y}_1 + \sqrt{\xi}(\mathbf{L\Sigma L}^T)^{1/2}\left[\sqrt{\mathbf{y}_1^T\mathbf{y}_1 + \eta}\mathbf{I}_p - \frac{\sqrt{\mathbf{y}_1^T\mathbf{y}_1 + \eta} - \sqrt{\eta}}{\mathbf{y}_1^T\mathbf{y}_1}\mathbf{y}_1\mathbf{y}_1^T\right]\mathbf{z}_0, \tag{1}$$

where $\xi \sim \chi^2_n$, $\mathbf{z}_0 \sim N_p(\mathbf{0}, \mathbf{I}_p)$,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mathbf{S}_1\mathbf{\Sigma}^{1/2}\boldsymbol{\mu} \\ \mathbf{S}_2\mathbf{\Sigma}^{1/2}\boldsymbol{\mu} \end{pmatrix}, \lambda\begin{pmatrix} \mathbf{S}_1\mathbf{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_1\mathbf{\Sigma}^2\mathbf{S}_2^T \\ \mathbf{S}_2\mathbf{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_2\mathbf{\Sigma}^2\mathbf{S}_2^T \end{pmatrix}\right) \quad \text{with } \eta = \mathbf{y}_2^T\mathbf{y}_2;$$

ξ *and* $\mathbf{z}_0$ *are independent of* **y**.

Proof. Since **A** and **z** are independently distributed, it follows that the conditional distribution of $\mathbf{LAz}|(\mathbf{z} = \mathbf{z}^*)$ is equal to the distribution of $\mathbf{LAz}^*$. Let $\tilde{\mathbf{L}} = (\mathbf{L}^T, \mathbf{z}^*)^T$. Then $\tilde{\mathbf{A}} = \tilde{\mathbf{L}}\mathbf{A}\tilde{\mathbf{L}}^T = \{\tilde{\mathbf{A}}_{ij}\}_{i,j=1,2}$ can be partitioned with $\tilde{\mathbf{A}}_{11} = \mathbf{LAL}^T$, $\tilde{\mathbf{A}}_{12} = \mathbf{LAz}^*$, $\tilde{\mathbf{A}}_{21} = \mathbf{z}^{*T}\mathbf{AL}^T$ and $\tilde{A}_{22} = \mathbf{z}^{*T}\mathbf{Az}^*$. Similarly, $\mathbf{H} = \tilde{\mathbf{L}}\mathbf{\Sigma}\tilde{\mathbf{L}}^T = \{\mathbf{H}_{ij}\}_{i,j=1,2}$ with $\mathbf{H}_{11} = \mathbf{L\Sigma L}^T$, $\mathbf{H}_{12} = \mathbf{L\Sigma z}^*$, $\mathbf{H}_{21} = \mathbf{z}^{*T}\mathbf{\Sigma L}^T$ and $H_{22} = \mathbf{z}^{*T}\mathbf{\Sigma z}^*$.

Because $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$ and $\text{rank}(\tilde{\mathbf{L}}) = p + 1 \le k$, we get from Theorem 3.2.5 of [18] that $\tilde{\mathbf{A}} \sim W_{p+1}(n, \mathbf{H})$. Furthermore, the application of Theorem 3.2.10 of [18] leads to

$$\tilde{\mathbf{A}}_{12}|\tilde{A}_{22}, \mathbf{z} = \mathbf{z}^* \sim N_p(\mathbf{H}_{12}H_{22}^{-1}\tilde{A}_{22}, \mathbf{H}_{11\cdot 2}\tilde{A}_{22}), \tag{2}$$

where $\mathbf{H}_{11\cdot 2} = \mathbf{H}_{11} - \mathbf{H}_{12}H_{22}^{-1}\mathbf{H}_{21}$.

Let $\xi = \tilde{A}_{22}/H_{22}$. Then

$$\mathbf{LAz} \left| \frac{\mathbf{z}^T\mathbf{Az}}{\mathbf{z\Sigma z}} = \xi, \mathbf{z} \right. \sim N_p(\xi\mathbf{L\Sigma z}, \xi(\mathbf{z}^T\mathbf{\Sigma z L\Sigma L}^T - \mathbf{L\Sigma zz}^T\mathbf{\Sigma L}^T)). \tag{3}$$

Because ξ and **z** are independent (cf., [18, Theorem 3.2.8]) the stochastic representation of **LAz** is

$$\mathbf{LAz} \overset{d}{=} \xi\mathbf{L\Sigma z} + \sqrt{\xi}(\mathbf{z}^T\mathbf{\Sigma z L\Sigma L}^T - \mathbf{L\Sigma zz}^T\mathbf{\Sigma L}^T)^{1/2}\mathbf{z}_0,$$

where $\xi \sim \chi^2_n$, $\mathbf{z}_0 \sim N_p(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\mathbf{\Sigma})$; ξ, $\mathbf{z}_0$, and **z** are mutually independent.

Next, the square root of $(\mathbf{z}^T\boldsymbol{\Sigma}\mathbf{z}\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T - \mathbf{L}\boldsymbol{\Sigma}\mathbf{z}\mathbf{z}^T\boldsymbol{\Sigma}\mathbf{L}^T)$ is calculated. Here, we apply the equality

$$(\mathbf{A} - \mathbf{b}\mathbf{b}^T)^{1/2} = \mathbf{A}^{1/2}(\mathbf{I}_p - c\mathbf{A}^{-1/2}\mathbf{b}\mathbf{b}^T\mathbf{A}^{-1/2})$$

with $c = \frac{1-\sqrt{1-\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}}}{\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}}$, $\mathbf{A} = \mathbf{z}^T\boldsymbol{\Sigma}\mathbf{z}\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T$, and $\mathbf{b} = \mathbf{L}\boldsymbol{\Sigma}\mathbf{z}$.

This leads to

$$\mathbf{L}\mathbf{A}\mathbf{z} \stackrel{d}{=} \xi\mathbf{L}\boldsymbol{\Sigma}\mathbf{z} + \sqrt{\xi}(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{1/2}\left[\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}\mathbf{z}}\mathbf{I}_p - \frac{\sqrt{\mathbf{z}^T\boldsymbol{\Sigma}\mathbf{z}} - \sqrt{\mathbf{z}^T(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^{1/2}\mathbf{Q}_1\boldsymbol{\Sigma}^{1/2})\mathbf{z}}}{\mathbf{z}^T\boldsymbol{\Sigma}^{1/2}\mathbf{Q}_1\boldsymbol{\Sigma}^{1/2}\mathbf{z}}\mathbf{S}_1\boldsymbol{\Sigma}^{1/2}\mathbf{z}\mathbf{z}^T\boldsymbol{\Sigma}^{1/2}\mathbf{S}_1^T\right]\mathbf{z}_0,$$

where $\mathbf{Q}_1 = \mathbf{S}_1^T\mathbf{S}_1$.

Let $\mathbf{t} = \boldsymbol{\Sigma}^{1/2}\mathbf{z} \sim N_k(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma}^2)$ and $\mathbf{S}_1 = (\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1/2}\mathbf{L}\boldsymbol{\Sigma}^{1/2}$. Then

$$\mathbf{L}\mathbf{A}\mathbf{z} \stackrel{d}{=} \xi\mathbf{L}\boldsymbol{\Sigma}^{1/2}\mathbf{t} + \sqrt{\xi}(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{1/2}\left(\sqrt{\mathbf{t}^T\mathbf{t}}\mathbf{I}_p - \frac{\sqrt{\mathbf{t}^T\mathbf{t}} - \sqrt{\mathbf{t}^T(\mathbf{I}_p - \mathbf{Q}_1)\mathbf{t}}}{\mathbf{t}^T\mathbf{Q}_1\mathbf{t}}\mathbf{S}_1\mathbf{t}\mathbf{t}^T\mathbf{S}_1^T\right)\mathbf{z}_0.$$

Since $\mathbf{S}_1$ is a $p \times k$ matrix with $\text{rank}(\mathbf{S}_1) = p$ and $\mathbf{Q}_1$ is a projection matrix with $\text{rank}(\mathbf{Q}_1) = p$, it follows that $\text{rank}(\mathbf{I}_p - \mathbf{Q}_1) = k - p$ (see [11, Theorem 12.3.4]). This implies that we can find a $(k-p) \times k$ matrix $\mathbf{S}_2 = (\mathbf{I}_p - \mathbf{Q}_1)^{1/2}$ such that $\mathbf{S}_2^T\mathbf{S}_2 = (\mathbf{I}_p - \mathbf{Q}_1)$ and $\text{rank}(\mathbf{S}_2) = k - p$. This justifies the transformation $\mathbf{y}_1 = \mathbf{S}_1\mathbf{t}$, $\mathbf{y}_2 = \mathbf{S}_2\mathbf{t}$, where $\mathbf{y}_1 \in \mathbb{R}^p$, $\mathbf{y}_2 \in \mathbb{R}^{k-p}$ and

$$\mathbf{y} = \begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix} \sim N_k\left(\begin{pmatrix}\mathbf{S}_1\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}\\ \mathbf{S}_2\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}\end{pmatrix}, \lambda\begin{pmatrix}\mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_2^T\\ \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_2^T\end{pmatrix}\right).$$

Because $\mathbf{t}^T\mathbf{t} = \mathbf{t}^T\mathbf{Q}_1\mathbf{t} + \mathbf{t}^T\mathbf{S}_2^T\mathbf{S}_2\mathbf{t}$, we get

$$\mathbf{L}\mathbf{A}\mathbf{z} \stackrel{d}{=} \xi(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{1/2}\mathbf{y}_1 + \sqrt{\xi}(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{1/2}\left[\sqrt{\mathbf{y}_1^T\mathbf{y}_1 + \eta}\mathbf{I}_p - \frac{\sqrt{\mathbf{y}_1^T\mathbf{y}_1 + \eta} - \sqrt{\eta}}{\mathbf{y}_1^T\mathbf{y}_1}\mathbf{y}_1\mathbf{y}_1^T\right]\mathbf{z}_0,$$

where $\eta = \mathbf{y}_2^T\mathbf{y}_2$.

