Exact Rényi and Kullback-Leibler Divergences Between Multivariate t-Distributions

Nizar Bouhlel, David Rousseau

Exact Rényi and Kullback-Leibler Divergences Between Multivariate t-Distributions

N. Bouhlel, *Member, IEEE* and D. Rousseau, *Member, IEEE*

***Abstract*—In this letter, we propose a closed-form expression of the Rényi divergence (RD) of order β between two zero-mean real multivariate t-distributions (MTDs). Such distribution has been deployed in several signal and image processing applications where heavy-tailed distribution is well-suited. Based on the computation of the multiple integral involved in the RD, the expression of the divergence is provided without resorting to the conventional time-consuming Monte Carlo (MC) integration technique. In addition, the Kullback-Leibler divergence (KLD) is deduced from RD. Finally, a comparison is made between the MC method and the numerical value of the RD expression to show how the former gives close approximations to the latter.**

***Index Terms*—Multivariate t-distribution, Rényi divergence of order β, Kullback-Leibler divergence, Lauricella series.**

I. Introduction

In the literature, numerous divergences have been proposed to measure the distance between probability distributions, such as Kullback-Leibler, Rényi [1], Bhattacharyya [2] and Hellinger [3], etc. The statistical distance has been extensively used in various applications such as image classification and texture retrieval [4], [5], change detection [6], speaker recognition and classification [7], [8], model selection [9], parameter estimation [10], etc. Explicit expressions of some of these divergences have been developed for most univariate distributions such as the Gamma, the Cauchy [11], the generalized Gamma [12] and the generalized Gaussian [13] ones. Also, expressions exist for some multivariate distributions such as Wishart [6], multivariate Gaussian and generalized Gaussian [14], and multivariate Cauchy [15]. However, for other multivariate distributions, the statistical distances require dealing with multiple integrals which are analytically intractable, and so, it is not easy to find closed-form expressions for all the divergences. As a consequence, different approximation techniques have been developed: Monte-Carlo (MC) integration technique [16], variational approximation [17], lower and upper bounds approximation [18], etc. The MC sampling can efficiently estimate these statistical distances and achieve high accuracy when a large number of samples is provided. Unfortunately, MC sampling is a time-consuming process and is not practically possible in many applications.

Among the aforementioned distributions, the multivariate t-distributions (MTD) [19] has received much attention and has been used in several signal and image processing applications where non-Gaussian statistics are necessary. Indeed, the MTD has been used for modeling the non-Gaussian heavy-tailed clutter measured by high-resolution radars [20], [21]. It is also used to model the wavelet coefficients for speckle denoising [22], hyperspectral anomaly detection [23], multi-target tracking [24], etc. The MTD belongs to the elliptical symmetric distributions [25] and has the multivariate Cauchy distribution (MCD) as special case. To our knowledge no closed-form expression existed for the RD and KLD between two MTDs. Therefore, the objective of this letter is to derive an expression for the RD between two zero-mean MTDs to bring solutions for future work on statistical signal processing, machine learning and other related fields in computer science. Moreover, we derive an analytic expression of the KLD since it is a special case of the RD avoiding approximation using expensive MC techniques. Finally, an implementation of the proposed RD expression is provided and a comparison is made with the MC sampling. Section II provides the closed-form expressions of the RD and KLD between two zero-mean MTDs. Section III presents the implementation of the divergence and a comparison with MC sampling method. A summary and conclusions are provided in the final section.

II. Multivariate t-Distribution and Rényi Divergence of Order β

A p-dimensional real random vector $\mathbf{X} = (X_1, ..., X_p)^T$ is said to have the p-variate t-distribution with degrees of freedom ν, mean vector $\boldsymbol{\mu}$, and correlation matrix $\boldsymbol{\Sigma}$ if its joint probability density function (pdf) is given by

$$f_{\mathbf{X}}(\boldsymbol{x}|\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}, p) = A\left[1 + \frac{1}{\nu}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right]^{-\frac{\nu+p}{2}} \quad (1)$$

where $A = \frac{\Gamma\left(\frac{\nu+p}{2}\right)|\boldsymbol{\Sigma}|^{-0.5}}{\Gamma(\nu/2)\nu^{\frac{p}{2}}\pi^{\frac{p}{2}}}$. The case $\nu = 1$ corresponds to MCD and $\nu \to \infty$ yields the multivariate normal distribution. Consider $\mathbf{X}^1$ and $\mathbf{X}^2$ be two real random vectors that follow zero-mean MTDs with pdfs $f_{\mathbf{X}^1}(\boldsymbol{x}|\nu_1, \boldsymbol{\Sigma}_1, p)$ and $f_{\mathbf{X}^2}(\boldsymbol{x}|\nu_2, \boldsymbol{\Sigma}_2, p)$ given by (1). The RD between $\mathbf{X}^1$ and $\mathbf{X}^2$ is given by

$$\begin{aligned} D_R^\beta(\mathbf{X}^1||\mathbf{X}^2) &= \frac{1}{\beta - 1}\ln\int_{\mathbb{R}^p} f^\beta_{\mathbf{X}^1}(\boldsymbol{x}|\nu_1, \boldsymbol{\Sigma}_1, p) f^{1-\beta}_{\mathbf{X}^2}(\boldsymbol{x}|\nu_2, \boldsymbol{\Sigma}_2, p)\mathrm{d}\boldsymbol{x} \\ &= \frac{1}{\beta - 1}\ln E_{\mathbf{X}^1}\left\{\left(\frac{f_{\mathbf{X}^1}(\boldsymbol{x}|\nu_1, \boldsymbol{\Sigma}_1, p)}{f_{\mathbf{X}^2}(\boldsymbol{x}|\nu_2, \boldsymbol{\Sigma}_2, p)}\right)^{\beta - 1}\right\}. \quad (2) \end{aligned}$$

