Evaluating Dissimilarities between Two Moving-Average Models: a Comparative Study between Jeffrey's Divergence and Rao Distance

Léo Legrand*† and Éric Grivel†
*Thales Systèmes Aéroportés, Pessac, FRANCE
†Bordeaux university - INP Bordeaux ENSEIRB-MATMECA - IMS - UMR CNRS 5218, Talence, FRANCE

***Abstract*—The autoregressive models (AR) and moving-average models (MA) are regularly used in signal processing. Previous works have been done on dissimilarity measures between AR models by using a Riemannian distance, the Jeffrey's divergence (JD) and the spectral distances such as the Itakura-Saito divergence. In this paper, we compare the Rao distance and the JD for MA models and more particularly in the case of 1st-order MA models for which an analytical expression of the inverse of the covariance matrix is available. More particularly, we analyze the advantages of the Rao distance use. Secondly, the simulation part compares both dissimilarity measures depending on the MA parameters but also on the number of data available.**

***Index Terms*—Jeffrey's divergence, Rao distance, moving-average models.**

I. INTRODUCTION

For many years, a great deal of interest has been paid to time series models, namely the autoregressive (AR) models and their variants such as the time-varying AR models (TVAR), the multivariate AR models as well as the moving-average models (MA). For the last decades, they have been used in a wide range of applications from speech analysis to radar processing. Several issues have been addressed and include the estimations of the model parameters from noisy observations [1], the model order selection and model comparison. In this latter case, some divergences that aim at measuring the similarity between distributions of samples can be considered. In [2], Magnant *et al.* have thus suggested computing the Jeffrey's divergence (JD), which is the symmetric Kullback Leibler (KL) divergence, between the distributions of the successive samples of two TVAR models; the authors have also extended their approach to classify more than two AR models in various subsets and to compare motion models [3]. There are some other dissimilarity measures such as the Hellinger distance and the Bhattacharyya divergence. The reader may refer to [4] for a comparative study between them.

However, information geometry can bring in much. Concerning model comparison, the authors in [5] and [6] have already analyzed the information geometry of covariance matrices of AR models. Its use has already been relevant in different fields from segmentation or classification [7], [8] to radar processing [9].

In this paper, we propose a study that is complementary to the above works. More particularly, our purpose is to compare the JD and the Rao distance [10] between two MA models. Analytical expressions of both dissimilarity measures are provided, depending on the MA parameters and the number of samples that are considered. Then, properties, comments and various examples are given. More particularly, it is known [11] that under some assumptions, the Rao distance can be deduced from the JD. We will see in this paper in which cases this is true. We will hence show when the Rao distance is of interest to compare MA models.

This paper is organized as follows: In Section II, we first recall correlation properties of MA models that are useful to get the JD and the Rao distance. Then, the JD and the Rao distance expressions are given for 1st-order MA models. In Section III, both measures are compared with each other by using simulations.

In the following, $\lfloor . \rfloor$ is the integer part, Tr denotes the trace, the upperscripts ^{T} and ^{H} respectively denote the transpose and the hermitian. $\text{diag}\left([m_1, \ \dots, \ m_n]\right)$ is a $n \times n$ diagonal matrix whose main diagonal is $m_1, \ \dots, \ m_n$. $x_{k_1:k_2} = (x_{k_1}, \dots, x_{k_2})$ denotes the collection of samples from time k_1 to k_2.

II. JEFFREY'S DIVERGENCE AND RAO DISTANCE EXPRESSIONS FOR 1ST-ORDER MA MODELS

A. MA model: some properties

Let us consider M 1st-order MA models, where the m^{th} one is defined as follows:

$$x_k = b_0^{(m)} u_k^{(m)} + b_1^{(m)} u_{k-1}^{(m)}, \text{ for } m = 1, \dots, M, \tag{1}$$

where $b_1^{(m)}$ is the 1st real MA parameter of the m^{th} model, $b_0^{(m)} = 1$ without loss of generality and the driving processes $\{u_k^{(m)}\}_{m=1,\dots,M}$ are uncorrelated zero-mean white Gaussian processes with variance $\sigma^2_{u^{(m)}}$. The resulting correlation function satisfies:

$$\begin{cases} r_0^{(m)} = \sigma^2_{u^{(m)}} (1 + {b_1^{(m)}}^2) \\ r_1^{(m)} = r_{-1}^{(m)} = \sigma^2_{u^{(m)}} b_1^{(m)} \\ r_\tau^{(m)} = \text{E}\left[x_k x_{k-\tau}^*\right] = 0 \text{ for } |\tau| \geq 2 \end{cases} \tag{2}$$

In the following, some properties about 1st-order MA models are provided. They will be required further.