To simplify the simulation procedures, Theorem 1 allows us to derive an integral representation of the density function of $\mathbf{L}\mathbf{A}\mathbf{z}$. Let $f_{\chi^2_s}(\cdot)$ denote the density of a χ^2-distribution with s degrees of freedom, while $f_{\chi^2_{s;\tau}}(\cdot)$ stands for the density of the non-central χ^2-distribution with s degrees of freedom and non-centrality parameter τ. Let $f_{N_q(\boldsymbol{\mu},\boldsymbol{\Sigma})}$ denote the density of the q-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let ${}_pF_q(\cdot,\cdot,\cdot)$ be a hypergeometric function of order p and q (see [17, p. 331]). Since $\mathbf{y}_2|\mathbf{y}_1 \sim N_{k-p}(\boldsymbol{\nu}, \lambda\boldsymbol{\Xi})$ with $\boldsymbol{\nu} = \mathbf{S}_2\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu} + (\mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_1^T)(\mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_1^T)^{-1}(\mathbf{y}_1 - \mathbf{S}_1\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu})$ and $\boldsymbol{\Xi} = \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_2^T - \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_1^T(\mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_1^T)^{-1}\mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_2^T$, the density function of η is given by (cf. [17, Theorem 4.2c.1])

$$f_{\eta|\mathbf{y}_1}(y) = \sum_{i=0}^{\infty} c_i \frac{i!}{2\beta\Gamma\left(\frac{k-p}{2}+i\right)}\left(\frac{y}{2\beta}\right)^{\frac{k-p}{2}-1} e^{-y/2\beta} L_i^{\left(\frac{k-p}{2}-1\right)}\left(\frac{y}{2\beta}\right), \quad y > 0,$$

where $\beta > 0$ is an arbitrary constant,

$$L_i^{\left(\frac{k-p}{2}-1\right)}\left(\frac{y}{2\beta}\right) = \frac{\left(-\frac{y}{2\beta}\right)^i}{i!}\, {}_2F_0\left(-i, -\frac{y}{2\beta} - i; -\frac{2\beta}{y}\right),$$

$$c_0 = 1, \qquad c_i = \frac{1}{i}\sum_{r=0}^{i-1} d_{i-r}c_r, \qquad d_j = \sum_{j_1=1}^{k-p}(1 - jb_{j_1}^2)(\alpha_{j_1})^{-j}, \quad j \geq 1.$$

The matrix $\mathbf{V}$ is a $(k-p) \times (k-p)$ orthogonal matrix which diagonalizes $\lambda\boldsymbol{\Xi}$, that is

$$\lambda\mathbf{V}^T\boldsymbol{\Xi}\mathbf{V} = \text{diag}(\alpha_1, \ldots, \alpha_{k-p}), \quad \mathbf{V}\mathbf{V}^T = \mathbf{I}_{k-p},$$

where $\alpha_1, \ldots, \alpha_{k-p}$ are the eigenvalues of $\lambda\boldsymbol{\Xi}$. Let

$$\mathbf{b} = \lambda^{-1/2}\mathbf{V}^T\boldsymbol{\Xi}^{-1/2}\boldsymbol{\nu}.$$

In Corollary 1, we present the exact density of $\mathbf{L}\mathbf{A}\mathbf{z}$.

Corollary 1. *Let* $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$, $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})$ *with* $\lambda > 0$ *and* $\boldsymbol{\Sigma}$ *positive definite. Assume that* $\mathbf{A}$ *and* $\mathbf{z}$ *are independent. Let* $\mathbf{L}$ *be a* $p \times k$ *constant matrix of rank* $p < k$ *and let* $\mathbf{S}_1 = (\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1/2}\mathbf{L}\boldsymbol{\Sigma}^{1/2}$, $\mathbf{S}_2 = (\mathbf{I}_p - \mathbf{S}_1^T\mathbf{S}_1)^{1/2}$. *Then the density of* $\mathbf{LAz}$ *is given by*

$$f_{\mathbf{LAz}}(\mathbf{x}) = \int_0^\infty \int_0^\infty \int_{\mathbb{R}^p} f_{N_p(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})}(\mathbf{x}|\xi = \upsilon, \mathbf{y}_1 = \mathbf{z}_1, \eta = z_2) f_{N_p(\mathbf{S}_1\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}, \lambda\mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_1^T)}(\mathbf{z}_1) f_{\chi^2_n}(\upsilon) f_{\eta|\mathbf{y}_1}(z_2|\mathbf{y}_1 = \mathbf{z}_1)\mathbf{d}\mathbf{z}_1\mathbf{d}z_2\mathbf{d}\upsilon,$$

where $\widetilde{\boldsymbol{\mu}} = \xi(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{1/2}\mathbf{y}_1$, $\widetilde{\boldsymbol{\Sigma}} = \xi(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{1/2}\left[(\mathbf{y}_1^T\mathbf{y}_1 + \eta)\mathbf{I}_p - \mathbf{y}_1\mathbf{y}_1^T\right](\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{1/2}$.

Proof. From the stochastic representation of $\mathbf{LAz}$ (see Theorem 1), we obtain that

$$\mathbf{LAz}|\xi, \mathbf{y}_1, \eta \sim N_p(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}), \tag{4}$$

where $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ are given in Corollary 1; $\xi \sim \chi^2_n$ and

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mathbf{S}_1\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu} \\ \mathbf{S}_2\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu} \end{pmatrix}, \lambda\begin{pmatrix} \mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_2^T \\ \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_2^T \end{pmatrix}\right) \tag{5}$$

with $\eta = \mathbf{y}_2^T\mathbf{y}_2$, and ξ and $\mathbf{y}$ independent. The unconditional density of $\mathbf{LAz}$ is obtained by constructing, first, the joint density of $\mathbf{LAz}$, ξ, $\mathbf{y}_1$, and η using (4) and (5) as well as the independence of ξ and $\mathbf{y}$, and then by integrating out the two random variables ξ and η as well as the random vector $\mathbf{y}_1$.

In the next corollary, we consider the special case when the covariance matrix $\boldsymbol{\Sigma}$ is the $k \times k$ identity matrix, i.e., $\boldsymbol{\Sigma} = \mathbf{I}_k$. In this case the stochastic representation of $\mathbf{LAz}$ as well as its density function simplify significantly.

Corollary 2. *Let* $\mathbf{A} \sim W_k(n, \mathbf{I}_k)$, $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\mathbf{I}_k)$ *with* $\lambda > 0$. *Assume that* $\mathbf{A}$ *and* $\mathbf{z}$ *are independent. Let* $\mathbf{L}$ *be a* $p \times k$ *matrix of constants,* $p < k$, *and let* $\widetilde{\mathbf{S}}_1 = (\mathbf{L}\mathbf{L}^T)^{-1/2}\mathbf{L}$, $\widetilde{\mathbf{S}}_2 = (\mathbf{I}_p - \widetilde{\mathbf{S}}_1^T\widetilde{\mathbf{S}}_1)^{1/2}$. *Then the density of* $\mathbf{LAz}$ *is given by*

$$\begin{aligned} f_{\mathbf{LAz}}(\mathbf{x}) = \frac{1}{\lambda}\int_0^\infty \int_0^\infty \int_{\mathbb{R}^p} & f_{N_p(\widetilde{\boldsymbol{\mu}}_1, \widetilde{\boldsymbol{\Sigma}}_1)}(\mathbf{x}|\xi = \upsilon, \mathbf{y}_1 = \mathbf{z}_1, \eta = z_2) f_{N_p(\widetilde{\mathbf{S}}_1\boldsymbol{\mu}, \lambda\widetilde{\mathbf{S}}_1^T\widetilde{\mathbf{S}}_1)}(\mathbf{z}_1) \\ & \times f_{\chi^2_n}(\upsilon) f_{\chi^2_{k-p;\delta^2}}(\lambda^{-1}z_2|\mathbf{y}_1 = \mathbf{z}_1)\mathbf{d}\mathbf{z}_1\mathbf{d}z_2\mathbf{d}\upsilon, \end{aligned}$$

where $\widetilde{\boldsymbol{\mu}}_1 = \xi(\mathbf{L}\mathbf{L}^T)^{1/2}\mathbf{y}_1$, $\widetilde{\boldsymbol{\Sigma}}_1 = \xi(\mathbf{L}\mathbf{L}^T)^{1/2}\left[(\mathbf{y}_1^T\mathbf{y}_1 + \eta)\mathbf{I}_p - \mathbf{y}_1\mathbf{y}_1^T\right](\mathbf{L}\mathbf{L})^{1/2}$, *and* $\delta^2 = \lambda^{-1}\boldsymbol{\mu}^T\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2\boldsymbol{\mu}$.