Consequently, the closed form expression of the RD between two zero-mean MTDs is given by (see Appendix A for demonstration)

$$\begin{aligned} D_R^\beta(\mathbf{X}^1||\mathbf{X}^2) = \frac{1}{\beta - 1}\Bigg[&\beta\ln\left(\frac{\Gamma\left(\frac{\nu_1+p}{2}\right)}{\Gamma\left(\frac{\nu_2+p}{2}\right)}\frac{\Gamma\left(\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)}\frac{\nu_2^{\frac{p}{2}}}{\nu_1^{\frac{p}{2}}}\right) + \ln\left(\frac{\Gamma\left(\frac{\nu_2+p}{2}\right)}{\Gamma\left(\frac{\nu_2}{2}\right)}\right) \\ &+ \ln\left(\frac{\Gamma(\delta_1 + \delta_2 - \frac{p}{2})}{\Gamma(\delta_1 + \delta_2)}\right) - \frac{\beta}{2}\sum_{i=1}^p \ln\lambda_i \\ &+ \ln F_D^{(p)}\Bigg(\delta_1, \underbrace{\frac{1}{2}, ..., \frac{1}{2}}_{p}; \delta_1 + \delta_2; 1 - \frac{\nu_2}{\nu_1}\frac{1}{\lambda_1}, ..., 1 - \frac{\nu_2}{\nu_1}\frac{1}{\lambda_p}\Bigg)\Bigg] \quad (3) \end{aligned}$$

where $\delta_1 = \frac{\nu_1+p}{2}\beta$ and $\delta_2 = \frac{\nu_2+p}{2}(1 - \beta)$ are notations used here to alleviate the writing of the equation, and $\lambda_1, ..., \lambda_p$

are the eigenvalues of the real matrix $\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}$ arranged in ascending order. The function $F_D^{(p)}$ represents the Lauricella D-hypergeometric series defined for p variables (see Appendix D). The series is convergent if $|1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i}|<1,\ i=1,...,p$. If not, other Lauricella series expressions can be provided using transformations (see Appendix D) that can guarantee the convergence of the series. These equivalent expressions will be presented in the section III. For the particular case corresponding to the MCD where $\nu_1=\nu_2=1$, the RD of order β is given by

$$D_R^\beta(\mathbf{X}^1||\mathbf{X}^2)=\frac{1}{\beta-1}\Big[-\frac{\beta}{2}\sum_{i=1}^{p}\ln\lambda_i + \ln F_D^{(p)}\Big(\frac{1+p}{2}\beta,\underbrace{\frac{1}{2},...,\frac{1}{2}}_{p};\frac{1+p}{2};1-\frac{1}{\lambda_1},...,1-\frac{1}{\lambda_p}\Big)\Big]. \quad (4)$$

The KLD is a particular case of the RD where $D_{KL}(\mathbf{X}^1||\mathbf{X}^2)=\lim_{\beta\to 1}D_R^\beta(\mathbf{X}^1||\mathbf{X}^2)$ [26]. As a consequence the KLD between two zero-mean MTDs is given by (see Appendix B for more details)

$$D_{KL}(\mathbf{X}^1||\mathbf{X}^2)=\ln\left(\frac{\Gamma(\frac{\nu_1+p}{2})}{\Gamma(\frac{\nu_2+p}{2})}\frac{\Gamma(\frac{\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})}\frac{\nu_2^{\frac{p}{2}}}{\nu_1^{\frac{p}{2}}}\right)+\frac{\nu_2-\nu_1}{2}\Big[\psi(\frac{\nu_1+p}{2}) - \psi(\frac{\nu_1}{2})\Big]-\frac{1}{2}\sum_{i=1}^{p}\ln\lambda_i-\frac{\nu_2+p}{2}\prod_{i=1}^{p}\left(\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i}\right)^{\frac{1}{2}}\frac{\partial}{\partial a}\Big\{F_D^{(p)}\Big(\frac{\nu_1+p}{2}, \frac{1}{2},...,\frac{1}{2};a+\frac{\nu_1+p}{2};1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_1},...,1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_p}\Big)\Big\}\Big|_{a=0}. \quad (5)$$

where $\psi(.)$ is the digamma function defined as the logarithmic derivative of the Gamma function. For the particular case of the MCD where $\nu_1=\nu_2=1$, the KLD is given by

$$D_{KL}(\mathbf{X}^1||\mathbf{X}^2)=-\frac{1}{2}\sum_{i=1}^{p}\ln\lambda_i-\frac{1+p}{2}\prod_{i=1}^{p}\lambda_i^{-\frac{1}{2}}\frac{\partial}{\partial a}\Big\{F_D^{(p)}\Big(\frac{1+p}{2}, \frac{1}{2},...,\frac{1}{2};a+\frac{1+p}{2};1-\frac{1}{\lambda_1},...,1-\frac{1}{\lambda_p}\Big)\Big\}\Big|_{a=0}. \quad (6)$$

The equation (6) has been demonstrated in [15].

