# Kullback–Leibler Divergence Between Multivariate Generalized Gaussian Distributions

Nizar Bouhlel, *Member, IEEE*, and Ali Dziri, *Member, IEEE*

***Abstract*—The Kullback–Leibler divergence (KLD) between two multivariate generalized Gaussian distributions (MGGDs) is a fundamental tool in many signal and image processing applications. Until now, the KLD of MGGDs has no known explicit form, and it is in practice either estimated using expensive Monte-Carlo stochastic integration or approximated. The main contribution of this letter is to present a closed-form expression of the KLD between two zero-mean MGGDs. Depending on the Lauricella series, a simple way of calculating numerically the KLD is exposed. Finally, we show that the approximation of the KLD by Monte-Carlo sampling converges to its theoretical value when the number of samples goes to the infinity.**

***Index Terms*—Multivariate generalized Gaussian distribution, Kullback–Leibler divergence, Lauricella function.**

## I. INTRODUCTION

THE multivariate generalized Gaussian distribution (MGGD) has been recently used in several signal and image processing applications. The univariate GGD has been extensively used to model accurately the magnitudes of wavelet detail coefficients [1], [2], and recently, it has been extended to the MGGD in order to capture the inter-band or/and intra-band dependencies in a wide-sense (multi-scale, multichannel, spatial or color dependencies). This property has been used for several image processing applications such as image de-noising [3]–[5], texture image retrieval and classification [6]–[8]. The Kullback-Leibler divergence (KLD) is a stochastic distance used to measure the similarity between two univariate or multivariate distributions. It is used in texture image retrieval. An analytical expression of the KLD between two univariate zero-mean GGDs was presented by Do and Vetterli in [1]. Recently, Verdoolaege *et al*. have derived an analytical expression of the KLD but limited to the case of bivariate GGD [6]. Bombrun *et al*. [8] have presented the KLD of the bivariate generalized Gamma times an Uniform distribution which is the generalization of the bivariate GGD. However, no closed-form expression existed for the KLD between two zero-mean MGGDs. The lack of an analytical expression for the KLD between two MGGDs leads to the development of several approximation techniques based on the numerical evaluation of the integral of the KLD. The most popular method is Monte-Carlo (MC) estimation technique [9]. It can efficiently estimate the KLD provided that a large number of independent and identically distributed samples is provided. Nevertheless, MC integration is a too slow process to be useful in many applications. Thus, other approximation techniques have been investigated [10], [11]. The main contribution of this letter is to derive a closed-form expression for the KLD between two zero-mean MGGDs in a general case. Besides, a simple way of evaluating the KLD is shown by using the Lauricella series. Finally, a comparison is made to show how the MC sampling method gives approximations close to the KLD theoretical value. Section II describes the MGGD and demonstrates the closed-form expression of the KLD between two zero-mean MGGDs. Section III presents a comparison with MC sampling method. Finally, Section IV concludes this study.

Manuscript received February 22, 2019; revised April 28, 2019; accepted May 1, 2019. Date of publication May 6, 2019; date of current version May 28, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mireille Boutin. *(Corresponding author: Nizar Bouhlel.)*

N. Bouhlel is with the LAB-STICC, ENSTA Bretagne, Brest 29200, France (e-mail: nizar.bouhlel@ensta-bretagne.fr).

A. Dziri is with the Conservatoire National des Arts et Metiers, Paris 75003, France (e-mail: ali.dziri@lecnam.net).

Digital Object Identifier 10.1109/LSP.2019.2915000

## II. MULTIVARIATE GENERALIZED GAUSSIAN DISTRIBUTION AND KULLBACK-LEIBLER DIVERGENCE

Let $\boldsymbol{X}$ be a random vector of $\mathbb{R}^p$ which follows the multivariate exponential power distribution, called also the MGGD, characterized by the following probability density function (pdf) given as follows [12]

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\beta) = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}\Gamma(\frac{p}{2\beta})2^{\frac{p}{2\beta}}}\frac{\beta}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}e^{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^{\beta}} \quad (1)$$

This is for any $\boldsymbol{x} \in \mathbb{R}^p$, where $p$ is the dimensionality of the probability space, $\beta$ is the shape parameter, $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ is the so-called dispersion matrix, which is a $(p \times p)$ positive definite symmetric matrix, and $\Gamma(.)$ is the Gamma function. When $\beta = 1$, the pdf corresponds to the multivariate Gaussian distribution (MGD). In this study, the mean vector $\boldsymbol{\mu}$ is assumed to be zero. In [12], Gómez *et al*. have shown that $\boldsymbol{X}$ is stochastically represented by: $\mathbf{X} = \tau\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{u}$ where $\boldsymbol{u}$ is a random vector uniformly distributed on the unit sphere in the $p$-dimensional Euclidean space, and $\tau$ is scalar positive random variable such that $\tau^{2\beta}$ follows a Gamma distribution with shape parameter 2 and scale parameter $p/(2\beta)$.