Proof. From Theorem 1 we get

$$\mathbf{LAz} \stackrel{d}{=} \xi(\mathbf{L}\mathbf{L}^T)^{1/2}\mathbf{y}_1 + \sqrt{\xi}(\mathbf{L}\mathbf{L}^T)^{1/2}\left[\sqrt{\mathbf{y}_1^T\mathbf{y}_1 + \eta}\,\mathbf{I}_p - \frac{\sqrt{\mathbf{y}_1^T\mathbf{y}_1 + \eta} - \sqrt{\eta}}{\mathbf{y}_1^T\mathbf{y}_1}\mathbf{y}_1\mathbf{y}_1^T\right]\mathbf{z}_0, \tag{6}$$

where $\mathbf{z}_0 \sim N_p(\mathbf{0}, \mathbf{I}_p)$, $\xi \sim \chi^2_n$,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mathbf{S}_1\boldsymbol{\mu} \\ \mathbf{S}_2\boldsymbol{\mu} \end{pmatrix}, \lambda\begin{pmatrix} \mathbf{S}_1\mathbf{S}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2\mathbf{S}_2^T \end{pmatrix}\right) \quad \text{with } \eta = \mathbf{y}_2^T\mathbf{y}_2; \tag{7}$$

$\mathbf{z}_0$, ξ, and $\mathbf{y}$ are independently distributed. Because the covariance matrix of $\mathbf{y}$ is block diagonal, and $\mathbf{y}_1$ and $\mathbf{y}_2$ are jointly multivariate normally distributed, it holds that $\mathbf{y}_1$ and $\mathbf{y}_2$ are independently distributed. Moreover, η is a function of $\mathbf{y}_2$ only and, hence, $\mathbf{y}_1$ and η are independent as well.

Next we prove that $\lambda^{-1}\eta \sim \chi^2_{k-p;\delta^2}$, where δ^2 is as in the statement of Theorem 1. First, we note that the matrix $\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2$ is an idempotent matrix, that is, $\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2 = \widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2$, because $\mathbf{Q}_1$ is an idempotent matrix and $\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2 = \mathbf{I}_k - \mathbf{Q}_1$. Furthermore,

(i) $\text{tr}[(\lambda^{-1}\mathbf{I}_{k-p})(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)] = \text{tr}(\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T) = \text{tr}(\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2) = \text{tr}(\mathbf{I}_k - \mathbf{Q}_1) = k - \text{tr}(\mathbf{Q}_1) = k - \text{tr}(\widetilde{\mathbf{S}}_1\widetilde{\mathbf{S}}_1^T) = k - \text{tr}\underbrace{[(\mathbf{L}\mathbf{L}^T)^{-1/2}\mathbf{L}\mathbf{L}^T(\mathbf{L}\mathbf{L}^T)^{-1/2}]}_{\mathbf{I}_p} = k - p$;

(ii) $(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)(\lambda^{-1}\mathbf{I}_{k-p})(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)(\lambda^{-1}\mathbf{I}_{k-p})(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T) = \lambda\widetilde{\mathbf{S}}_2\underbrace{\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2}_{\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2}\widetilde{\mathbf{S}}_2^T = (\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)(\lambda^{-1}\mathbf{I}_{k-p})(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)$;

(iii) $(\boldsymbol{\mu}^T\widetilde{\mathbf{S}}_2^T)(\lambda^{-1}\mathbf{I}_{k-p})(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)(\lambda^{-1}\mathbf{I}_{k-p})(\widetilde{\mathbf{S}}_2\boldsymbol{\mu}) = \lambda^{-1}\boldsymbol{\mu}^T\underbrace{\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2}_{\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2}\boldsymbol{\mu} = (\boldsymbol{\mu}^T\widetilde{\mathbf{S}}_2^T)(\lambda^{-1}\mathbf{I}_{k-p})(\widetilde{\mathbf{S}}_2\boldsymbol{\mu})$;

(iv) $[\boldsymbol{\mu}^T\widetilde{\mathbf{S}}_2^T][(\lambda^{-1}\mathbf{I}_{k-p})(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)]^2 = \lambda^{-1}\boldsymbol{\mu}^T\underbrace{\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2}_{\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2}\widetilde{\mathbf{S}}_2^T = (\boldsymbol{\mu}^T\widetilde{\mathbf{S}}_2^T)(\lambda^{-1}\mathbf{I}_{k-p})(\lambda\widetilde{\mathbf{S}}_2\widetilde{\mathbf{S}}_2^T)$.

Now, the application of Theorem 5.1.3 of [17] shows that $\lambda^{-1}\eta$ is a $\chi^2_{k-p;\delta^2}$ variable with $\delta^2 = \lambda^{-1}\boldsymbol{\mu}^T\widetilde{\mathbf{S}}_2^T\widetilde{\mathbf{S}}_2\boldsymbol{\mu}$. Using this result and the stochastic representation (6), we get the statement of Corollary 2.

Next, we consider the special case when $p = 1$, that is, when $\mathbf{L} = \mathbf{l}^T$ is a vector. The stochastic representation and the density function of $\mathbf{l}^T\mathbf{Az}$ are obtained by applying Theorem 1 and Corollary 1. Note that the density function is given by three univariate integrals of known univariate density functions.

Corollary 3. *Let* $\mathbf{A} \sim W_k(n, \boldsymbol{\Sigma})$ *and* $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma})$, *with* $\lambda > 0$ *and* $\boldsymbol{\Sigma}$ *positive definite. Assume that* $\mathbf{A}$ *and* $\mathbf{z}$ *are independent. Let* $\mathbf{l}$ *be a k-dimensional vector of constants and let* $\mathbf{S}_1 = (\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l})^{-1/2}\mathbf{l}^T\boldsymbol{\Sigma}^{1/2}$ *and* $\mathbf{S}_2 = (\mathbf{I} - \mathbf{S}_1^T\mathbf{S}_1)^{1/2}$. *Then*

(a) *the stochastic representation of* $\mathbf{l}^T\mathbf{Az}$ *is given by*

$$\mathbf{l}^T\mathbf{Az} \stackrel{d}{=} (\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l})^{1/2}[\xi y_1 - \sqrt{\eta\xi}z_0],$$

where $\xi \sim \chi^2_n$, $z_0 \sim N(0, 1)$,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mathbf{S}_1\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu} \\ \mathbf{S}_2\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu} \end{pmatrix}, \lambda\begin{pmatrix} \mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_2^T \\ \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_1^T & \mathbf{S}_2\boldsymbol{\Sigma}^2\mathbf{S}_2^T \end{pmatrix}\right) \quad \textit{with}$$

$\eta = \mathbf{y}_2^T\mathbf{y}_2$;

ξ *and* z_0 *are independent of* $\mathbf{y}$;

(b) *the density of* $\mathbf{l}^T\mathbf{Az}$ *is given by*

$$f_{\mathbf{l}^T\mathbf{Az}}(x) = \int_0^\infty\int_0^\infty\int_{-\infty}^\infty f_{N((\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l})^{1/2}\xi y_1,(\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l})\xi\eta)}(x|\xi = \upsilon, y_1 = z_1, \eta = z_2)$$
$$\times f_{N(\mathbf{S}_1\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu},\lambda\mathbf{S}_1\boldsymbol{\Sigma}^2\mathbf{S}_1^T)}(z_1)f_{\chi^2_n}(\upsilon)f_\eta(z_2|y_1 = z_1)\mathbf{d}z_1\mathbf{d}z_2\mathbf{d}\upsilon.$$

Proof. The results follow directly from Theorem 1 and Corollary 1 as a special case with $p = 1$ and $\mathbf{L} = \mathbf{l}^T$.

A particularly interesting result is obtained from Corollary 3 if $\boldsymbol{\Sigma} = \mathbf{I}_k$. Then the stochastic representation of $\mathbf{l}^T\mathbf{Az}$ is

$$\mathbf{l}^T\mathbf{Az} \stackrel{d}{=} \xi(\mathbf{l}^T\mathbf{l})^{1/2}y_1 + \sqrt{\xi\eta}(\mathbf{l}^T\mathbf{l})^{1/2}z_0, \tag{8}$$

where $y_1 \sim N((\mathbf{l}^T\mathbf{l})^{-1/2}\mathbf{l}^T\boldsymbol{\mu}, \lambda)$ and $z_0 \sim N(0, 1)$. Moreover, the random variables y_1, z_0, ξ, and η are mutually independently distributed.

From (8) we get

$$\mathbf{l}^T\mathbf{Az}|\xi, \eta \sim N(\xi\mathbf{l}^T\boldsymbol{\mu}, (\lambda\xi^2 + \xi\eta)\mathbf{l}^T\mathbf{l}), \tag{9}$$

and hence the unconditional density of $\mathbf{l}^T\mathbf{Az}$ is

$$f_{\mathbf{l}^T\mathbf{Az}}(x) = \frac{1}{\lambda}\int_0^{+\infty}\int_0^{+\infty} f_{N(\xi\mathbf{l}^T\boldsymbol{\mu},(\lambda\xi^2+\xi\eta)\mathbf{l}^T\mathbf{l})}(x|\xi = \upsilon,\ \eta = z_2)f_{\chi^2_n}(\upsilon)f_{\chi^2_{k-p;\delta^2}}(\lambda^{-1}z_2)\mathbf{d}z_2\mathbf{d}\upsilon \tag{10}$$

which is a two-dimensional integral representation. The density function can be easily computed numerically, since it is given as a double integral of well-known univariate densities. Moreover, (9) shows that in order to generate $\mathbf{l}^T\mathbf{Az}$, we need to simulate three independently distributed univariate random variables. This result speeds up the simulation significantly, especially for larger values of k. Without the representation (9), we would have to simulate $k + \frac{k(k+1)}{2}$ random variables.