III. Implementation and comparison with Monte-Carlo technique

As mentioned in section II, other Lauricella series expressions can be provided using some transformations. These new expressions differ in the region of convergence. Therefore, for a valid expression of the RD, the convergence of the Lauricella series need to be guaranteed. Three cases can be identified.

A. Case $(\nu_1/\nu_2)\lambda_p>\ldots>(\nu_1/\nu_2)\lambda_1>1$

The Lauricella series in (3) is convergent since $|1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i}|<1, i=1,...,p$.

B. Case $1>(\nu_1/\nu_2)\lambda_p>\ldots>(\nu_1/\nu_2)\lambda_1$

Thanks to the Lauricella transformations given in Appendix D, we can provide a new convergent form of the Lauricella series. Indeed, using the transformation (33), the Lauricella series in (3) can be transformed as follows

$$F_D^{(p)}\Big(\delta_1,\frac{1}{2},...,\frac{1}{2};\delta_1+\delta_2;1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_1},...,1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_p}\Big) =\prod_{i=1}^{p}(\frac{\nu_1}{\nu_2}\lambda_i)^{\frac{1}{2}}F_D^{(p)}\Big(\delta_2,\frac{1}{2},...,\frac{1}{2};\delta_1+\delta_2;1-\frac{\nu_1}{\nu_2}\lambda_1,...,1-\frac{\nu_1}{\nu_2}\lambda_p\Big) \quad (7)$$

TABLE I

Computation of $F_D^{(p)}(.)$ and ${}_2F_1(.)$, $p=3,\beta=0.5,\nu_1=\nu_2=1$

		$N=20$		$N=30$	
$1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda}$	${}_2F_1(.)$	$F_D^{(p)}$	$\|\epsilon\|$	$F_D^{(p)}$	$\|\epsilon\|$
0.1	1.0818	1.0818	6.6613e-16	1.0818	6.6613e-16
0.3	1.3015	1.3015	3.4061e-13	1.3015	8.8817e-16
0.5	1.6568	1.6568	2.8780e-8	1.6568	1.6404e-11
0.7	2.3592	2.3591	8.3425e-5	2.3592	1.4177e-6
0.9	4.8050	4.7097	95321e-2	4.7828	2.2172e-2

C. Case $(\nu_1/\nu_2)\lambda_p>1$ and $(\nu_1/\nu_2)\lambda_1<1$

This case guarantees that $0\le 1-\lambda_j/\lambda_p<1$, $j=1,..,p-1$ and $0\le 1-(\nu_2/\nu_1)1/\lambda_p<1$. The Lauricella series in (3) can be transformed using (35). The new one is as follows

$$F_D^{(p)}\Big(\delta_1,\frac{1}{2},...,\frac{1}{2};\delta_1+\delta_2;1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_1},...,1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_p}\Big) =(\frac{\nu_2}{\nu_1}\frac{1}{\lambda_p})^{\delta_2}\prod_{i=1}^{p}(\frac{\nu_1}{\nu_2}\lambda_i)^{\frac{1}{2}}F_D^{(p)}\Big(\delta_2,\frac{1}{2},...,\frac{1}{2},\delta_1+\delta_2-\frac{p}{2};\delta_1+\delta_2; 1-\frac{\lambda_1}{\lambda_p},...,1-\frac{\lambda_{p-1}}{\lambda_p},1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_p}\Big). \quad (8)$$

Practically, to implement the RD, the infinite sums of the Lauricella series are replaced by finite sums. Thus, the indices m_i in (30) satisfy $0\le m_i\le N$ where N is the upper bound chosen to achieve a desired precision.

For the particular case $\lambda_1=...=\lambda_p=\lambda$, the Lauricella series becomes equal to the Gauss hypergeometric function, $F_D^{(p)}(\delta_1,\frac{1}{2},...,\frac{1}{2};\delta_1+\delta_2;1-\frac{\nu 2}{\nu_1}\frac{1}{\lambda},...,1-\frac{\nu 2}{\nu_1}\frac{1}{\lambda})={}_2F_1(\delta_1,\frac{p}{2},\delta_1+\delta_2,1-\frac{\nu 2}{\nu_1}\frac{1}{\lambda})$ allowing thus to compare the accuracy of calculation between them. Table I shows the absolute value of the error $|\epsilon|$ computed between $F_D^{(p)}(.)$ and ${}_2F_1(.)$ where $p=3,\beta=0.5,N=\{20,30\}$ and $\nu_1=\nu_2=1$. With these particular values, ${}_2F_1(1,\frac{3}{2},2,1-1/\lambda)=2/(\lambda^{-1}+\lambda^{-1/2})$. It is clear to see that $|\epsilon|$ is low and increases when $1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda}$ is close to 1.