Let $\mathbf{X}^1$ and $\mathbf{X}^2$ be two random vectors that follow MGGDs with pdfs $f_{\mathbf{X}^1}(\boldsymbol{x}|\boldsymbol{\Sigma}_1,\beta_1)$ and $f_{\mathbf{X}^2}(\boldsymbol{x}|\boldsymbol{\Sigma}_2,\beta_2)$ given by (1). The KLD between the two zero-mean MGGDs, is given by

$$\begin{aligned}\mathrm{KL}(\mathbf{X}^1||\mathbf{X}^2) &= \int_{\mathbb{R}^p}\ln\left(\frac{f_{\mathbf{X}^1}(\boldsymbol{x}|\boldsymbol{\Sigma}_1,\beta_1)}{f_{\mathbf{X}^2}(\boldsymbol{x}|\boldsymbol{\Sigma}_2,\beta_2)}\right)f_{\mathbf{X}^1}(\boldsymbol{x}|\boldsymbol{\Sigma}_1,\beta_1)\mathrm{d}\boldsymbol{x}\\ &= E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^1}(\mathbf{X})\} - E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^2}(\mathbf{X})\}. \quad (2)\end{aligned}$$

The symmetric KL similarity measure between $\mathbf{X}^1$ and $\mathbf{X}^2$ is $d_{\mathrm{KL}}(\mathbf{X}^1,\mathbf{X}^2) = \mathrm{KL}(\mathbf{X}^1||\mathbf{X}^2) + \mathrm{KL}(\mathbf{X}^2||\mathbf{X}^1)$. In order to compute the KLD, we have to derive the analytical expressions of $E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^1}(\mathbf{X})\}$ and $E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^2}(\mathbf{X})\}$ which depend respectively on $E_{\mathbf{X}^1}\{(\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X})^{\beta_1}\}$ and $E_{\mathbf{X}^1}\{(\mathbf{X}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\}$.

The first expectation is given in a simple form as follows (see Appendix A): $E_{\mathbf{X}^1}\{(\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X})^{\beta_1}\} = p/\beta_1$. The second expectation is given as follows (see Appendix B)

$$E_{\mathbf{X}^1}\{(\mathbf{X}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\} = 2^{\frac{\beta_2}{\beta_1}}\frac{\Gamma\left(\frac{\beta_2}{\beta_1}+\frac{p}{2\beta_1}\right)}{\Gamma(\frac{p}{2\beta_1})}\lambda_p^{\beta_2} \times F_D^{(p-1)}\left(-\beta_2, \underbrace{\frac{1}{2},\dots,\frac{1}{2}}_{p-1}; \frac{p}{2}; 1-\frac{\lambda_{p-1}}{\lambda_p},\dots,1-\frac{\lambda_1}{\lambda_p}\right) \quad (3)$$

where $\lambda_1,\dots,\lambda_p$ are the eigenvalues of the real matrix $\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}$, and $F_D^{(p-1)}(.)$ represents the Lauricella D-hypergeometric function defined for $p-1$ variables (see Appendix C).

Consequently, the closed form expression of the KLD between two zero-mean MGGDs, is given by

$$\mathrm{KL}(\mathbf{X}^1||\mathbf{X}^2) = \ln\left(\frac{\beta_1|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}\Gamma(\frac{p}{2\beta_2})}{\beta_2|\boldsymbol{\Sigma}_2|^{-\frac{1}{2}}\Gamma(\frac{p}{2\beta_1})}\right) + \frac{p}{2}\left(\frac{1}{\beta_2}-\frac{1}{\beta_1}\right) \times \ln 2 - \frac{p}{2\beta_1} + 2^{\frac{\beta_2}{\beta_1}-1}\frac{\Gamma(\frac{\beta_2}{\beta_1}+\frac{p}{2\beta_1})}{\Gamma(\frac{p}{2\beta_1})}\lambda_p^{\beta_2} \times F_D^{(p-1)}\left(-\beta_2, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda_{p-1}}{\lambda_p},\dots,1-\frac{\lambda_1}{\lambda_p}\right). \quad (4)$$

The expression of $\mathrm{KL}(\mathbf{X}^2||\mathbf{X}^1)$ can be easily deduced from $\mathrm{KL}(\mathbf{X}^1||\mathbf{X}^2)$ as follows

$$\mathrm{KL}(\mathbf{X}^2||\mathbf{X}^1) = \ln\left(\frac{\beta_2|\boldsymbol{\Sigma}_2|^{-\frac{1}{2}}\Gamma(\frac{p}{2\beta_1})}{\beta_1|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}\Gamma(\frac{p}{2\beta_2})}\right) + \frac{p}{2}\left(\frac{1}{\beta_1}-\frac{1}{\beta_2}\right) \times \ln 2 - \frac{p}{2\beta_2} + 2^{\frac{\beta_1}{\beta_2}-1}\frac{\Gamma(\frac{\beta_1}{\beta_2}+\frac{p}{2\beta_2})}{\Gamma(\frac{p}{2\beta_2})}\lambda_p^{-\beta_1} \times F_D^{(p-1)}\left(-\beta_1, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda_p}{\lambda_{p-1}},\dots,1-\frac{\lambda_p}{\lambda_1}\right). \quad (5)$$

Lauricella [13] gave several transformation formulas (see Appendix C), of which the relation (41) is used

$$\lambda_p^{-\beta_1}F_D^{(p-1)}\left(-\beta_1, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda_p}{\lambda_{p-1}},\dots,1-\frac{\lambda_p}{\lambda_1}\right) = \lambda_1^{-\beta_1}F_D^{(p-1)}\left(-\beta_1, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda_1}{\lambda_p},\dots,1-\frac{\lambda_1}{\lambda_2}\right). \quad (6)$$