2.1. *The approximate density function*

The stochastic representation of $\mathbf{LAz}$ given in Theorem 1 allows us to approximate its distribution by generating independent samples from ξ, $\mathbf{z}_0$, and $\mathbf{y}$ and then computing $\mathbf{LAz}$. The resulting values of $\mathbf{LAz}$ are used to estimate the density function, e.g., by a histogram or a kernel density estimator, or to estimate the cumulative distribution function using its sample counterpart.

In this section we suggest a further approximation of the density of $\mathbf{LAz}$. The approach is based on the theory of the Gaussian integral (cf., [3, p. 4]) and the third order Taylor series approximation. From (3) we obtain the unconditional density of $\mathbf{LAz}$ expressed as

$$f_{\mathbf{LAz}}(\mathbf{x}) = \int_0^{+\infty}\int_{\mathbb{R}^k} f_{\tilde{\mathbf{A}}_{12}}(\mathbf{x}|\mathbf{z} = \mathbf{z}^*, \xi = \upsilon)f_{\mathbf{z}}(\mathbf{z}^*)f_\xi(\upsilon)\mathbf{d}\mathbf{z}^*\mathbf{d}\upsilon.$$

Let

$$I(\mathbf{x},\upsilon) = \int_{\mathbb{R}^k} f_{\tilde{\mathbf{A}}_{12}}(\mathbf{x}|\mathbf{z}=\mathbf{z}^*,\xi=\upsilon) f_{\mathbf{z}}(\mathbf{z}^*)\mathbf{dz}^*$$
$$= \lambda^{-k/2}\frac{\det(\boldsymbol{\Sigma})^{-1/2}}{\upsilon^{p/2}(2\pi)^{(p+k)/2}}\int_{\mathbb{R}^k}\exp\left(-\frac{(\mathbf{z}^*-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}^*-\boldsymbol{\mu})}{2\lambda}\right)$$
$$\times \frac{\det(\mathbf{H}_{11\cdot 2})^{-1/2}}{(\mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{z}^*)^{p/2}}\exp\left(-\frac{(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}\mathbf{z}^*)^T(\mathbf{H}_{11\cdot 2})^{-1}(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}\mathbf{z}^*)}{2\upsilon\mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{z}^*}\right)\mathbf{dz}^*, \tag{11}$$

where $\mathbf{H}_{11\cdot 2}$ was defined in the Proof of Theorem 1 (Section 2).

Let $\mathbf{C}_1 = (\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1}$, $\mathbf{C}_2 = \mathbf{C}_1\mathbf{L}\boldsymbol{\Sigma}$, and $\mathbf{D} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{L}^T\mathbf{C}_2$. Using the identities (cf. [11, Theorem 18.2.8])

$$(\mathbf{H}_{11\cdot 2})^{-1} = (\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1} + \frac{(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1}\mathbf{L}\boldsymbol{\Sigma}\mathbf{z}^*\mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{L}^T(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1}}{\mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{z}^* - \mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{L}^T(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1}\mathbf{L}\boldsymbol{\Sigma}\mathbf{z}^*},$$
$$\det(\mathbf{H}_{11\cdot 2}) = \det(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)\frac{\mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{z}^* - \mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{L}^T(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1}\mathbf{L}\boldsymbol{\Sigma}\mathbf{z}^*}{\mathbf{z}^{*T}\boldsymbol{\Sigma}\mathbf{z}^*},$$

and making the transformation $\mathbf{t} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{z}^*-\boldsymbol{\mu})$ in (11) with the Jacobian $\det(\boldsymbol{\Sigma})^{1/2}$, we get

$$I(\mathbf{x},\upsilon) = \lambda^{-k/2}\frac{\det(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1/2}}{\upsilon^{p/2}(2\pi)^{(p+k)/2}}\int_{\mathbb{R}^k}\exp\left(-\frac{\mathbf{t}^T\mathbf{t}}{2\lambda}\right)[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{1}{2}}[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{\frac{1-p}{2}}$$
$$\times\exp\left(-\frac{(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu}))^T\mathbf{C}_1(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu}))}{2\upsilon(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})}\right.$$
$$\left.-\frac{[(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu}))^T\mathbf{C}_2(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^2}{2\upsilon[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})][(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]}\right)\mathbf{dt}$$
$$= \lambda^{-k/2}\frac{\det(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^T)^{-1/2}}{\upsilon^{p/2}(2\pi)^{(p+k)/2}}\int_{\mathbb{R}^k}\exp\left(-\frac{\mathbf{t}^T\mathbf{t}}{2\lambda}\right)g(\mathbf{x},\upsilon,\mathbf{t})\mathbf{dt},$$

where $g(\mathbf{x},\upsilon,\mathbf{t}) = h(\mathbf{t})\exp[-\frac{1}{2\upsilon}f(\mathbf{x},\upsilon,\mathbf{t})]$ with

$$h(\mathbf{t}) = [(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{1}{2}}[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{p-1}{2}},$$
$$f(\mathbf{x},\upsilon,\mathbf{t}) = f_1(\mathbf{x},\upsilon,\mathbf{t}) + f_2(\mathbf{x},\upsilon,\mathbf{t}),$$

where

$$f_1(\mathbf{x},\upsilon,\mathbf{t}) = \frac{(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu}))^T\mathbf{C}_1(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu}))}{(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})},$$
$$f_2(\mathbf{x},\upsilon,\mathbf{t}) = \frac{[(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu}))^T\mathbf{C}_2(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^2}{[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})][(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]}.$$

We approximate the function $g(\mathbf{x},\upsilon,\mathbf{t})$ by its Taylor series at $\mathbf{t}=\mathbf{0}$. We make the approximation of the third order, which has sufficient precision (see Section 3). Since the odd derivatives are equal to zero, we obtain the approximation

$$g(\mathbf{x},\upsilon,\mathbf{t}) = \omega_0(\mathbf{x},\upsilon) + \frac{1}{2}\mathbf{t}^T\omega_2(\mathbf{x},\upsilon)\mathbf{t} + o(\|\mathbf{t}\|^4), \tag{12}$$

where

$$\omega_0(\mathbf{x},\upsilon) = g(\mathbf{x},\upsilon,\mathbf{0}) = h(\mathbf{0})\exp\left(-\frac{1}{2\upsilon}f(\mathbf{x},\upsilon,\mathbf{0})\right),$$
$$\omega_2(\mathbf{x},\upsilon) = \left.\frac{\partial^2 g(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}}$$
$$= \exp\left(-\frac{1}{2\upsilon}f(\mathbf{x},\upsilon,\mathbf{0})\right)\left(\left.\frac{\partial^2 h(\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}} - \frac{1}{2\upsilon}h(\mathbf{0})\left.\frac{\partial^2 f(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}}\right.$$
$$-\frac{1}{2\upsilon}\left.\frac{\partial f(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}}\left.\frac{\partial h(\mathbf{t})}{\partial\mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}} - \frac{1}{2\upsilon}\left.\frac{\partial h(\mathbf{t})}{\partial\mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}}\left.\frac{\partial f(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}}$$
$$\left.+\frac{1}{4\upsilon^2}h(\mathbf{0})\left.\frac{\partial f(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}}\left.\frac{\partial f(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}}\right).$$

Let $\mathbf{F} = \boldsymbol{\Sigma}^{1/2}\mathbf{D}\boldsymbol{\Sigma}^{1/2}$, $\mathbf{P} = -\upsilon\mathbf{L}\boldsymbol{\Sigma}^{3/2}$, $\mathbf{R} = \boldsymbol{\Sigma}\mathbf{L}^T\mathbf{C}_2$, $\mathbf{C}_3 = \mathbf{P}^T\mathbf{C}_1\mathbf{P}$, $\mathbf{C}_4 = \boldsymbol{\Sigma}^{1/2}\mathbf{R}\boldsymbol{\Sigma}^{1/2}$, $\mathbf{r}_1 = \boldsymbol{\Sigma}^{1/2}\mathbf{D}\boldsymbol{\mu}$, $\mathbf{r}_2 = \boldsymbol{\Sigma}^{3/2}\boldsymbol{\mu}$, $\mathbf{r}_3 = \boldsymbol{\Sigma}^{-1/2}\mathbf{D}\boldsymbol{\mu}$, $\mathbf{r}_4 = \mathbf{P}\mathbf{C}_1\mathbf{v}$, $\mathbf{v} = \mathbf{x} - \upsilon\mathbf{L}\boldsymbol{\Sigma}\boldsymbol{\mu}$, $\mathbf{r}_5 = \boldsymbol{\Sigma}^{1/2}\mathbf{C}_2^T\mathbf{x}$, $\mathbf{r}_6 = \boldsymbol{\Sigma}^{1/2}\mathbf{R}\boldsymbol{\mu}$, $a = \boldsymbol{\mu}^T\mathbf{D}\boldsymbol{\mu}$, $b = \boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}$, $c = \mathbf{v}^T\mathbf{C}_1\mathbf{v}$, $d = \mathbf{v}^T\mathbf{C}_2\boldsymbol{\mu}$. The approximate density of $\mathbf{LAz}$ is given in Theorem 2.