<u>Property 1</u>: the $k \times k$ covariance matrices $\{Q_k^{(m)}\}_{m=1,\dots,M}$ of the 1st-order MA models are tridiagonal and their inverses are symmetric. The analytical expression of each element of $(Q_k^{(m)})^{-1}$ at the i^{th} row and the j^{th} column satisfies for $|b_1^{(m)}| \neq 1$ and $(i,j) \in [1, \ k]^2$ [12]:

$$\left(Q_k^{(m)}\right)^{-1}_{i,j} = \left(Q_k^{(m)}\right)^{-1}_{j,i} = \tag{3}$$
$$\frac{1 + {b_1^{(m)}}^2}{r_0^{(m)} \cdot (1 - {b_1^{(m)}}^2)} \cdot \left[\left(-b_1^{(m)}\right)^{|i-j|} - \left(-b_1^{(m)}\right)^{2k-i-j+2} - \frac{\left(-b_1^{(m)}\right)^{i+j} \left(1 - \left(b_1^{(m)}\right)^{2k-2i+2}\right)\left(1 - \left(b_1^{(m)}\right)^{2k-2j+2}\right)}{1 - \left(b_1^{(m)}\right)^{2k+2}} \right].$$

Otherwise, when $|b_1^{(m)}| = 1$, the elements of $(Q_k^{(m)})^{-1}$ satisfy[1] [13]:

$$\left(Q_k^{(m)}\right)^{-1}_{i,j} = \left(Q_k^{(m)}\right)^{-1}_{j,i} = \begin{cases} \frac{(-1)^{j-i}}{\sigma^2_{u^{(m)}}} \frac{i(k+1-j)}{k+1}, \ i \leq j \\ \frac{(-1)^{i-j}}{\sigma^2_{u^{(m)}}} \frac{j(k+1-i)}{k+1}, \ i \geq j \end{cases} \tag{4}$$

[1]this result can be retrieved by continuity from (3)

Property 2: depending on the MA parameters, the correlation functions and the power spectral densities (PSDs) of MA models whose MA parameters are different can satisfy some specific relations.

1) *Inverse zeros:* If the m^{th} and the n^{th} 1st-order MA models are defined as follows:

$$b_1^{(n)} = \frac{1}{b_1^{(m)}}, \tag{5}$$

it can be shown that for any τ:

$$r_\tau^{(m)} = b_1^{(m)2} \frac{\sigma^2_{u^{(m)}}}{\sigma^2_{u^{(n)}}} r_\tau^{(n)} \text{ and } \frac{r_1^{(m)}}{r_0^{(m)}} = \frac{r_1^{(n)}}{r_0^{(n)}}. \tag{6}$$

Therefore, $r_\tau^{(m)} = r_\tau^{(n)}$ if:

$$\sigma^2_{u^{(m)}} = \frac{1}{b_1^{(m)2}} \sigma^2_{u^{(n)}}. \tag{7}$$

Studying the above case is motivated by the following reason: the MA models can be seen as the output of a finite-inpulse-response filtering of zero-mean white sequences with variances $\sigma^2_{u^{(l)}}$, with $l = m, n$. Their transfer functions are defined by:

$$H^{(m)}(z) = 1 + b_1^{(m)} z^{-1} \text{ and } H^{(n)}(z) = 1 + \frac{1}{b_1^{(m)}} z^{-1}. \tag{8}$$