Substituting (6) in (5), and using (4), the symmetric KL similarity measure $d_{\mathrm{KL}}(\mathbf{X}^1,\mathbf{X}^2)$ between $\mathbf{X}^1$ and $\mathbf{X}^2$ becomes

$$d_{\mathrm{KL}}(\mathbf{X}^1,\mathbf{X}^2) = -\frac{p}{2\beta_1} - \frac{p}{2\beta_2} + 2^{\frac{\beta_2}{\beta_1}-1}\frac{\Gamma(\frac{\beta_2}{\beta_1}+\frac{p}{2\beta_1})}{\Gamma(\frac{p}{2\beta_1})} \times \lambda_p^{\beta_2}F_D^{(p-1)}\left(-\beta_2, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda_{p-1}}{\lambda_p},\dots,1-\frac{\lambda_1}{\lambda_p}\right) + 2^{\frac{\beta_1}{\beta_2}-1}\frac{\Gamma(\frac{\beta_1}{\beta_2}+\frac{p}{2\beta_2})}{\Gamma(\frac{p}{2\beta_2})}\lambda_1^{-\beta_1}F_D^{(p-1)}\left(-\beta_1, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda_1}{\lambda_p}, 1-\frac{\lambda_1}{\lambda_{p-1}},\dots,1-\frac{\lambda_1}{\lambda_2}\right). \quad (7)$$

When $\beta_1 = \beta_2 = 1$, the KLD expression corresponds to that of the MGD. The following result can then be deduced

$$E_{\mathbf{X}^1}\{\mathbf{X}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{X}\} = \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) = p\lambda_p F_D^{(p-1)}\left(-1, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda_{p-1}}{\lambda_p},\dots,1-\frac{\lambda_1}{\lambda_p}\right) \quad (8)$$

TABLE I
COMPUTATION OF $F_D^{(p-1)}(.)$ AND ${}_2F_1(.)$ WHEN $p = 3, \beta_2 = 0.47$

| | | $N_{max} = 10$ | | $N_{max} = 20$ | |
|---|---|---|---|---|---|
| $1-\frac{\lambda}{\lambda_p}$ | ${}_2F_1(.)$ | $F_D^{(p-1)}$ | $\lvert\epsilon\rvert$ | $F_D^{(p-1)}$ | $\lvert\epsilon\rvert$ |
| 0.1 | 0.9679 | 0.9679 | 8.4377e-15 | 0.9679 | 1.1102e-16 |
| 0.3 | 0.8990 | 0.8990 | 2.0138e-9 | 0.8990 | 2.9976e-15 |
| 0.5 | 0.8215 | 0.8215 | 8.0969e-7 | 0.8215 | 1.7730e-10 |
| 0.7 | 0.7306 | 0.7306 | 5.5265e-5 | 0.7306 | 3.8219e-7 |
| 0.9 | 0.6124 | 0.6145 | 2.1147e-3 | 0.6126 | 2.2566e-4 |
| 0.99 | 0.5312 | 0.5459 | 1.4686e-2 | 0.5371 | 5.8421e-3 |

TABLE II
PARAMETERS $(\boldsymbol{\Sigma}_1, \beta_1)$ AND $(\boldsymbol{\Sigma}_2, \beta_2)$ USED TO COMPUTE KLD

| | $\beta$ | $\boldsymbol{\Sigma}$ | $\Sigma_{11}$ , $\Sigma_{22}$ , $\Sigma_{33}$ , $\Sigma_{12}$ , $\Sigma_{13}$ , $\Sigma_{23}$ |
|---|---|---|---|
| $\beta_1$ | 0.74 | $\boldsymbol{\Sigma}_1$ | 0.8, 0.2, 0.2, 0.3, 0.2, 0.1 |
| $\beta_2$ | 0.55 | $\boldsymbol{\Sigma}_2$ | 1 , 0.5, 0.7, 0.3, 0.2, 0.1 |

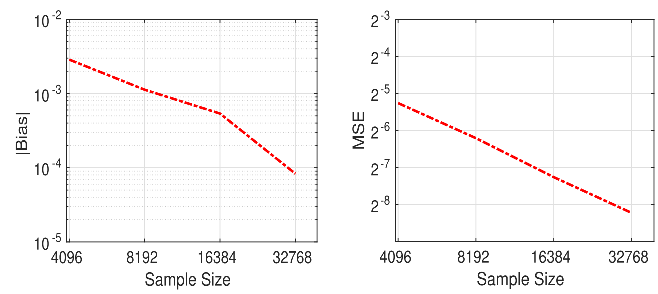


Fig. 1. Bias (left) and MSE (right) of the difference between the approximated and theoretical symmetric KL. The parameter $N_{\max}$ is fixed to 20.