In the following, we compare the MC technique with our numerical approximation of the closed-form expression of the RD. The MC method involves sampling a large number of samples to compute the sum instead of the integral. Here, the experiment is repeated 2000 times for each sample size. The following values are chosen for the experiment: $p=3$, ($\nu_1=2$, $\Sigma_{11}=2$, $\Sigma_{22}=2$, $\Sigma_{33}=2$, $\Sigma_{12}=1.2$, $\Sigma_{13}=0.4$, $\Sigma_{23}=0.6$) and ($\nu_2=4$, $\Sigma_{11}=1$, $\Sigma_{22}=1$, $\Sigma_{33}=1$, $\Sigma_{12}=0.3$, $\Sigma_{13}=0.1$, $\Sigma_{23}=0.4$). We recall that our numerical approximation method only depends on N. However the MC method depends on the sample size. For a value of $N=30$, the accuracy of the numerical approximation is evaluated to 10^{-9}. The accuracy can be further improved by increasing the value of N, but it will increase the computation time. Figure 1 shows the absolute value of bias, the mean square error (MSE), and the box plot of the difference between the numerical approximation of the RD of order $\beta=1/2$ and the MC method, as a function of the sample sizes. It is clearly seen that the bias and the MSE decrease when the sample size increases. And so, for large sample sizes, and therefore longer simulation times, the MC technique becomes close to our numerical approximation results. Furthermore, the box plot of the error may show this behavior.

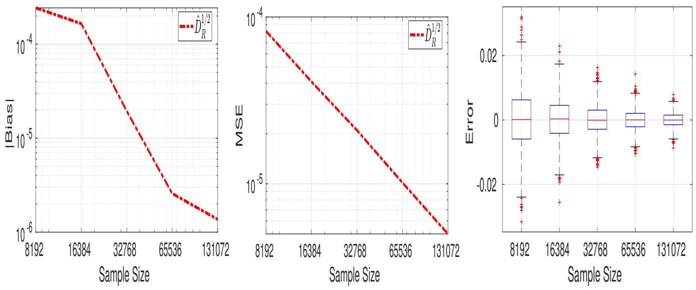

Fig. 1. Bias (**left**) and MSE (**middle**) of the difference between our numerical approximation and MC sampling of RD of order $\beta = 0.5$ for MTD, and Box plot of the error (**right**).

In addition, the computing time of the proposed approximation and the classical MC sampling method performed by simulations are compared for a similar accuracy. Then, for a precision of 10^{-6}, the calculation time of our method is 6 time smaller than MC method. This demonstrates the advantages of the proposed method: the analytical expression of the divergence and the computational interest.

IV. Conclusion

In this paper, a closed-form expression of the RD and KLD between two zero-mean MTDs have been derived. The divergence expressions depend on the Lauricella D-hypergeometric series. We have also proposed an efficient implementation of the divergence to guarantee the convergence of the Lauricella series. For a large number of samples, the MC sampling method is close to numerical approximated RD value. The proposed numerical approximation of the expression of the RD between two zero-mean MTDs distribution is efficient in terms of computation time and accuracy.

Appendix A
Expression of $D_R^{\beta}(\mathbf{X}^1||\mathbf{X}^2)$

The pdf of the $\mathbf{X}^i$ is defined for $i = 1, 2$ by

$$f_{\mathbf{X}^i}(\boldsymbol{x}|\nu_i, \boldsymbol{\Sigma}_i, p) = A_i \left[1 + \frac{1}{\nu_i}\boldsymbol{x}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}\right]^{-\frac{\nu_i+p}{2}} \tag{9}$$

and $A_i = \frac{\Gamma(\frac{\nu_i+p}{2})}{\Gamma(\frac{\nu_i}{2})\nu_i^{\frac{p}{2}}\pi^{\frac{p}{2}}|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}}$. We denote by $\mathbb{H} = E_{\mathbf{X}^1}\left\{\left(\frac{f_{\mathbf{X}^1}(\boldsymbol{x}|\nu_1,\boldsymbol{\Sigma}_1,p)}{f_{\mathbf{X}^2}(\boldsymbol{x}|\nu_2,\boldsymbol{\Sigma}_2,p)}\right)^{\beta-1}\right\}$ developed as follows

$$\mathbb{H} = A_1^{\beta}A_2^{1-\beta}\int_{\mathbb{R}^p}\left(1+\frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{x}}{\nu_1}\right)^{-\delta_1}\left(1+\frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{x}}{\nu_2}\right)^{-\delta_2}\mathrm{d}\boldsymbol{x}. \tag{10}$$

Consider transformation $\boldsymbol{y} = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{x}$ where $\boldsymbol{y} = [y_1, y_2, ..., y_p]^T$. The Jacobian determinant is given by $\mathrm{d}\boldsymbol{y} = |\boldsymbol{\Sigma}_1|^{-1/2}\mathrm{d}\boldsymbol{x}$ (Theorem 1.12 in [27]) and matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1^{\frac{1}{2}}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1^{\frac{1}{2}}$ is a real symmetric matrix given that $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are real symmetric matrices. Accordingly, the expectation is evaluated as follows

$$\mathbb{H} = \frac{A_1^{\beta}A_2^{1-\beta}}{|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}}\int_{\mathbb{R}^p}\left(1+\frac{\boldsymbol{y}^T\boldsymbol{y}}{\nu_1}\right)^{-\delta_1}\left(1+\frac{\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}}{\nu_2}\right)^{-\delta_2}\mathrm{d}\boldsymbol{y}. \tag{11}$$