Then, $H^{(n)}(z)$ can be rewritten as follows:

$$\begin{aligned} |b_1^{(m)}| H^{(n)}(z) &= \frac{|b_1^{(m)}|}{-b_1^{(m)}} \frac{-b_1^{(m)} - z^{-1}}{1 + b_1^{(m)} z^{-1}} \left(1 + b_1^{(m)} z^{-1}\right) \\ &= G(-b_1^{(m)}, z) H^{(m)}(z). \end{aligned} \tag{9}$$

In (9), $G(-b_1^{(m)}, z)$ is a Blaschke product [14] which can be seen as the transfer function of an all-pass filter.
By introducing the PSDs for both models defined by $S_{xx}^{(l)}(\theta) = \sigma^2_{u^{(l)}} |H^{(l)}(z)|^2_{|z=e^{j\theta}}$, $l = m, n$ with θ the normalized angular frequency, the above equation (9) leads to $S_{xx}^{(m)}(\theta) = S_{xx}^{(n)}(\theta)$ when the variances $\{\sigma^2_{u^{(l)}}\}_{l=m,n}$ satisfy (7).

2) *Opposites zeros:* If $b_1^{(n)} = -b_1^{(m)}$, then $r_\tau^{(n)} = (-1)^{|\tau|} r_\tau^{(m)}$ and $S_{xx}^{(m)}(\theta) = S_{xx}^{(n)}(\theta + \pi)$.

3) When $b_1 = 1$ (resp. -1), $S_{xx}(\theta) = 0$ for $\theta = \pm\pi$ (resp. 0).
In the next subsections, let us express the JD and the Rao distance between two 1st-order MA models.

B. Jeffrey's divergence between two 1st-order MA models

1) Analytical expression of the JD:

To analyze the dissimilarities in time between two MA models, we propose to compute the expression of the JD between the joint distributions of the values $x_{1:k}$ of the m^{th} and the n^{th} MA models, denoted $p_m(x_{1:k})$ and $p_n(x_{1:k})$ respectively.
As the JD is the symmetrized version of the KL divergence, let us recall the KL divergence between two multivariate normal densities $\mathcal{N}_m(\underline{\mu}_m, Q_m)$ and $\mathcal{N}_n(\underline{\mu}_n, Q_n)$:

$$\begin{aligned} KL(\mathcal{N}_m, \mathcal{N}_n) = &\frac{1}{2}\Big[\text{Tr}(Q_n^{-1} Q_m) - k - \ln\frac{\det Q_m}{\det Q_n} \\ &+ (\underline{\mu}_n - \underline{\mu}_m)^T Q_n^{-1} (\underline{\mu}_n - \underline{\mu}_m)\Big]. \end{aligned} \tag{10}$$

Due to (1), $p_l(x_{1:k}) = \mathcal{N}\left(\mathbf{0}_{k\times 1}, Q_k^{(l)}\right)$ for $l = m, n$ and the JD can be then deduced as follows:

$$\begin{aligned} JD_{mn}(k) &\triangleq JD\left(p_m(x_{1:k}), p_n(x_{1:k})\right) \\ &\underset{(10)}{=} -k + \frac{1}{2}\left[\text{Tr}\left(Q_k^{(n)^{-1}} Q_k^{(m)}\right) + \text{Tr}\left(Q_k^{(m)^{-1}} Q_k^{(n)}\right)\right]. \end{aligned} \tag{11}$$

Remark: by introducing the eigenvalues $\{\lambda_j^{(l)}\}_{j=1,\ldots,k}$ of $Q_k^{(l)}$:

$$JD_{mn}(k) = \frac{1}{2}\sum_{j=1}^{k}\left(\sqrt{\frac{\lambda_j^{(n)}}{\lambda_j^{(m)}}} - \sqrt{\frac{\lambda_j^{(m)}}{\lambda_j^{(n)}}}\right)^2. \tag{12}$$

In the following, our purpose is to express the JD for successive values of k:
When $k = 1$, as the variance of the MA sample $x_1^{(l)}$ is equal to $r_0^{(l)}$, for $l = m, n$, this leads to:

$$JD_{mn}(1) = -1 + \frac{1}{2}\left[\frac{r_0^{(m)}}{r_0^{(n)}} + \frac{r_0^{(n)}}{r_0^{(m)}}\right]. \tag{13}$$