## III. COMPARISON WITH MONTE-CARLO TECHNIQUE

The difficulty in calculating the KLD lies in the calculation of the Lauricella function. To reach this goal, we have used the Lauricella series given by (37), which is easy to compute and implement. The only precaution to be taken into account is the convergence of series. Accordingly, the eigenvalues of $\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}$ are rearranged in a descending order $\lambda_p > \lambda_{p-1} > \cdots > \lambda_1 > 0$ in order to guarantee that $0 \le 1 - \lambda_j/\lambda_p < 1$, $j = 1,\dots,p-1$ and $0 \le 1 - \lambda_1/\lambda_i < 1$, $i = 2,\dots,p$. Finally, in practice, the indices of the series evolve from 0 to $N_{\max}$ instead of infinity. The latter is chosen to ensure a good approximation of the Lauricella function. In the case where $\lambda_1 = \cdots = \lambda_{p-1} = \lambda$, the Lauricella function is equivalent to the Gauss hypergeometric function given as follows $F_D^{(p-1)}(-\beta_2, \frac{1}{2},\dots,\frac{1}{2}; \frac{p}{2}; 1-\frac{\lambda}{\lambda_p},\dots,1-\frac{\lambda}{\lambda_p}) = {}_2F_1(-\beta_2, \frac{p-1}{2}, \frac{p}{2}, 1-\frac{\lambda}{\lambda_p})$. This relation allows us to compare the computational accuracy of the Lauricella series with respect to the Gauss function. The latter, denoted by ${}_2F_1$, is provided in Matlab by the command *hypergeom*. Table I shows the computation of $F_D^{(p-1)}(.)$ and ${}_2F_1(.)$, along with the absolute value of the error $|\epsilon|$, where $p = 3$, $\beta_2 = 0.47$, $N_{\max} = \{10, 20\}$. We note here that most values of $\beta$ encountered in practical applications belong to the interval [0,1] [14]. We clearly see that the error is reasonably low and increases for values of $1 - \lambda/\lambda_p$ close to 1, as expected. In the following, we focus on the MC sampling method to approximate the KLD value. It consists in sampling a large number of samples and using them to compute the summation, rather than the integral. Here, for each sample size, the experiment is repeated for 2000 times. The parameters $(\boldsymbol{\Sigma}_1, \beta_1)$ and $(\boldsymbol{\Sigma}_2, \beta_2)$ are given in Table II. Fig. 1 depicts the absolute value of bias and mean square error

(MSE) of the difference between the symmetric KL approximated value and its theoretical one, given versus the sample sizes. As the sample size increases, the bias and the MSE decrease. Accordingly, the approximated value will be very close to the theoretical KLD when the number of samples is very large. The computation time of the proposed approximation and the classical MC sampling method are recorded using Matlab on a 2.6 GHz processor with 16 GB of memory. For our proposed approximation, the computation time is evaluated to 0.39 and 0.69 seconds for $N_{\max} = \{10, 20\}$, respectively. The value of $N_{\max}$ can be increased to further improve the accuracy, but it will increase the computation time. For the MC sampling method, it is about 7.53 and 15.20 seconds for sample sizes chosen equal to $\{32768, 2 \times 32768\}$, respectively.

## IV. Conclusion

In this letter, we have derived a closed-form expression of the KLD between two zero-mean MGGDs. The similarity measure depends on the Lauricella D-hypergeometric function. We have also proposed a simple approach to compute easily the Lauricella function. For a large number of samples, we have shown by using the MC sampling method that the approximated KLD value converges to its theoretical one.

## Appendix A
## Expression of $E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_1^{-1}\mathbf{X})^{\beta_1}\}$

To demonstrate the result, we exploit the stochastic representation of $\boldsymbol{X}$ given by Gómez *et al.* where $\mathbf{X} = \tau\mathbf{\Sigma}_1^{\frac{1}{2}}\boldsymbol{u}$. Then, $E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_1^{-1}\mathbf{X})^{\beta_1}\} = E_{\mathbf{X}^1}\{(\tau\boldsymbol{u}^T\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Sigma}_1^{-1}\tau\mathbf{\Sigma}_1^{\frac{1}{2}}\boldsymbol{u})^{\beta_1}\} = E_{\mathbf{X}^1}\{(\tau^2\boldsymbol{u}^T\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_1^{\frac{1}{2}}\boldsymbol{u})^{\beta_1}\}$. We recall here that $\mathbf{\Sigma}_1$ is a positive definite symmetric matrix and $\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_1^{\frac{1}{2}} = \mathbf{I}_p$. As a consequence, $E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_1^{-1}\mathbf{X})^{\beta_1}\} = E_{\mathbf{X}^1}\{(\tau^2\boldsymbol{u}^T\boldsymbol{u})^{\beta_1}\} = E_{\mathbf{X}^1}\{\tau^{2\beta_1}(\boldsymbol{u}^T\boldsymbol{u})^{\beta_1}\}$. Since $(\boldsymbol{u}^T\boldsymbol{u})^{\beta_1}$ and $\tau^{2\beta_1}$ are independents, and $E_{\mathbf{X}^1}\{(\boldsymbol{u}^T\boldsymbol{u})^{\beta_1}\} = 1$, the expectation $E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_1^{-1}\mathbf{X})^{\beta_1}\} = E_{\mathbf{X}^1}\{\tau^{2\beta_1}\} = p/\beta_1$.

## Appendix B
## Expression of $E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\}$

The expectation $E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\}$ is computed as follows

$$E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\} = A\int_{\mathbb{R}^p}(\boldsymbol{x}^T\mathbf{\Sigma}_2^{-1}\boldsymbol{x})^{\beta_2}e^{-\frac{1}{2}(\boldsymbol{x}^T\mathbf{\Sigma}_1^{-1}\boldsymbol{x})^{\beta_1}}\mathrm{d}\boldsymbol{x} \quad (9)$$