Theorem 2. *Let* $\mathbf{A} \sim W_k(n, \mathbf{\Sigma})$, $n > k$, *and* $\mathbf{z} \sim N_k(\boldsymbol{\mu}, \lambda\mathbf{\Sigma})$ *with* $\lambda > 0$. *Let* $\mathbf{\Sigma}$ *be a positive definite covariance matrix. Suppose* $\mathbf{A}$ *and* $\mathbf{z}$ *are independent, and let* $\mathbf{L}$ *be a* $p \times k$ *constant matrix of rank* $p < k$. *Then the approximate density function of* $\mathbf{LAz}$ *is given by*

$$f_{\mathbf{LAz}}(\mathbf{x}) \approx \frac{\det(\mathbf{L\Sigma L}^T)^{-1/2}}{(2\pi)^{p/2}} \int_0^\infty \upsilon^{-p/2} f_{\chi^2_n}(\upsilon) \quad \omega_0(\mathbf{x}, \upsilon) + \frac{1}{2\lambda}\mathrm{tr}(\omega_2(\mathbf{x}, \upsilon)) \quad d\upsilon, \tag{13}$$

where

$$\omega_0(\mathbf{x}, \upsilon) = a^{-1/2} b^{(1-p)/2} \exp \ -\frac{1}{2\upsilon} \ \frac{c}{b} - \frac{d^2}{ab} \quad ,$$

$$\begin{aligned}
\omega_2(\mathbf{x}, \upsilon) \ = \ & \exp \ -\frac{1}{2\upsilon} \ \frac{c}{b} - \frac{d^2}{ab} \qquad (p-1)a^{-3/2} b^{(1-p)/2-2} \mathbf{r}_2[(p+1)a\mathbf{r}_2 + b\mathbf{r}_1]^T \\
& - a^{-3/2} b^{(1-p)/2} \mathbf{F} + (1-p) a^{-1/2} b^{(1-p)/2-1} \mathbf{\Sigma}^2 + a^{-5/2} b^{(1-p)/2-1} \mathbf{r_1} [3b\mathbf{r}_1 - (1-p)a\mathbf{r}_2]^T \\
& + 2\{b^{-2}(b\mathbf{r}_4 + c\mathbf{r}_2) + (ab)^{-2}[abd(\mathbf{r}_5 - 2\upsilon\mathbf{r}_6) - d^2(a\mathbf{r}_2 + b\mathbf{r}_1)]\} \\
& \times \ \frac{1}{4\upsilon^2} a^{-1/2} b^{(1-p)/2} [2b^{-2}(b\mathbf{r}_4 + c\mathbf{r}_2) + 2ab^{-2}[abd(\mathbf{r}_5 - 2\upsilon\mathbf{r}_6) - d^2(a\mathbf{r}_2 + b\mathbf{r}_1)]]^T \\
& - \frac{1}{2\upsilon} a^{-3/2} b^{(1-p)/2-1} [-b\mathbf{r}_1 + (1-p)a\mathbf{r}_2]^T \quad - \frac{1}{\upsilon} a^{-3/2} b^{(1-p)/2-1} [-b\mathbf{r}_1 + (1-p)a\mathbf{r}_2] \\
& \times [b^{-2}(b\mathbf{r}_4 + c\mathbf{r}_2) + (ab)^{-2}[abd(\mathbf{r}_5 - 2\upsilon\mathbf{r}_6) - d^2(a\mathbf{r}_2 + b\mathbf{r}_1)]]^T \\
& - \frac{1}{\upsilon} a^{-1/2} b^{(1-p)/2} \quad b^{-3}[b^2\mathbf{C}_2 - bc\mathbf{\Sigma}^2 - 2b(\mathbf{r}_2\mathbf{r}_4^T + \mathbf{r}_4\mathbf{r}_2^T) + 4c\mathbf{r}_2\mathbf{r}_2^T] \\
& + (ab)^{-2} \quad \frac{ab}{2}(\mathbf{r}_5 - 2\upsilon\mathbf{r}_6) + d(a\mathbf{r}_2 + b\mathbf{r}_1) \ \ [\mathbf{r}_5 - 2\upsilon\mathbf{r}_6]^T - 2abd\upsilon\mathbf{C}_4 \\
& - 2d[\mathbf{r}_5 - 2\upsilon\mathbf{r}_6][a\mathbf{r}_2 + b\mathbf{r}_1]^T - d^2[2\mathbf{r}_2\mathbf{r}_1^T + a\mathbf{\Sigma}^2 + 2\mathbf{r}_1\mathbf{r}_2^T + b\mathbf{F}] \\
& - 2(ab)^{-3}[a\mathbf{r}_2 + b\mathbf{r}_1][abd(\mathbf{r}_5 - 2\upsilon\mathbf{r}_6) - d^2(a\mathbf{r}_2 + b\mathbf{r}_1)]^T \qquad .
\end{aligned}$$

The Proof of Theorem 2 is given in the Appendix. Note that $f_{\mathbf{LAz}}(\mathbf{x})$ is not necessarily a density function, since the moments of even order greater than or equal to four, which are obviously non-negative, are not used in the Taylor series approximation as given in (13). As a result, it holds that $\int_{\mathbb{R}^p} f_{\mathbf{LAz}}(\mathbf{x})\mathbf{dx} < 1$. For this reason, a standardized version of (13) should be used:

$$\widetilde{f}_{\mathbf{LAz}}(\mathbf{x}) = \frac{f_{\mathbf{LAz}}(\mathbf{x})}{\int_{\mathbb{R}^p} f_{\mathbf{LAz}}(\mathbf{x})\mathbf{dx}}. \tag{14}$$

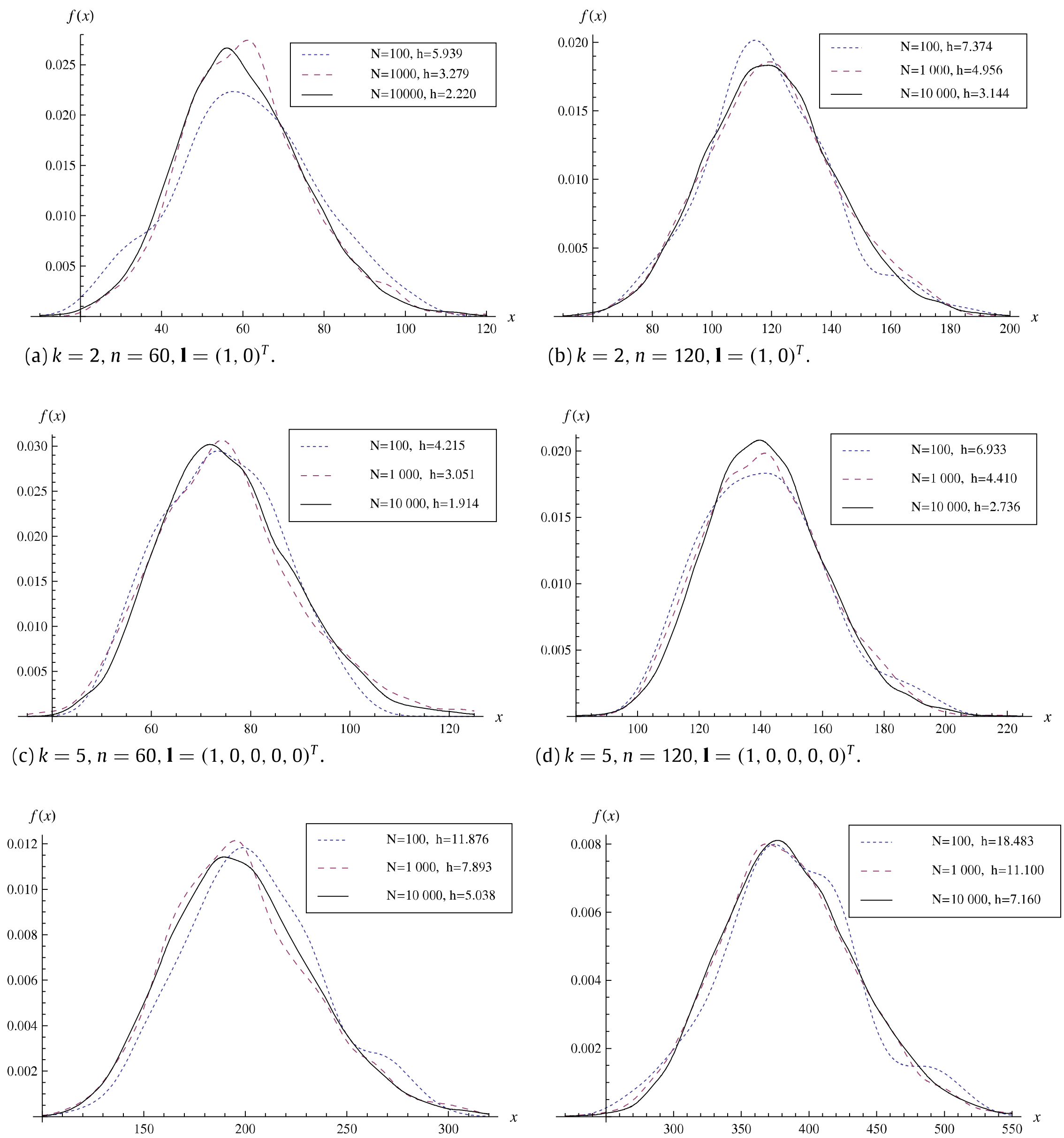
3. Numerical illustration

In this section, the approximate densities provided in Section 2 are compared with the exact one. The comparison is made for $p = 1$, $\mathbf{\Sigma} = \mathbf{I}_k$, and $\boldsymbol{\mu} = (1, \ldots, 1)^T$. In this case the exact density is given as the two-dimensional integral in (10). This integral can be easily evaluated using any mathematical software, e.g., Mathematica. The first approximate density is obtained by generating a sample of size $N \in \{100, 1000, 10\,000\}$ from the product in the stochastic representation of Corollary 3. The sample is used for nonparametric kernel density estimation with Gaussian kernel. We refer to this approach as an approximate density based on the Monte Carlo simulation. The second approximation is calculated following the expression provided in Theorem 2 and its standardized version (14). This approximate density function is called the Taylor series approximation.