When $k > 1$, it can be easily shown that:

$$\begin{aligned} \text{Tr}\left(\left(Q_k^{(n)}\right)^{-1} Q_k^{(m)}\right) &= r_0^{(m)} \sum_{i=1}^{k} \left(Q_k^{(n)}\right)^{-1}_{i,i} \\ &+ r_1^{(m)}\left[\sum_{i=1}^{k-1}\left(Q_k^{(n)}\right)^{-1}_{i,i+1} + \sum_{i=2}^{k}\left(Q_k^{(n)}\right)^{-1}_{i,i-1}\right]. \end{aligned} \tag{14}$$

Combining (11) and (14), and using the symmetric property of the inverse of the correlation matrices, one has:

$$\begin{aligned} JD_{mn}(k) = &-k + \frac{1}{2}\left[\frac{r_0^{(m)}}{r_0^{(n)}} A_k^{(n)} + \frac{r_0^{(n)}}{r_0^{(m)}} A_k^{(m)}\right] \\ &+ \frac{r_1^{(m)}}{r_0^{(n)}} B_k^{(n)} + \frac{r_1^{(n)}}{r_0^{(m)}} B_k^{(m)}, \end{aligned} \tag{15}$$

where $A_k^{(l)} = r_0^{(l)} \sum_{i=1}^{k} \left(Q_k^{(l)}\right)^{-1}_{i,i}$, $l = m, n$, satisfies when $|b_1^{(l)}| \neq 1$:

$$A_k^{(l)} \underset{(3)}{=} \frac{\left(1 + b_1^{(l)2}\right)\left(k\left(1 + b_1^{(l)2k+2}\right) - \frac{2b_1^{(l)2}\left(1 - b_1^{(l)2k}\right)}{\left(1 - b_1^{(l)2}\right)}\right)}{\left(1 - b_1^{(l)2}\right)\left(1 - b_1^{(l)2k+2}\right)}, \tag{16}$$

and when $|b_1^{(l)}| = 1$:

$$A_k^{(l)} \underset{(4)}{=} \frac{k(k+2)}{3}. \tag{17}$$

It should be noted that $A_k^{(l)}$ only depends on $b_1^{(l)}$ and k. Replacing $b_1^{(l)}$ by its opposite or its inverse leads to the same value of $A_k^{(l)}$. According to (16) and (17), $A_1^{(l)} = 1$. Nevertheless, according to (11) and (15):

$$JD_{ll}(k) \underset{(11)}{=} 0 \underset{(15)}{=} -k + A_k^{(l)} + 2\frac{r_1^{(l)}}{r_0^{(l)}} B_k^{(l)}. \tag{18}$$

Therefore $B_k^{(l)} = r_0^{(l)} \sum_{i=1}^{k-1} \left(Q_k^{(l)}\right)^{-1}_{i,i+1}$ with $l = m, n$ can be expressed by using $A_k^{(l)}$ as follows:

$$B_k^{(l)} = \frac{1}{2}\frac{r_0^{(l)}}{r_1^{(l)}}\left[k - A_k^{(l)}\right]. \tag{19}$$

Combining (15) and (19) leads to:

$$\begin{aligned} JD_{mn}(k) = &k\left[-1 + \frac{1}{2}\left(\frac{r_1^{(m)}}{r_1^{(n)}} + \frac{r_1^{(n)}}{r_1^{(m)}}\right)\right] \\ &+ \frac{1}{2}\left[\left(\frac{r_0^{(m)}}{r_0^{(n)}} - \frac{r_1^{(m)}}{r_1^{(n)}}\right) A_k^{(n)} + \left(\frac{r_0^{(n)}}{r_0^{(m)}} - \frac{r_1^{(n)}}{r_1^{(m)}}\right) A_k^{(m)}\right]. \end{aligned} \tag{20}$$

If $k = 1$, (20) reduces to (13).

2) Symmetry properties, features and other comments: Given (15)-(20), some remarks and comments can be highlighted.

Remark 1: it should be noted that when $b_1^{(m)}$ and $b_1^{(n)}$ are replaced by their opposites in (20), the JD is unchanged. In addition, if $b_1^{(m)} = -b_1^{(n)}$, replacing the MA parameters by their inverses does not change the JD.