where $A = \frac{\Gamma(\frac{p}{2})\beta_1}{2^{\frac{p}{2\beta_1}}\pi^{\frac{p}{2}}\Gamma(\frac{p}{2\beta_1})|\mathbf{\Sigma}_1|^{\frac{1}{2}}}$. Consider the transformation $\boldsymbol{y} = \mathbf{\Sigma}_1^{-1/2}\boldsymbol{x}$ where $\boldsymbol{y} = [y_1, y_2, \ldots, y_p]^T$. The Jacobian determinant is given by $\mathrm{d}\boldsymbol{y} = |\mathbf{\Sigma}_1|^{-1/2}\mathrm{d}\boldsymbol{x}$ (Theorem 1.12 in [15]) and the matrix $\mathbf{\Sigma} = \mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1^{\frac{1}{2}}$ is a real symmetric matrix since $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are real symmetric matrices. Then, the expectation is evaluated as follows

$$E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\} = A|\mathbf{\Sigma}_1|^{\frac{1}{2}}\int_{\mathbb{R}^p}(\boldsymbol{y}^T\mathbf{\Sigma}\boldsymbol{y})^{\beta_2}e^{-\frac{1}{2}(\boldsymbol{y}^T\boldsymbol{y})^{\beta_1}}\mathrm{d}\boldsymbol{y} \quad (10)$$

The matrix $\mathbf{\Sigma}$ can be diagonalized by an orthogonal matrix $\mathbf{P}$ where $\mathbf{P}^{-1} = \mathbf{P}^T$ and $\mathbf{\Sigma} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ where $\mathbf{D}$ is a diagonal matrix composed of the eigenvalues of $\mathbf{\Sigma}$. Consider that $\boldsymbol{y}^T\mathbf{\Sigma}\boldsymbol{y} = \mathrm{tr}(\mathbf{\Sigma}\boldsymbol{y}\boldsymbol{y}^T) = \mathrm{tr}(\mathbf{P}\mathbf{D}\mathbf{P}^T\boldsymbol{y}\boldsymbol{y}^T) = \mathrm{tr}(\mathbf{D}\mathbf{P}^T\boldsymbol{y}\boldsymbol{y}^T\mathbf{P})$, and let $\boldsymbol{z} = \mathbf{P}^T\boldsymbol{y}$ with $\boldsymbol{z} = [z_1, z_2, \ldots, z_p]^T$ be a transformation where the Jacobian determinant is given by $\mathrm{d}\boldsymbol{z} = |\mathbf{P}^T|\mathrm{d}\boldsymbol{y} = \mathrm{d}\boldsymbol{y}$. Using the fact that $\mathrm{tr}(\mathbf{D}\mathbf{P}^T\boldsymbol{y}\boldsymbol{y}^T\mathbf{P}) = \mathrm{tr}(\mathbf{D}\boldsymbol{z}\boldsymbol{z}^T) = \boldsymbol{z}^T\mathbf{D}\boldsymbol{z}$ and $\boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{z}^T\mathbf{P}^T\mathbf{P}\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}$, then previous expectation is given as follows

$$\begin{aligned}E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\} &= A|\mathbf{\Sigma}_1|^{\frac{1}{2}}\int_{\mathbb{R}^p}(\boldsymbol{z}^T\mathbf{D}\boldsymbol{z})^{\beta_2}e^{-\frac{1}{2}(\boldsymbol{z}^T\boldsymbol{z})^{\beta_1}}\mathrm{d}\boldsymbol{z}\\ &= A|\mathbf{\Sigma}_1|^{\frac{1}{2}}\int_{\mathbb{R}}..\int_{\mathbb{R}}\left(\sum_{i=1}^p\lambda_i z_i^2\right)^{\beta_2}e^{-\frac{1}{2}\left(\sum_{i=1}^p z_i^2\right)^{\beta_1}}\mathrm{d}z_1\ldots\mathrm{d}z_p\end{aligned} \quad (11)$$

where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$. Let the independent real variables $z_1, \ldots, z_p$ be transformed to the general polar coordinates $r, \theta_1, \ldots, \theta_{p-1}$ as follows, where $r > 0$, $-\pi/2 < \theta_j \leq \pi/2$, $j = 1, \ldots, p-2$, $-\pi < \theta_{p-1} \leq \pi$ [15],

$$z_1 = r\sin\theta_1 \quad (12)$$

$$z_2 = r\cos\theta_1\sin\theta_2 \quad (13)$$

$$z_j = r\cos\theta_1\cos\theta_2\ldots\cos\theta_{j-1}\sin\theta_j, j = 2, 3, \ldots, p-1 \quad (14)$$

$$z_p = r\cos\theta_1\cos\theta_2\ldots\cos\theta_{p-1}. \quad (15)$$

The Jacobian determinant according to theorem (1.24) in [15] is

$$\mathrm{d}z_1\ldots\mathrm{d}z_p = r^{p-1}\prod_{j=1}^{p-1}|\cos\theta_j|^{p-j-1}\mathrm{d}r\mathrm{d}\theta_1\ldots\mathrm{d}\theta_{p-1} \quad (16)$$

It is clear that with the previous transformation, we get $\sum_{i=1}^p z_i^2 = r^2$ and the multiple integral in (11) is then given as follows

$$\begin{aligned}&= A|\mathbf{\Sigma}_1|^{\frac{1}{2}}\int_0^{+\infty}r^{2\beta_2+p-1}e^{-\frac{1}{2}r^{2\beta_1}}\mathrm{d}r\int_{-\pi/2}^{\pi/2}..\int_{-\pi}^{\pi}\\ &\prod_{j=1}^{p-1}|\cos\theta_j|^{p-j-1}(\lambda_1\sin^2\theta_1 + \cdots + \lambda_p\cos^2\theta_1\\ &\ldots\cos^2\theta_{p-1})^{\beta_2}\mathrm{d}\theta_1\ldots\mathrm{d}\theta_{p-1}\end{aligned} \quad (17)$$