Knowing that (Definition 1.1.5 [28])

$$\left(1+\frac{\boldsymbol{y}^T\boldsymbol{y}}{\nu_1}\right)^{-\delta_1} = \frac{1}{\Gamma(\delta_1)}\int_0^{+\infty} t^{\delta_1-1}e^{-t(1+\frac{\boldsymbol{y}^T\boldsymbol{y}}{\nu_1})}\mathrm{d}t, \tag{12}$$

$$\left(1+\frac{1}{\nu_2}\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right)^{-\delta_2} = \frac{1}{\Gamma(\delta_2)}\int_0^{+\infty} x^{\delta_2-1}e^{-x(1+\frac{1}{\nu_2}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y})}\mathrm{d}x, \tag{13}$$

the expectation expression is given by

$$\mathbb{H} = \frac{A_1^{\beta}A_2^{1-\beta}}{|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}\Gamma(\delta_1)\Gamma(\delta_2)}\int_0^{+\infty}\int_0^{+\infty} t^{\delta_1-1}x^{\delta_2-1}e^{-(t+x)} \left(\int_{\mathbb{R}^p} e^{-(\frac{t}{\nu_1}\boldsymbol{y}^T\boldsymbol{y}+\frac{x}{\nu_2}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y})}\mathrm{d}\boldsymbol{y}\right)\mathrm{d}t\mathrm{d}x. \tag{14}$$

Using the transformation $\boldsymbol{z} = (\frac{t}{\nu_1})^{1/2}\boldsymbol{y}$ with the Jacobian determinant given by $\mathrm{d}\boldsymbol{z} = (\frac{t}{\nu_1})^{p/2}\mathrm{d}\boldsymbol{y}$, the last multiple integral is given as follows

$$\int_{\mathbb{R}^p} e^{-(\frac{t}{\nu_1}\boldsymbol{y}^T\boldsymbol{y}+\frac{x}{\nu_2}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y})}\mathrm{d}\boldsymbol{y} = \left(\frac{t}{\nu_1}\right)^{-\frac{p}{2}}\int_{\mathbb{R}^p} e^{-\boldsymbol{z}^T(\boldsymbol{I}_p+\frac{\nu_1}{\nu_2}\frac{x}{t}\boldsymbol{\Sigma})\boldsymbol{z}}\mathrm{d}\boldsymbol{z}.$$

where $\boldsymbol{I}_p$ is the identity matrix of size p. Using the following property which states that if $\boldsymbol{x}$ is a $p\times 1$ vector of real variables and $\mathbf{A}$ a $p\times p$ symmetric positive definite matrix of constants the following integral is true [27]

$$\int_{\boldsymbol{x}} e^{-\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}}\mathrm{d}\boldsymbol{x} = \pi^{\frac{p}{2}}|\mathbf{A}|^{-\frac{1}{2}} \tag{15}$$

the new expectation as a consequence is given by

$$\mathbb{H} = \frac{A_1^{\beta}A_2^{1-\beta}\pi^{\frac{p}{2}}}{|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}\Gamma(\delta_1)\Gamma(\delta_2)}\mathbb{J} \tag{16}$$

where the multiple integral $\mathbb{J}$ is defined

$$\mathbb{J} = \int_0^{+\infty}\int_0^{+\infty} t^{\delta_1-1}x^{\delta_2-1}e^{-(t+x)}\left(\frac{t}{\nu_1}\right)^{-\frac{p}{2}}|\boldsymbol{I}_p+\frac{\nu_1}{\nu_2}\frac{x}{t}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\mathrm{d}t\mathrm{d}x.$$

Let $\lambda_1, ..., \lambda_p$ be the eigenvalues of $\boldsymbol{\Sigma}$. The following determinant is given as follows [27]

$$|\boldsymbol{I}_p+\frac{\nu_1}{\nu_2}\frac{x}{t}\boldsymbol{\Sigma}| = \prod_{i=1}^{p}(1+\frac{\nu_1}{\nu_2}\frac{x}{t}\lambda_i). \tag{17}$$

The following change of variables $u = x/t$ and $v = x + t$, allows an expression of the Jacobian determinant defined by $\mathrm{d}u\mathrm{d}v = (1+u)^2/v\mathrm{d}x\mathrm{d}t$. Accordingly, one can write after development

$$\mathbb{J} = \nu_1^{\frac{p}{2}}\Gamma\left(\delta_1+\delta_2-\frac{p}{2}\right)\int_0^{+\infty} u^{\delta_2-1}(1+u)^{-(\delta_1+\delta_2-\frac{p}{2})} \times\prod_{i=1}^{p}(1+\frac{\nu_1}{\nu_2}u\lambda_i)^{-\frac{1}{2}}\mathrm{d}u. \tag{18}$$