In the first numerical study, we investigate the performance of the approximation based on the third order Taylor series approach as given in (13). We compare the square under the density with the target value of unity. This comparison allows us to conclude how much information is ignored by deleting the moments of order higher than two, which are obviously positive. A large deviation from unity is a signal that important information is present in the higher moments and the performance of the approximate density function could be improved by including higher order terms in the Taylor series expansion.

The results are presented in Tables 1 and 2 for several values of n and k with $p = 1$ and $\mathbf{l} = (1, 0, \ldots, 0)^T$ in Table 1 as well as the vector $\mathbf{l}$ having ones in the odd places and zeros otherwise in Table 2. The results of Table 1 demonstrate a very good performance of the approximate density function based on the second order Taylor series approximation. Moreover, we do not observe good performance for the second choice of $\mathbf{l}$ in Table 2 for larger values of k. The values given in the table suggest that the Taylor series of higher order should be considered in this case.

Next, we investigate the impact of the sample size N on the performance of the Monte Carlo approximation. The results are provided in Fig. 1. Here, the kernel density estimators are given for the samples of $N \in \{100, 1000, 10\,000\}$



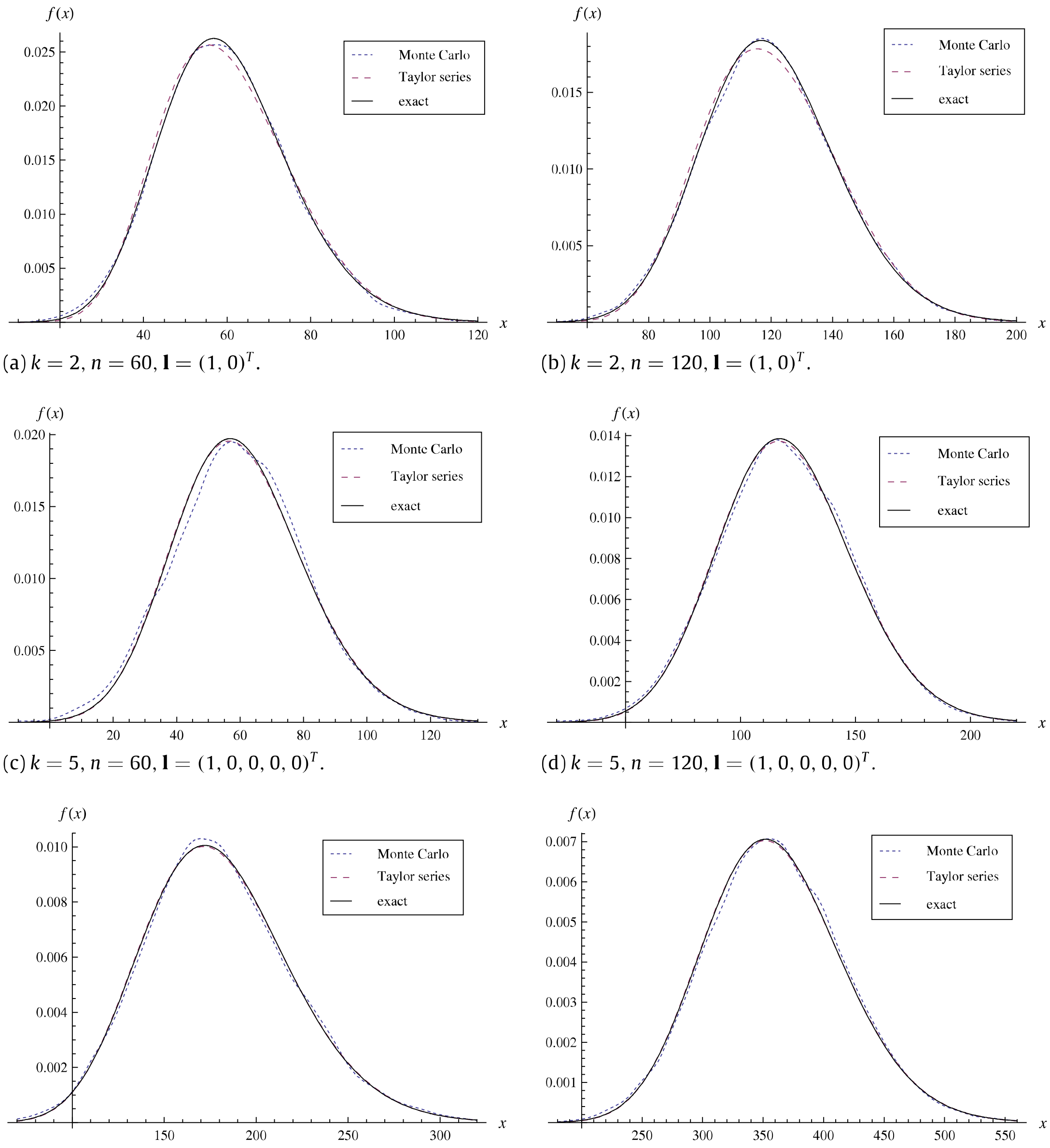
(a) $k = 2$, $n = 60$, $\mathbf{l} = (1, 0)^T$.
(b) $k = 2$, $n = 120$, $\mathbf{l} = (1, 0)^T$.
(c) $k = 5$, $n = 60$, $\mathbf{l} = (1, 0, 0, 0, 0)^T$.
(d) $k = 5$, $n = 120$, $\mathbf{l} = (1, 0, 0, 0, 0)^T$.
(e) $k = 5$, $n = 60$, $\mathbf{l} = (1, 0, 1, 0, 1)^T$.
(f) $k = 5$, $n = 120$, $\mathbf{l} = (1, 0, 1, 0, 1)^T$.

Fig. 1. The kernel density estimator for $k \in \{2, 5\}$, $n \in \{60, 120\}$, $\mathbf{l} \in \{(1, 0)^T, (1, 0, 0, 0, 0)^T, (1, 0, 1, 0, 1)^T\}$, and $N \in \{100, 1000, 10\,000\}$. The values of the bandwidth parameter h are obtained by minimizing the cross validation and they are provided in the legends.

Table 1
Area under the approximate density function based on the Taylor series approximation as given in Theorem 2. $n \in \{30, 60, 120, 250\}$, $k \in \{2, 4, 6, 8, 10\}$, $p = 1$, $\mathbf{l} = (1, 0, \ldots, 0)^T$.

n	k				
	2	4	6	8	10
30	1.000000	1.000000	1.000000	1.000000	1.000000
60	1.000000	0.999913	1.000000	1.000000	1.000000
120	1.000000	1.000000	0.999995	0.999930	1.000000
250	1.000000	0.999999	0.999998	0.999998	1.000000

elements. In all the cases considered, the Gaussian kernel is used with the bandwidth parameter h chosen by cross validation. It is provided in the legend of each plot. The results are given for several values of $k \in \{2, 5\}$, $n \in \{60, 120\}$, and $\mathbf{l} \in \{(1, 0)^T, (1, 0, 0, 0, 0)^T, (1, 0, 1, 0, 1)^T\}$. We observe a considerable improvement in the shapes of the estimated densities if a sample of larger size is used. Moreover, independently of the values of k and n, a good performance is reached by the samples of size $N = 10\,000$, which are used in Fig. 2, where the two approximate densities are compared with the exact one as given in (10).



(a) $k=2, n=60, \mathbf{l}=(1,0)^T$. (b) $k=2, n=120, \mathbf{l}=(1,0)^T$.

(c) $k=5, n=60, \mathbf{l}=(1,0,0,0,0)^T$. (d) $k=5, n=120, \mathbf{l}=(1,0,0,0,0)^T$.

(e) $k=5, n=60, \mathbf{l}=(1,0,1,0,1)^T$. (f) $k=5, n=120, \mathbf{l}=(1,0,1,0,1)^T$.

Fig. 2. The exact density and its two approximations as given in Section 2 ($k \in \{2,5\}$, $n \in \{60,120\}$ and $\mathbf{l} \in \{(1,0)^T, (1,0,0,0,0)^T, (1,0,1,0,1)^T\}$). Samples of $N = 10\,000$ observations are used in the calculation of the Monte Carlo approximate density.