The first integral has the following simplified expression

$$\int_0^{+\infty}r^{2\beta_2+p-1}e^{-\frac{1}{2}r^{2\beta_1}}\mathrm{d}r = \frac{2^{\left(\frac{\beta_2}{\beta_1}+\frac{p}{2\beta_1}\right)}}{2\beta_1}\Gamma\left(\frac{\beta_2}{\beta_1}+\frac{p}{2\beta_1}\right) \quad (18)$$

By replacing the expression of $\sin^2\theta_j$ by $1 - \cos^2\theta_j$, for $j = 1, \ldots, p-1$, we have the following expression

$$\begin{aligned}&\lambda_1\sin^2\theta_1 + \cdots + \lambda_p\cos^2\theta_1\ldots\cos^2\theta_{p-1} = \lambda_1 + (\lambda_2 - \lambda_1)\\ &\cos^2\theta_1 + \cdots + (\lambda_p - \lambda_{p-1})\cos^2\theta_1\cos^2\theta_2\ldots\cos^2\theta_{p-1}.\end{aligned} \quad (19)$$

Let $x_i = \cos^2\theta_i$ be a transformation to use where $\mathrm{d}x_i = 2x_i^{1/2}(1 - x_i)^{1/2}\mathrm{d}\theta_i$. Then the multiple integral over all $\theta_j$, $j = 1, \ldots, p-1$ is given as follows

$$\begin{aligned}&\int_{-\pi/2}^{\pi/2}..\int_{-\pi}^{\pi}\left(\lambda_1\sin^2\theta_1 + \cdots + \lambda_p\cos^2\theta_1\ldots\cos^2\theta_{p-1}\right)^{\beta_2}\\ &\times\prod_{j=1}^{p-1}|\cos\theta_j|^{p-j-1}\,\mathrm{d}\theta_1\ldots\mathrm{d}\theta_{p-1} = 2\int_0^1..\int_0^1\prod_{j=1}^{p-1}\\ &x_j^{\frac{p-j}{2}-1}(1 - x_j)^{-\frac{1}{2}}[\lambda_1 + (\lambda_2 - \lambda_1)x_1 + \cdots + (\lambda_p - \lambda_{p-1})\\ &x_1x_2\ldots x_{p-1}]^{\beta_2}\mathrm{d}x_1\ldots\mathrm{d}x_{p-1}\end{aligned} \quad (20)$$

Let $x_i' = 1 - x_i, i = 1, .., p-1$ be transformations to use. Then

$$(\lambda_2 - \lambda_1)x_1 = (\lambda_2 - \lambda_1)(1 - x_1') \quad (21)$$

$$(\lambda_3 - \lambda_2)x_1x_2 = (\lambda_3 - \lambda_2)(1 - x_1')[1 - x_2'] \quad (22)$$

$$(\lambda_p - \lambda_{p-1})\prod_{i=1}^{p-1}x_i = (\lambda_p - \lambda_{p-1})\prod_{i=1}^{p-1}(1 - x_i') \quad (23)$$

Adding equations from (21) to (23), we can state that

$$\lambda_1 + (\lambda_2 - \lambda_1)x_1 + \cdots + (\lambda_p - \lambda_{p-1})x_1 x_2 \ldots x_{p-1}$$
$$= \lambda_p - (\lambda_p - \lambda_1)x'_1 - (\lambda_p - \lambda_2)(1 - x'_1)x'_2 - (\lambda_p - \lambda_3)$$
$$\times (1 - x'_1)(1 - x'_2)x'_3 - \cdots - (\lambda_p - \lambda_{p-1})(1 - x'_1)$$
$$\ldots (1 - x'_{p-2})x'_{p-1} = \mathbf{I} \tag{24}$$

Then, the multiple integral (20) is given as follows

$$= 2\int_0^1 .. \int_0^1 \left(\prod_{j=1}^{p-1}(1 - x'_j)^{\frac{p-j}{2}-1} x_j^{'-\frac{1}{2}}\right) \mathbf{I}^{\beta_2} \mathrm{d}x'_1 \ldots \mathrm{d}x'_{p-1} \tag{25}$$

Let the real variables $x'_1, x'_2, \ldots, x'_{p-1}$ be transformed to the real variables $u_1, u_2, \ldots, u_{p-1}$ as follows,

$$u_1 = x'_1 \tag{26}$$
$$u_2 = (1 - x'_1)x'_2 = (1 - u_1)x'_2 \tag{27}$$
$$u_3 = (1 - x'_1)(1 - x'_2)x'_3 = (1 - u_1 - u_2)x'_3 \tag{28}$$
$$u_{p-1} = \prod_{i=1}^{p-2}(1 - x'_i)x'_{p-1} = \left(1 - \sum_{i=1}^{p-2} u_i\right) x'_{p-1}. \tag{29}$$

The Jacobian determinant is given by

$$\mathrm{d}u_1 \ldots \mathrm{d}u_{p-1} = \prod_{j=1}^{p-1}\left(1 - \sum_{i=1}^{j-1} u_i\right) \mathrm{d}x'_1 \ldots \mathrm{d}x'_{p-1}. \tag{30}$$