By applying another transformation $y = 1/(1+u)$, the last equation is given by

$$\mathbb{J} = \nu_1^{\frac{p}{2}}\Gamma\left(\delta_1+\delta_2-\frac{p}{2}\right)\prod_{i=1}^{p}\left(\frac{\nu_1}{\nu_2}\lambda_i\right)^{-\frac{1}{2}}\int_0^1 y^{\delta_1-1}(1-y)^{\delta_2-1} \times\prod_{i=1}^{p}\left(1-(1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i})y\right)^{-\frac{1}{2}}\mathrm{d}y. \tag{19}$$

The last integral represents the Lauricella D-hypergeometric function, denoted $F_D^{(p)}(.)$. The integral representation of $F_D^{(p)}(.)$ is given by (32) in Appendix D. Consequently, (19) is given by

$$\mathbb{J} = \nu_1^{\frac{p}{2}} \Gamma\left(\delta_1+\delta_2-\frac{p}{2}\right) \prod_{i=1}^{p} \left(\frac{\nu_1}{\nu_2}\lambda_i\right)^{-\frac{1}{2}} \frac{\Gamma(\delta_1)\Gamma(\delta_2)}{\Gamma(\delta_1+\delta_2)}$$
$$\times F_D^{(p)}\left(\delta_1, \frac{1}{2}, ..., \frac{1}{2}; \delta_1+\delta_2; 1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_1}, ..., 1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_p}\right). \quad (20)$$

Plugging the expression of $\mathbb{J}$ in (16), and using the notation $B_i = 1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i}, i=1,...p$ to alleviate the writing of the equation, the new expression of the expectation is given by

$$\mathbb{H} = \left(\frac{\Gamma(\frac{\nu_1+p}{2})}{\Gamma(\frac{\nu_2+p}{2})}\frac{\Gamma(\frac{\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})}\frac{\nu_2^{\frac{p}{2}}}{\nu_1^{\frac{p}{2}}}\right)^{\beta} \frac{\Gamma(\frac{\nu_2+p}{2})}{\Gamma(\frac{\nu_2}{2})}\frac{\Gamma(\delta_1+\delta_2-\frac{p}{2})}{\Gamma(\delta_1+\delta_2)} \prod_{i=1}^{p} \lambda_i^{-\frac{\beta}{2}}$$
$$\times F_D^{(p)}\Big(\delta_1, \underbrace{\frac{1}{2}, ..., \frac{1}{2}}_{p}; \delta_1+\delta_2; B_1, ..., B_p\Big). \quad (21)$$

As a consequence, the RD of order β is given by

$$D_R^{\beta}(\mathbf{X}^1||\mathbf{X}^2) = \frac{1}{\beta-1}\ln\mathbb{H}. \quad (22)$$

Appendix B
Expression of $D_{KL}(\mathbf{X}^1||\mathbf{X}^2)$

The KLD between $\mathbf{X}^1$ and $\mathbf{X}^2$ is given by

$$D_{KL}(\mathbf{X}^1||\mathbf{X}^2) = \lim_{\beta\to 1}\frac{1}{\beta-1}\ln\mathbb{H} = \frac{\partial}{\partial\beta}\left\{\ln\mathbb{H}\right\}\Big|_{\beta=1}. \quad (23)$$

By using the expression (3) of DR, it is easy to prove that

$$\frac{\partial}{\partial\beta}\left\{\ln\frac{\Gamma(\delta_1+\delta_2-\frac{p}{2})}{\Gamma(\delta_1+\delta_2)}\right\}\Big|_{\beta=1} = \frac{\nu_1-\nu_2}{2}\left[\psi(\frac{\nu_1}{2})-\psi(\frac{\nu_1+p}{2})\right] \quad (24)$$

and, since

$$F_D^{(p)}\left(\frac{\nu_1+p}{2}, \frac{1}{2}, ..., \frac{1}{2}; \frac{\nu_1+p}{2}; B_1, ..., B_p\right) = \prod_{i=1}^{p}\left(\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i}\right)^{-\frac{1}{2}}$$

where $B_i = 1-\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i}, i=1,...,p$, we have the following

$$\frac{\partial}{\partial\beta}\left\{\ln F_D^{(p)}\left(\delta_1, \frac{1}{2}, ..., \frac{1}{2}; \delta_1+\delta_2; B_1, ..., B_p\right)\right\}\Big|_{\beta=1}$$
$$= \prod_{i=1}^{p}\left(\frac{\nu_2}{\nu_1}\frac{1}{\lambda_i}\right)^{\frac{1}{2}} \frac{\partial}{\partial\beta}\left\{F_D^{(p)}\left(\delta_1, \frac{1}{2}, ..., \frac{1}{2}; \delta_1+\delta_2; B_1, ..., B_p\right)\right\}\Big|_{\beta=1}. \quad (25)$$

Using the following relation (demonstration in Appendix C)

$$\frac{\partial}{\partial\beta}\left\{F_D^{(p)}\left(\delta_1, \frac{1}{2}, ..., \frac{1}{2}; \delta_1+\delta_2; B_1, ..., B_p\right)\right\}\Big|_{\beta=1}$$
$$= -\frac{\nu_2+p}{2}\frac{\partial}{\partial a}\left\{F_D^{(p)}\left(\frac{\nu_1+p}{2}, \frac{1}{2}, ..., \frac{1}{2}; a+\frac{\nu_1+p}{2}; B_1, ..., B_p\right)\right\}\Big|_{a=0} \quad (26)$$

we can finally establish the expression of KLD for two zero-mean real MTDs.