Remark 2: $JD_{mn}(k+1) - JD_{mn}(k)$ is not a constant and varies over time. By exploiting (20), one obtains the following expression:

$$\begin{aligned} JD_{mn}(k+1) - JD_{mn}(k) &= -1 + \frac{1}{2}\left(\frac{r_1^{(m)}}{r_1^{(n)}} + \frac{r_1^{(n)}}{r_1^{(m)}}\right) \\ &+ \frac{1}{2}\Big[\left(\frac{r_0^{(m)}}{r_0^{(n)}} - \frac{r_1^{(m)}}{r_1^{(n)}}\right)\left(A_{k+1}^{(n)} - A_k^{(n)}\right) \\ &+ \left(\frac{r_0^{(n)}}{r_0^{(m)}} - \frac{r_1^{(n)}}{r_1^{(m)}}\right)\left(A_{k+1}^{(m)} - A_k^{(m)}\right)\Big]. \end{aligned} \tag{21}$$

In (21), the dependancy of the JD increment over time only depends on $A_{k+1}^{(l)} - A_k^{(l)}$, with $l = m, n$. Furthermore, when k is high enough, according to (16) and (17), $A_{k+1}^{(l)} - A_k^{(l)}$ becomes:
when $|b_1^{(l)}| \neq 1$:

$$A_{k+1}^{(l)} - A_k^{(l)} \approx E_k^{(l)} = \frac{1 + {b_1^{(l)}}^2}{|1 - {b_1^{(l)}}^2|}, \tag{22}$$

and when $|b_1^{(l)}| = 1$:

$$A_{k+1}^{(l)} - A_k^{(l)} = E_k^{(l)} = \frac{2k+3}{3}. \tag{23}$$

Thus, $JD_{mn}(k+1) - JD_{mn}(k)$ can be approximated by the following expression:

$$\begin{aligned} &JD_{mn}(k+1) - JD_{mn}(k) \approx \Delta JD_{mn}(k) \\ &= -1 + \frac{1}{2}\left(\frac{r_1^{(m)}}{r_1^{(n)}} + \frac{r_1^{(n)}}{r_1^{(m)}}\right) \\ &+ \frac{1}{2}\left[\left(\frac{r_0^{(m)}}{r_0^{(n)}} - \frac{r_1^{(m)}}{r_1^{(n)}}\right)E_k^{(n)} + \left(\frac{r_0^{(n)}}{r_0^{(m)}} - \frac{r_1^{(n)}}{r_1^{(m)}}\right)E_k^{(m)}\right]. \end{aligned} \tag{24}$$

If both $|b_1^{(m)}| \neq 1$ and $|b_1^{(n)}| \neq 1$, then $\Delta JD_{mn}(k)$ is constant and does not vary over the time k. In this case, (24) leads to:

$$JD_{mn}(k) \approx JD_{mn}(k_0) + (k - k_0)\Delta JD_{mn}, \tag{25}$$

with k_0 the instant from which $JD_{mn}(k_0+1) - JD_{mn}(k_0)$ is approximatively equal to ΔJD_{mn}. When $k \to +\infty$, the following approximation can be done:

$$\lim_{k \to +\infty} JD_{mn}(k) = k\Delta JD_{mn}. \tag{26}$$

The behaviour of the JD is closely related to the increment when k is high. This provides some asymptotical properties when studying the JD between two MA models.

Remark 3: When the models $^{(m)}$ and $^{(n)}$ satisfy (5), *i.e.* they have inverse MA parameters or equivalently inverse zeros, combining (6) and (20) leads to:

$$JD_{mn}(k) = k\left[-1 + \frac{1}{2}\left({b_1^{(m)}}^2 \frac{\sigma_{u^{(m)}}^2}{\sigma_{u^{(n)}}^2} + \frac{1}{b_1^{(m)}}^2 \frac{\sigma_{u^{(n)}}^2}{\sigma_{u^{(m)}}^2}\right)\right]. \tag{27}$$

In addition, if the driving-process-variance property (7) is satisfied, then:

$$JD_{mn}(k) = 0. \tag{28}$$

In this case, although the MA parameters are not the same, the PSDs are the same and the JD between both models is equal to zero.