As a consequence, the new domain is $\Delta = \{(u_1, u_2, \ldots, u_{p-1}) \in \mathbb{R}^{p-1} | 0 \le u_1 \le 1, 0 \le u_2 \le 1 - u_1, 0 \le u_3 \le 1 - u_1 - u_2, \ldots, \text{and } 0 \le u_{p-1} \le 1 - u_1 - u_2 \cdots - u_{p-2}\}$, and the multiple integral (25) is given as follows

$$= 2\int_\Delta .. \int\left(\prod_{j=1}^{p-1}\left(1 - \frac{u_j}{1 - \sum_{i=1}^{j-1} u_i}\right)^{\frac{p-j}{2}-1}\left(\frac{u_j}{1 - \sum_{i=1}^{j-1} u_i}\right)^{-\frac{1}{2}}\right)$$
$$\times \left(\lambda_p - \sum_{i=1}^{p-1}(\lambda_p - \lambda_i)u_i\right)^{\beta_2} \prod_{j=1}^{p-1}\left(1 - \sum_{i=1}^{j-1} u_i\right)^{-1} \mathrm{d}u_1 \ldots \mathrm{d}u_{p-1} \tag{31}$$
$$= 2\int_\Delta .. \int\left(\prod_{j=1}^{p-1} u_j^{-\frac{1}{2}}\left(1 - \sum_{i=1}^{j} u_i\right)^{\frac{p-j}{2}-1}\left(1 - \sum_{i=1}^{j-1} u_i\right)^{-\frac{p-j}{2}+\frac{1}{2}}\right)$$
$$\times \left(\lambda_p - \sum_{i=1}^{p-1}(\lambda_p - \lambda_i)u_i\right)^{\beta_2} \mathrm{d}u_1 \ldots \mathrm{d}u_{p-1} \tag{32}$$
$$= 2\lambda_p^{\beta_2}\int_\Delta .. \int\left(\prod_{j=1}^{p-1} u_j^{-\frac{1}{2}}\right)\left(1 - \sum_{i=1}^{p-1} u_i\right)^{\frac{p}{2}-\frac{p-1}{2}-1}$$
$$\times \left(1 - \sum_{i=1}^{p-1}\left(1 - \frac{\lambda_i}{\lambda_p}\right)u_i\right)^{\beta_2} \mathrm{d}u_1 \ldots \mathrm{d}u_{p-1}. \tag{33}$$

Using the definition of the Lauricella function given by relation (39), the last expression is given as follows

$$= 2\lambda_p^{\beta_2}\frac{\Gamma(\frac{1}{2})^p}{\Gamma(\frac{p}{2})} F_D^{(p-1)}\left(-\beta_2, \underbrace{\frac{1}{2}, \ldots, \frac{1}{2}}_{p-1}; \frac{p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \ldots, 1 - \frac{\lambda_{p-1}}{\lambda_p}\right) \tag{34}$$

From (11), (18) and (34), we conclude that

$$E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\} = 2^{\frac{\beta_2}{\beta_1}}\frac{\Gamma(\frac{\beta_2}{\beta_1} + \frac{p}{2\beta_1})}{\Gamma(\frac{p}{2\beta_1})}\lambda_p^{\beta_2}$$
$$\times F_D^{(p-1)}\left(-\beta_2, \frac{1}{2}, \ldots, \frac{1}{2}; \frac{p}{2}; 1 - \frac{\lambda_{p-1}}{\lambda_p}, \ldots, 1 - \frac{\lambda_1}{\lambda_p}\right). \tag{35}$$

*1) **Case of** $p = 2$:* The expectation is given by

$$E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\} = \lambda_2^{\beta_2} 2^{\frac{\beta_2}{\beta_1}}\frac{\Gamma(\frac{1+\beta_2}{\beta_1})}{\Gamma(\frac{1}{\beta_1})}\, {}_2F_1\left(-\beta_2, \frac{1}{2}; 1; 1 - \frac{\lambda_1}{\lambda_2}\right)$$

where ${}_2F_1$ is the Gauss's hypergeometric function.

*2) **Case of** $p = 3$:* The expectation is then given by

$$E_{\mathbf{X}^1}\{(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X})^{\beta_2}\} = \lambda_3^{\beta_2} 2^{\frac{\beta_2}{\beta_1}}\frac{\Gamma\left(\frac{\beta_2}{\beta_1} + \frac{3}{2\beta_1}\right)}{\Gamma\left(\frac{3}{2\beta_1}\right)}$$
$$\times {}_2F_1\left(-\beta_2, \frac{1}{2}, \frac{1}{2}; \frac{3}{2}; 1 - \frac{\lambda_1}{\lambda_3}, 1 - \frac{\lambda_2}{\lambda_3}\right). \tag{36}$$

where $F_1$ is the Appell's hypergeometric function.