Table 2
Area under the approximate density function based on the Taylor series approximation as given in Theorem 2. $n \in \{30, 60, 120, 250\}$, $k \in \{2, 4, 6, 8, 10\}$, $p = 1$. The vector $\mathbf{l}$ has ones in the odd places and zeros otherwise.

n	k				
	2	4	6	8	10
30	0.999999	0.707107	0.577350	0.500000	0.447211
60	0.999998	0.707107	0.577350	0.500000	0.447214
120	0.999999	0.707102	0.577342	0.499837	0.447266
250	0.999999	0.707105	0.577350	0.499983	0.447214

The results of Fig. 2 demonstrate a good performance of both approximate densities for all considered values of k, n, and $\mathbf{l}$. Here, the exact density is shown by the solid line, the Taylor series approximation by the long-dashed line, and the Monte Carlo approximation by the short-dashed line. We observe that the Monte Carlo approximation performs slightly better for $k = 2$, while for $k = 5$ the approximate density based on the second order Taylor series approximation is better. Nevertheless, note that a better performance of both approximate densities can be achieved if, for the Monte Carlo approximation, samples of larger size are used, and for the Taylor series expansion, a higher order is taken. Hence, we can conclude that both approximate densities perform very well for the values of the parameters considered, but we are not able to recommend a uniformly superior method.

4. Summary

In this paper we analyze the product of a Wishart matrix and a Gaussian vector. We provide a very useful stochastic representation of this product, which is later used to derive the exact density function. The stochastic representation allows us to simulate samples of an arbitrary size which are used in the calculation of the kernel density function, an approximation of the exact one. Moreover, we suggest a further approximation that is based on the third order Taylor series expansion. In the numerical study, we document the good performance of both approximate densities.

The results obtained possess a number of interesting applications. For example, they are useful in the derivation of the posterior distribution of functions of the inverse covariance matrix and mean vector in the Bayesian approach. The suggested approach can be used to derive the posterior of the elements of the discriminant function as well as the posterior for the weights of the tangency portfolio. Since the Wishart distribution is used, it allows us to deal with the case when the sample size is smaller than the dimension of the covariance matrix. In this case similar results can be derived by applying the distributional properties of the singular Wishart distribution instead of the Wishart distribution as discussed by [1]. This generalization has not been treated in the present paper and is left for future research.

Acknowledgments

The authors appreciate the financial support of the German Science Foundation (DFG), projects BO3521/2 and OK103/1, "Wishart Processes in Statistics and Econometrics: Theory and Applications". The authors are grateful to the referees and the editor for their suggestions, which have improved the presentation of the paper.

Appendix

Proof of Theorem 2. Let $\mathbf{F} = \mathbf{\Sigma}^{1/2}\mathbf{D}\mathbf{\Sigma}^{1/2},\ \mathbf{r}_1 = \mathbf{\Sigma}^{1/2}\mathbf{D}\boldsymbol{\mu},\ \mathbf{r}_2 = \mathbf{\Sigma}^{3/2}\boldsymbol{\mu},\ \mathbf{r}_3 = \mathbf{\Sigma}^{-1/2}\mathbf{D}\boldsymbol{\mu},\ a = \boldsymbol{\mu}^T\mathbf{D}\boldsymbol{\mu},\ b = \boldsymbol{\mu}^T\mathbf{\Sigma}\boldsymbol{\mu}$. Now

$$\frac{\partial h(\mathbf{t})}{\partial \mathbf{t}} = h_1(\mathbf{t}) + h_2(\mathbf{t})$$

with

$$\begin{aligned} h_1(\mathbf{t}) &= -[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{3}{2}}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{p-1}{2}}[\mathbf{F}\mathbf{t}+\mathbf{r}_1],\\ h_2(\mathbf{t}) &= (1-p)[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{p-1}{2}-1}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{1}{2}}[\mathbf{\Sigma}^2\mathbf{t}+\mathbf{r}_2]. \end{aligned}$$

Then

$$\begin{aligned} \frac{\partial h_1(\mathbf{t})}{\partial \mathbf{t}^T} &= (\mathbf{F}\mathbf{t}+\mathbf{r}_1)\Big\{3[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{5}{2}}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{\frac{1-p}{2}}\\ &\quad\times[\mathbf{F}\mathbf{t}+\mathbf{r}_1]^T + (p-1)[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{\frac{1-p}{2}-1}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{3}{2}}[\mathbf{\Sigma}^2\mathbf{t}+\mathbf{r}_2]^T\Big\}\\ &\quad-[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{3}{2}}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{p-1}{2}}\mathbf{F} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial h_2(\mathbf{t})}{\partial \mathbf{t}^T} &= (p-1)[\mathbf{\Sigma}^2\mathbf{t}+\mathbf{r}_2]\Big\{(p+1)[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{\frac{1-p}{2}-2}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{1}{2}}[\mathbf{\Sigma}^2\mathbf{t}+\mathbf{r}_2]^T\\ &\quad+[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{\frac{1-p}{2}-1}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{3}{2}}[\mathbf{F}\mathbf{t}+\mathbf{r}_1]^T\Big\}\\ &\quad+(1-p)[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{\Sigma}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{\frac{1-p}{2}-1}[(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\mathbf{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^{-\frac{1}{2}}\mathbf{\Sigma}^2. \end{aligned}$$

Putting these results together, we get

$$\left.\frac{\partial h(\mathbf{t})}{\partial \mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}} = -a^{-3/2}b^{(1-p)/2-1}[b\mathbf{r}_1 + (p-1)a\mathbf{r}_2]$$

and

$$\begin{aligned} \left.\frac{\partial^2 h(\mathbf{t})}{\partial \mathbf{t}\partial \mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}} &= (p-1)a^{-3/2}b^{(1-p)/2-2}\mathbf{r}_2[(p+1)a\mathbf{r}_2 + b\mathbf{r}_1]^T + (1-p)a^{-1/2}b^{(1-p)/2-1}\mathbf{\Sigma}^2\\ &\quad+a^{-5/2}b^{(1-p)/2-1}\mathbf{r}_1[3b\mathbf{r}_1 + (p-1)a\mathbf{r}_2]^T - a^{-3/2}b^{(1-p)/2}\mathbf{F}. \end{aligned}$$

Next, we calculate the partial derivatives of $f_1(\mathbf{x},\upsilon,\mathbf{t})$ and $f_2(\mathbf{x},\upsilon,\mathbf{t})$. Let $\mathbf{P} = -\upsilon\mathbf{L}\boldsymbol{\Sigma}^{3/2}$, $\mathbf{R} = \boldsymbol{\Sigma}\mathbf{L}^T\mathbf{C}_2$, $\mathbf{C}_3 = \mathbf{P}^T\mathbf{C}_1\mathbf{P}$, $\mathbf{C}_4 = \boldsymbol{\Sigma}^{1/2}\mathbf{R}\boldsymbol{\Sigma}^{1/2}$, $\mathbf{v} = \mathbf{x} - \upsilon\mathbf{L}\boldsymbol{\Sigma}\boldsymbol{\mu}$, $\mathbf{r}_4 = \mathbf{P}^T\mathbf{C}_1\mathbf{v}$, $\mathbf{r}_5 = \boldsymbol{\Sigma}^{1/2}\mathbf{C}_2^T\mathbf{x}$, $\mathbf{r}_6 = \boldsymbol{\Sigma}^{1/2}\mathbf{R}\boldsymbol{\mu}$, $c = \mathbf{v}^T\mathbf{C}_1\mathbf{v}$, $d = \mathbf{v}^T\mathbf{C}_2\boldsymbol{\mu}$. Then

$$\frac{\partial f_1(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}} = 2\frac{\overbrace{[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})][\mathbf{C}_3\mathbf{t}+\mathbf{r}_4]}^{\varphi_1(\mathbf{x},\upsilon,\mathbf{t})} - \overbrace{[(\mathbf{P}\mathbf{t}+\mathbf{v})^T\mathbf{C}_1(\mathbf{P}\mathbf{t}+\mathbf{v})][\boldsymbol{\Sigma}^2\mathbf{t}+\mathbf{r}_2]}^{\varphi_2(\mathbf{x},\upsilon,\mathbf{t})}}{\underbrace{[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]^2}_{\varphi_3(\mathbf{t})}}$$

and

$$\frac{\partial^2 f_1(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T} = \frac{1}{\varphi_3^2(\mathbf{t})}\left(\left(\frac{\partial\varphi_1(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T} - \frac{\partial\varphi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T}\right)\varphi_3(\mathbf{t}) - [\varphi_1(\mathbf{x},\upsilon,\mathbf{t}) - \varphi_2(\mathbf{x},\upsilon,\mathbf{t})]\frac{\partial\varphi_3(\mathbf{t})}{\partial\mathbf{t}^T}\right),$$

where

$$\frac{\partial\varphi_1(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T} = [(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]\mathbf{C}_3 + 2[\mathbf{C}_3\mathbf{t}+\mathbf{r}_4][\boldsymbol{\Sigma}^2\mathbf{t}+\mathbf{r}_2]^T,$$
$$\frac{\partial\varphi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T} = [(\mathbf{P}\mathbf{t}+\mathbf{v})^T\mathbf{C}_1(\mathbf{P}\mathbf{t}+\mathbf{v})]\boldsymbol{\Sigma}^2 + 2[\boldsymbol{\Sigma}^2\mathbf{t}+\mathbf{r}_2][\mathbf{C}_3\mathbf{t}+\mathbf{r}_4]^T,$$
$$\frac{\partial\varphi_3(\mathbf{t})}{\partial\mathbf{t}^T} = 4[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})][\boldsymbol{\Sigma}^2\mathbf{t}+\mathbf{r}_2]^T.$$