Appendix C

We use the following notation $\alpha = \sum_{i=1}^{p} m_i$ to alleviate the writing. Using the Lauricella series definition (30), and the Pochhammer symbol (31), the derivative of the Lauricella series with respect to β is given as follows

$$\frac{\partial}{\partial\beta}\left\{F_D^{(p)}\left(\delta_1, \frac{1}{2}, ..., \frac{1}{2}; \delta_1+\delta_2; B_1, ..., B_p\right)\right\}$$
$$= \sum_{\substack{m_1,...,\\m_p=0}}^{+\infty}\frac{\partial}{\partial\beta}\left\{\frac{(\delta_1)_\alpha}{(\delta_1+\delta_2)_\alpha}\right\}\prod_{i=1}^{p}\left(\frac{1}{2}\right)_{m_i}\frac{B_i^{m_i}}{m_i!}. \quad (27)$$

By using the following property $\frac{\partial}{\partial c}(c)_k = (c)_k[\psi(c+k)-\psi(c)]$, which is easy to prove using the derivative of Pochhammer (31), we can deduce that

$$\frac{\partial}{\partial\beta}\left\{\frac{(\delta_1)_\alpha}{(\delta_1+\delta_2)_\alpha}\right\}\Big|_{\beta=1} = \left(\frac{\nu_2+p}{2}\right)\left[\psi(\frac{\nu_1+p}{2}+\alpha)-\psi(\frac{\nu_1+p}{2})\right]$$
$$= -\left(\frac{\nu_2+p}{2}\right)\frac{\partial}{\partial a}\left\{\frac{(\frac{\nu_1+p}{2})_\alpha}{(a+\frac{\nu_1+p}{2})_\alpha}\right\}\Big|_{a=0}. \quad (28)$$

As a consequence,

$$\frac{\partial}{\partial\beta}\left\{F_D^{(p)}\left(\delta_1, \frac{1}{2}, ..., \frac{1}{2}; \delta_1+\delta_2; B_1, ..., B_p\right)\right\}\Big|_{\beta=1}$$
$$= -\left(\frac{\nu_2+p}{2}\right)\sum_{\substack{m_1,...,\\m_p=0}}^{+\infty}\frac{\partial}{\partial a}\left\{\frac{(\frac{\nu_1+p}{2})_\alpha}{(a+\frac{\nu_1+p}{2})_\alpha}\right\}\Big|_{a=0}\prod_{i=1}^{p}\left(\frac{1}{2}\right)_{m_i}\frac{B_i^{m_i}}{m_i!}$$
$$= -\frac{\nu_2+p}{2}\frac{\partial}{\partial a}\left\{F_D^{(p)}\left(\frac{\nu_1+p}{2}, \frac{1}{2}, ..., \frac{1}{2}; a+\frac{\nu_1+p}{2}; B_1, ..., B_p\right)\right\}\Big|_{a=0} \quad (29)$$

Appendix D
Lauricella series

The Lauricella series $F_D^{(n)}$ is given as follows [28], [14]

$$F_D^{(n)}(a, b_1, ..., b_n; c; x_1, ..., x_n)$$
$$= \sum_{\substack{m_1,...,\\m_n=0}}^{+\infty}\frac{(a)_{m_1+...+m_n}}{(c)_{m_1+...+m_n}}(b_1)_{m_1}...(b_n)_{m_n}\frac{x_1^{m_1}}{m_1!}...\frac{x_n^{m_n}}{m_n!} \quad (30)$$

where $|x_1|, ..., |x_n| < 1$. The Pochhammer symbol $(q)_i$ is defined as follows

$$(q)_i = q(q+1)...(q+i-1) = \frac{\Gamma(q+i)}{\Gamma(q)} \quad \text{if} \quad i = 1, 2, ... \quad (31)$$

Lauricella F_D can be expressed as a one-dimensional Euler integral for any number of variables n. It is defined when $\mathrm{Re}(a) > 0$ and $\mathrm{Re}(c-a) > 0$ by

$$F_D^{(n)}(a, b_1, ..., b_n; c; x_1, ..., x_n) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 u^{a-1}\times$$
$$(1-u)^{c-a-1}(1-ux_1)^{-b_1}...(1-ux_n)^{-b_n}\mathrm{d}u. \quad (32)$$