C. Rao distance between two 1^{st}-order MA models

1) Detailed Rao distance expression: The square of the Rao distance between the $k \times k$ correlation matrices of two MA models $^{(m)}$ and $^{(n)}$ is given by:

$$d_{mn}^2(k) = \mathrm{Tr}\left[\left(\ln\left(\left(Q_k^{(m)}\right)^{-1/2} Q_k^{(n)} \left(Q_k^{(m)}\right)^{-1/2}\right)\right)^2\right]. \tag{29}$$

In the above equation, the correlation matrices can be rewritten by using their diagonalized forms.

According to [15], the h^{th} eigenvalue of a tridiagonal Toeplitz matrix is given for $l = m, n$ by:

$$\lambda_h^{(l)} = r_0^{(l)} + 2 \cdot r_1^{(l)} \cdot \cos\left(\frac{h\pi}{k+1}\right), \tag{30}$$

with $h = 1, \ldots, k$. The corresponding eigenvector $\mathbf{v}_h$ does not depend on the model and satisfies:

$$\begin{aligned} \mathbf{v}_h &= [v_{h,1}, \ldots, v_{h,k}]^T \\ &= \left[\sin\left(\frac{h\pi}{k+1}\right), \sin\left(\frac{2h\pi}{k+1}\right), \ldots, \sin\left(\frac{hk\pi}{k+1}\right)\right]^T. \end{aligned} \tag{31}$$

Then, given (30) and (31), the Rao distance (29) can be expressed as follows:

$$d_{mn}^2(k) = \sum_{i=1}^{k}\sum_{j=1}^{k} \ln^2\left(\frac{\lambda_j^{(n)}}{\lambda_j^{(m)}}\right) \cdot \frac{v_{j,i}v_{i,j}}{||\mathbf{v}_i|| \cdot ||\mathbf{v}_j||}. \tag{32}$$

By developping and simplifying (32), this leads to:

$$d_{mn}^2(k) = \sum_{j=1}^{k} \ln^2\left(\frac{\lambda_j^{(n)}}{\lambda_j^{(m)}}\right). \tag{33}$$

At this stage, by combining (30) and (33), one obtains:

$$\begin{aligned} d_{mn}^2(k) &= k\ln^2\left(\frac{r_0^{(n)}}{r_0^{(m)}}\right) \\ &+ 2\ln\left(\frac{r_0^{(n)}}{r_0^{(m)}}\right)\sum_{h=1}^{k}\ln\left(\frac{1 + \frac{2r_1^{(n)}}{r_0^{(n)}}\cos\left(\frac{h\pi}{k+1}\right)}{1 + \frac{2r_1^{(m)}}{r_0^{(m)}}\cos\left(\frac{h\pi}{k+1}\right)}\right) \\ &+ \sum_{h=1}^{k}\ln^2\left(\frac{1 + \frac{2r_1^{(n)}}{r_0^{(n)}}\cos\left(\frac{h\pi}{k+1}\right)}{1 + \frac{2r_1^{(m)}}{r_0^{(m)}}\cos\left(\frac{h\pi}{k+1}\right)}\right). \end{aligned} \tag{34}$$

Considering (30) and by using Taylor series expansion, (34) becomes:

$$\begin{aligned} d_{mn}^2(k) &= k\ln^2\left(\frac{r_0^{(n)}}{r_0^{(m)}}\right) \\ &+ 2\ln\left(\frac{r_0^{(n)}}{r_0^{(m)}}\right)\left[\sum_{p=1}^{+\infty}\frac{-1}{2p}\left[\left(\frac{2r_1^{(n)}}{r_0^{(n)}}\right)^{2p} - \left(\frac{2r_1^{(m)}}{r_0^{(m)}}\right)^{2p}\right]S_{2p,k}\right] \\ &+ \sum_{p_1=1}^{+\infty}\sum_{p_2=1}^{+\infty}\frac{T_{2p_1,2p_2}^{mn}S_{2p_1+2p_2,k}}{4p_1p_2} \\ &+ \sum_{p_1=0}^{+\infty}\sum_{p_2=0}^{+\infty}\frac{T_{2p_1+1,2p_2+1}^{mn}S_{2p_1+2p_2+2,k}}{(2p_1+1)(2p_2+1)}, \end{aligned} \tag{35}$$