## Appendix C
## Lauricella Function

In 1893, G. Lauricella [13] investigated the properties of series $F_D^{(n)}$ of $n$ variables. When $n = 2$, the function coincides with Appell's, $F_1$. When $n = 1$, it coincides with Gauss' ${}_2F_1$. The Lauricella series $F_D^{(n)}$ is given as follows

$$F_D^{(n)}(a, b_1, \ldots, b_n; x_1, \ldots, x_n) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty}$$
$$\frac{(a)_{m_1+\cdots+m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c)_{m_1+\cdots+m_n}}\frac{x_1^{m_1}}{m_1!}\cdots\frac{x_n^{m_n}}{m_n!} \tag{37}$$

where $|x_1|, \ldots, |x_n| < 1$. The Pochhammer symbol $(q)_i$ indicates the i-th rising factorial of $q$, i.e.

$$(q)_i = \frac{\Gamma(q+i)}{\Gamma(q)} \quad \text{if} \quad i = 1, 2, \ldots \tag{38}$$

If $i = 0$, $(q)_i = 1$. The function $F_D^{(n)}(.)$ can be expressed in terms of multiple integrals as follows [16]

$$F_D^{(n)}(a, b_1, \ldots, b_n; c; x_1, \ldots, x_n) = \frac{\Gamma(c)}{\Gamma(c - \sum_{i=1}^n b_i)\prod_{i=1}^n \Gamma(b_i)}$$
$$\times \int_\Omega \cdots \int \prod_{i=1}^n u_i^{b_i - 1}\left(1 - \sum_{i=1}^n u_i\right)^{c - \sum_{i=1}^n b_i - 1}$$
$$\times \left(1 - \sum_{i=1}^n x_i u_i\right)^{-a}\prod_{i=1}^n \mathrm{d}u_i \tag{39}$$

where $\Omega = \{(u_1, u_2, \ldots, u_n) | 0 \le u_i \le 1, i = 1, \ldots, n$, and $0 \le u_1 + u_2 + \cdots + u_n \le 1\}$, $\mathrm{Re}(b_i) > 0$ for $i = 1, \ldots, n$ and $\mathrm{Re}(c - b_1 - \cdots - b_n) > 0$. Lauricella has given the following two transformation formulas [17]:

$$F_D^{(n)}(a, b_1, \ldots, b_n; c; x_1, \ldots, x_n) = \prod_{i=1}^n (1 - x_i)^{-b_i}$$
$$\times F_D^{(n)}\left(c - a, b_1, \ldots, b_n; c; \frac{x_1}{x_1 - 1}, \ldots, \frac{x_n}{x_n - 1}\right) \tag{40}$$
$$= (1 - x_1)^{-a} F_D^{(n)}\left(a, c - \sum_{i=1}^n b_i, b_2, \ldots, b_n; c; \frac{x_1}{x_1 - 1},\right.$$
$$\left.\frac{x_1 - x_2}{x_1 - 1}, \ldots, \frac{x_1 - x_n}{x_1 - 1}\right) \tag{41}$$

## REFERENCES

[1] M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized Gaussian density and Kullback–Leibler distance,” *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[2] S. Cui and M. Datcu, “Statistical wavelet subband modeling for multitemporal SAR change detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1095–1109, Aug. 2012.

[3] L. Sendur and I. W. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.

[4] C. Dongwook and D. B. Tien, “Multivariate statistical modeling for image denoising using wavelet transforms,” *Signal Process., Image Commun.*, vol. 20, no. 1, pp. 77–89, 2005.

[5] C. Dongwook, D. B. Tien, and C. Guangyi, “Image denoising based on wavelet shrinkage using neighbor and level dependency,” *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 7, no. 3, pp. 299–311, 2009.

[6] G. Verdoolaege, Y. Rosseel, M. Lambrechts, and P. Scheunders, “Wavelet-based colour texture retrieval using the Kullback–Leibler divergence between bivariate generalized Gaussian models,” in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 265–268.

[7] G. Verdoolaege and P. Scheunders, “Geodesics on the manifold of multivariate generalized gaussian distributions with an application to multicomponent texture discrimination,” *Int. J. Comput. Vis.*, vol. 95, no. 3, pp. 265–286, May 2011.

[8] L. Bombrun and Y. Berthoumieu, “Multivariate texture retrieval using the Kullback–Leibler divergence between bivariate generalized Gamma times an Uniform distribution,” in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 2413–2416.

[9] S. Cui and M. Datcu, “Comparison of Kullback–Leibler divergence approximation methods between Gaussian mixture models for satellite image retrieval,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 3719–3722.

[10] F. Nielsen and K. Sun, “Guaranteed bounds on the Kullback-Leibler divergence of univariate mixtures,” *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1543–1546, Nov. 2016.

[11] E. Ghodhbani, M. Kaaniche, and A. Benazza-Benyahia, “Close approximation of Kullback–Leibler divergence for sparse source retrieval,” *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 745–749, May 2019.

[12] E. Gómez, M. Gomez-Viilegas, and J. Marin, “A multivariate generalization of the power exponential family of distributions,” *Commun. Statist. Theory Methods*, vol. 27, no. 3, pp. 589–600, 1998.

[13] G. Lauricella, “Sulle funzioni ipergeometriche a piu variabili,” *Rendiconti del Circolo Matematico di Palermo*, vol. 7, no. 1, pp. 111–158, Dec. 1893.

[14] F. Pascal, L. Bombrun, J. Tourneret, and Y. Berthoumieu, “Parameter estimation for multivariate generalized Gaussian distributions,” *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5960–5971, Dec. 2013.

[15] A. M. Mathai, *Jacobians of Matrix Transformations and Functions of Matrix Argument*. Singapore: World Scientific, 1997.

[16] A. Hattori and T. Kimura, “On the Euler integral representations of hypergeometric functions in several variables,” *J. Math. Soc. Japan*, vol. 26, no. 1, pp. 1–16, 1974.

[17] H. Exton, *Multiple Hypergeometric Functions and Applications*. Hoboken, NJ, USA: Wiley, 1976.