Several transformations can be applied as follows [29], [14]

$$F_D^{(n)}(a, b_1, ..., b_n; c; x_1, ..., x_n)$$
$$= \prod_{i=1}^{n}(1-x_i)^{-b_i}F_D^{(n)}(c-a, b_1, ..., b_n; c; \frac{x_1}{x_1-1}, ..., \frac{x_n}{x_n-1}) \quad (33)$$
$$= (1-x_1)^{-a}F_D^{(n)}(a, c-\sum_{i=1}^{n}b_i, b_2, ..., b_n; c; \frac{x_1}{x_1-1}, \frac{x_1-x_2}{x_1-1}, ..., \frac{x_1-x_n}{x_1-1}) \quad (34)$$
$$= (1-x_n)^{c-a}\prod_{i=1}^{n}(1-x_i)^{-b_i}F_D^{(n)}(c-a, b_1, \ldots, b_{n-1}, c-\sum_{i=1}^{n}b_i; c; \frac{x_1-x_n}{x_1-1}, \ldots, \frac{x_{n-1}-x_n}{x_{n-1}-1}, x_n). \quad (35)$$

REFERENCES

[1] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, Calif., 1961, vol. 1, pp. 547–561, University of California Press.

[2] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. on Communication Technology*, vol. 15, no. 1, pp. 52–60, February 1967.

[3] L. Giet and M. Lubrano, "A minimum Hellinger distance estimator for stochastic differential equations: An application to statistical inference for continuous time interest rate models," *Computational Statistics & Data Analysis*, vol. 52, no. 6, pp. 2945 – 2965, 2008.

[4] K. Roland and U. Andreas, "Lightweight probabilistic texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 241–253, 2010.

[5] Wagner B. Silva, Corina C. Freitas, Sidnei J. S. Sant'Anna, and Alejandro C. Frery, "Classification of segments in PolSAR imagery by minimum stochastic distances between Wishart distributions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1263–1273, 2013.

[6] N. Bouhlel and S. Méric, "Multilook polarimetric SAR change detection using stochastic distances between matrix-variate $\mathcal{G}_d^0$ distributions," *IEEE Trans. on Geosci. Remote Sens.*, vol. 58, no. 10, pp. 6823–6843, 2020.

[7] S. R. Kadiri and P. Alku, "Excitation features of speech for speaker-specific emotion detection," *IEEE Access*, vol. 8, pp. 60382–60391, 2020.

[8] M. Kim, Y. Kim, J. Yoo, J. Wang, and H. Kim, "Regularized speaker adaptation of KL-HMM for dysarthric speech recognition," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 9, pp. 1581–1591, 2017.

[9] A. Seghouane and S. Amari, "The aic criterion and symmetrizing the Kullback–Leibler divergence," *IEEE Trans. on Neural Networks*, vol. 18, no. 1, pp. 97–106, 2007.

[10] Y. Petetin, Y. Janati, and F. Desbouvries, "Structured variational bayesian inference for Gaussian state-space models with regime switching," *IEEE Signal Process. Lett.*, vol. 28, pp. 1953–1957, 2021.

[11] F. Chyzak and F. Nielsen, "A closed-form formula for the Kullback-Leibler divergence between Cauchy distributions," *arXiv:1905.10965*, 2019.

[12] S. K. Choy and C. S. Tong, "Statistical wavelet subband characterization based on generalized Gamma density and its application in texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 281–289, Feb 2010.

[13] 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.

[14] N. Bouhlel and A. Dziri, "Kullback–leibler divergence between multivariate generalized Gaussian distributions," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1021–1025, 2019.

[15] N. Bouhlel and D. Rousseau, "A generic formula and some special cases for the Kullback-Leibler divergence between central multivariate Cauchy distributions," *Entropy*, vol. 24, no. 6, 2022.

[16] S. Cui and M. Datcu, "Comparison of Kullback-Leibler divergence approximation methods between Gaussian mixture models for satellite image retrieval," in *2015 IEEE IGARSS*, July 2015, pp. 3719–3722.

[17] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *2007 IEEE ICASSP '07*, 2007, vol. 4, pp. IV–317–IV–320.

[18] 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.

[19] K. Samuel and N. Saralees, *Multivariate T-Distributions and Their Applications*, Cambridge University Press, 2004.

[20] A. Balleri, A. Nehorai, and J. Wang, "Maximum likelihood estimation for compound-Gaussian clutter with inverse gamma texture," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 43, no. 2, pp. 775–779, 2007.

[21] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 876–879, 2012.

[22] J. J. Ranjani and S. J. Thiruvengadam, "Generalized sar despeckling based on dtcwt exploiting interscale and intrascale dependences," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 552–556, 2011.

[23] K. Kayabol, E. B. Aytekin, S. Arisoy, and E. E. Kuruoglu, "Skewed t-distribution for hyperspectral anomaly detection based on autoencoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[24] D. Peng, J. Zhongliang, L. Henry, S. Kai, and L. Minzhe, "The labeled multi-bernoulli filter for multitarget tracking with glint noise," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 55, no. 5, pp. 2253–2268, 2019.

[25] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. on Signal Process.*, vol. 60, no. 11, pp. 5597–5625, 2012.

[26] Leandro Pardo, *Statistical Inference Based on Divergence Measures*, Chapman and Hall/CRC, New York, 2005.

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

[28] A.M. Mathai and H.J. Haubold, *Special Functions for Applied Scientists*, Springer, 2008.

[29] H. Exton, *Multiple Hypergeometric Functions and Applications*, Wiley, New York, 1976.