Thus,

$$\left.\frac{\partial f_1(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}} = \frac{2}{b^2}(b\mathbf{r}_4 + c\mathbf{r}_2),$$
$$\left.\frac{\partial^2 f_1(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}} = \frac{2}{b^3}(b^2\mathbf{C}_2 - bc\boldsymbol{\Sigma}^2 - 2b(\mathbf{r}_2\mathbf{r}_4^T + \mathbf{r}_4\mathbf{r}_2^T) + 4c\mathbf{r}_2\mathbf{r}_2^T).$$

Similarly, we obtain

$$f_2(\mathbf{x},\upsilon,\mathbf{t}) = \frac{[\overbrace{(\mathbf{x}-\upsilon\mathbf{L}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu}))^T\mathbf{C}_2(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})}^{\psi_2(\mathbf{x},\upsilon,\mathbf{t})}]^2}{\underbrace{[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})][(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]}_{\psi_1(\mathbf{t})}} = \frac{\psi_2^2(\mathbf{x},\upsilon,\mathbf{t})}{\psi_1(\mathbf{t})}.$$

Hence,

$$\frac{\partial f_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}} = \frac{2\psi_1(\mathbf{t})\psi_2(\mathbf{x},\upsilon,\mathbf{t})\frac{\partial\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}} - \psi_2^2(\mathbf{x},\upsilon,\mathbf{t})\frac{\partial\psi_1(\mathbf{t})}{\partial\mathbf{t}}}{\psi_1^2(\mathbf{t})}$$

and

$$\begin{aligned}\frac{\partial^2 f_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T} &= 2\psi_1^{-2}(t)\left\{\left[\psi_1(\mathbf{t})\frac{\partial\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}} + \psi_2(\mathbf{x},\upsilon,\mathbf{t})\frac{\partial\psi_1(\mathbf{t})}{\partial\mathbf{t}}\right]\frac{\partial\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T}\right.\\ &\quad\left. + 2\psi_2(\mathbf{x},\upsilon,\mathbf{t})\psi_1(\mathbf{t})\frac{\partial^2\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T} - 2\psi_2(\mathbf{x},\upsilon,\mathbf{t})\frac{\partial\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}}\frac{\partial\psi_1(\mathbf{t})}{\partial\mathbf{t}^T} - \psi_2^2(\mathbf{x},\upsilon,\mathbf{t})\frac{\partial^2\psi_1(\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T}\right\}\\ &\quad - 2\psi_1^{-3}(\mathbf{t})\frac{\partial\psi_1(\mathbf{t})}{\partial\mathbf{t}}\left(2\psi_1(\mathbf{t})\psi_2(\mathbf{x},\upsilon,\mathbf{t})\frac{\partial\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}^T} - \psi_2^2(\mathbf{x},\upsilon,\mathbf{t})\frac{\partial\psi_1(\mathbf{t})}{\partial\mathbf{t}^T}\right),\end{aligned}$$

where

$$\frac{\partial\psi_1(\mathbf{t})}{\partial\mathbf{t}} = 2[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})][\boldsymbol{\Sigma}^2\mathbf{t}+\mathbf{r}_2] + 2[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]\times[\mathbf{F}\mathbf{t}+\mathbf{r}_1],$$
$$\begin{aligned}\frac{\partial^2\psi_1(\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T} &= 4[\boldsymbol{\Sigma}^2\mathbf{t}+\mathbf{r}_2][\mathbf{F}\mathbf{t}+\mathbf{r}_1]^T + 2[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\mathbf{D}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]\boldsymbol{\Sigma}^2\\ &\quad + 4[\mathbf{F}\mathbf{t}+\mathbf{r}_1][\boldsymbol{\Sigma}^2\mathbf{t}+\mathbf{r}_2]^T + 2[(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}+\boldsymbol{\mu})]\mathbf{F},\end{aligned}$$
$$\frac{\partial\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}} = \boldsymbol{\Sigma}^{1/2}\mathbf{C}_2^T\mathbf{x} - 2\upsilon\mathbf{C}_4\mathbf{t} - 2\upsilon\mathbf{r}_6,$$
$$\frac{\partial^2\psi_2(\mathbf{x},\upsilon,\mathbf{t})}{\partial\mathbf{t}\partial\mathbf{t}^T} = -2\upsilon\mathbf{C}_4.$$

Now consider the identities

$$\psi_1(\mathbf{0}) = ab, \qquad \psi_2(\mathbf{x}, \upsilon, \mathbf{0}) = d,$$

$$\left.\frac{\partial \psi_1(\mathbf{t})}{\partial \mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}} = 2a\mathbf{r}_2 + 2b\mathbf{r}_1, \qquad \left.\frac{\partial \psi_2(\mathbf{x}, \upsilon, \mathbf{t})}{\partial \mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}} = \mathbf{r}_5 - 2\upsilon \mathbf{r}_6,$$

$$\left.\frac{\partial^2 \psi_1(\mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}} = 4\mathbf{r}_2\mathbf{r}_1^T + 2a\mathbf{\Sigma}^2 + 4\mathbf{r}_1\mathbf{r}_1^T + 2b\mathbf{F}, \qquad \left.\frac{\partial^2 \psi_2(\mathbf{x}, \upsilon, \mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}} = -2\upsilon \mathbf{C_4}.$$

We then get

$$\left.\frac{\partial f_2(\mathbf{x}, \upsilon, \mathbf{t})}{\partial \mathbf{t}}\right|_{\mathbf{t}=\mathbf{0}} = 2(ab)^{-2}[abd(\mathbf{r}_5 - \upsilon \mathbf{r}_6) - d^2(a\mathbf{r}_2 + b\mathbf{r}_1)]$$

and

$$\left.\frac{\partial^2 f_2(\mathbf{x}, \upsilon, \mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^T}\right|_{\mathbf{t}=\mathbf{0}} = 2(ab)^{-2}\left\{\left[\frac{ab}{2}(\mathbf{r}_5 - 2\upsilon \mathbf{r}_6) + d(a\mathbf{r}_2 + b\mathbf{r}_1)\right]\left(\mathbf{r}_5 - 2\upsilon \mathbf{r}_6\right)^T - 2abd\upsilon \mathbf{C}_4 \right.$$
$$\left. - 2d[\mathbf{r}_5 - 2\upsilon \mathbf{r}_6][a\mathbf{r}_2 + b\mathbf{r}_1]^T - d^2[2\mathbf{r}_2\mathbf{r}_1^T + a\mathbf{\Sigma}^2 + 2\mathbf{r}_1\mathbf{r}_2^T + b\mathbf{F}]\right\}$$
$$- 2(ab)^{-3}[a\mathbf{r}_2 + b\mathbf{r}_1][abd(\mathbf{r}_5 - 2\upsilon \mathbf{r}_6) - d^2(a\mathbf{r}_2 + b\mathbf{r}_1)]^T.$$

Putting all this together, we get the statement of Theorem 2.

References

[1] T. Bodnar, Y. Okhrin, Properties of the partitioned singular, inverse and generalized wishart distributions, Journal of Multivariate Analysis 99 (2008) 2389–2405.
[2] T. Bodnar, Y. Okhrin, On the product of inverse wishart and normal distributions with applications to discriminant analysis and portfolio theory, Scandinavian Journal of Statistics 38 (2011) 311–331.
[3] E. Brézin, Introduction to Statistical Field Theory, Cambridge University Press, 2010.
[4] M. Britten-Jones, The sampling error in estimates of mean–variance efficient portfolio weights, Journal of Finance 54 (1999) 655–671.
[5] J. Dickey, Matrix variate generalization of the multivariate *t*-distribution, Annals of Mathematical Statistics 38 (1967) 511–518.
[6] M. Drton, H. Massam, I. Olkin, Moments of minors of wishart matrices, Annals of Statistics 36 (2008) 2261–2283.
[7] D. Geiger, D. Heckerman, A characterization of the bivariate wishart distribution, Probability and Mathemathical Statistics 18 (1998) 119–131.
[8] D. Geiger, D. Heckerman, Parameter priors for directed acyclic graphical models and the characterization of several probability distributions, Annals of Statistics 30 (2002) 1412–1440.
[9] A. Gupta, D. Nagar, Matrix Variate Distributions, Chapman and Hall/CRC, 2000.
[10] A. Gupta, T. Varga, Elliptically Contoured Models in Statistics, Kluwer, 1993.
[11] D. Harville, Matrix Algebra from Statistician's Perspective, Springer, 1997.
[12] S. Khan, On the generalized beta integral with matrix argument, Pakistan Journal of Statistics 16 (2000) 163–167.
[13] S. Khan, Distribution of sum of squares and products for the generalized multilinear matrix-*t* model, Journal of Multivariate Analysis 83 (2002) 124–140.
[14] S. Khan, Predictive distribution of regression vector and residual sum of squares for normal multiple regression model, Communications in Statistics—Theory and Methods 33 (2004) 2423–2443.
[15] S. Khan, Prediction distribution of future regression and residual sum of squares matrices for multivariate simple regression model with correlated normal responses, Journal of Applied Probability and Statistics 1 (2006) 15–30.
[16] H. Massam, J.W. owski, The matsumoto-yor property and the structure of the wishart distribution, Journal of Multivariate Analysis 97 (2006) 103–123.
[17] A.M. Mathai, S.B. Provost, Quadratic Forms in Random Variables, Marcel Dekker, 1992.
[18] R. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, 1982.