with:

$$T_{a,b}^{mn} = \left[\frac{2r_1^{(n)}}{r_0^{(n)}}\right]^{a+b} + \left[\frac{2r_1^{(m)}}{r_0^{(m)}}\right]^{a+b} - 2\left[\frac{2r_1^{(n)}}{r_0^{(n)}}\right]^{a}\left[\frac{2r_1^{(m)}}{r_0^{(m)}}\right]^{b},$$

$$S_{2a,k} = \sum_{h=1}^{k} \cos^{2a}\left(\frac{h\pi}{k+1}\right).$$

By using the binomial theorem, it can be shown that:

$$S_{2a,k} = \left[-1 + \frac{k+1}{2^{2a}} \sum_{\beta=0}^{\beta_m} \binom{2a}{a-\beta(k+1)}\right],$$

where β_m is equal to $\left\lfloor \frac{a}{(k+1)} \right\rfloor$.

2) Symmetry, features and other comments:

Remark 4: once again, it should be noted that when $b_1^{(m)}$ and $b_1^{(n)}$ are replaced by their opposites in (35), the square of the Rao distance is the same.

Remark 5: if k is high enough, when $a \leq k$, $\beta_m = 0$:

$$S_{2a,k} = \left[-1 + \frac{k+1}{2^{2a}} \binom{2a}{a}\right]. \quad (36)$$

Otherwise when a and k becomes high:

$$S_{2a,k} \approx \left[-1 + \frac{k+1}{2^{2a}} \binom{2a}{a}\right]. \quad (37)$$

Finally, by exploiting (35), (36) and (37):

$$d_{mn}^2(k+1) - d_{mn}^2(k) \approx \Delta d_{mn}^2, \quad (38)$$

$$\text{with } \Delta d_{mn}^2 = \ln^2\left(\frac{r_0^{(n)}}{r_0^{(m)}}\right) \quad (39)$$

$$+ 2\ln\left(\frac{r_0^{(n)}}{r_0^{(m)}}\right)\left[\sum_{p=1}^{+\infty} \frac{-(2p)!}{2p\, p!\, p!}\left[\left(\frac{r_1^{(n)}}{r_0^{(n)}}\right)^{2p} - \left(\frac{r_1^{(m)}}{r_0^{(m)}}\right)^{2p}\right]\right]$$

$$+ \sum_{p_1=1}^{+\infty}\sum_{p_2=1}^{+\infty} \frac{T_{2p_1,2p_2}^{mn}(2p_1+2p_2)!}{4p_1p_2\left((p_1+p_2)!\right)^2 2^{2p_1+2p_2}}$$

$$+ \sum_{p_1=0}^{+\infty}\sum_{p_2=0}^{+\infty} \frac{T_{2p_1+1,2p_2+1}^{mn}(2p_1+2p_2+2)!}{(2p_1+1)(2p_2+1)\left((p_1+p_2+1)!\right)^2 2^{2p_1+2p_2+2}}.$$

Also, the square of the Rao distance can be approximated by a linear function of this increment:

$$d_{mn}^2(k) \approx d_{mn}^2(k_1) + (k-k_1)\Delta d_{mn}^2, \quad (40)$$

with k_1 the instant from which $d_{mn}^2(k_1+1) - d_{mn}^2(k_1) \approx \Delta d_{mn}^2$. When $k \to +\infty$, the Rao distance can be approximated by the following equation:

$$\lim_{k\to+\infty} d_{mn}^2(k) = k\Delta d_{mn}^2. \quad (41)$$

According to the above expressions (40)-(41), the increment study gives some indications about the behaviour of the Rao distance, especially when k is getting higher and higher.

D. Links and differences between the JD and the Rao distance

In many applications, the Rao distance is deduced from the JD. Indeed, for distributions that lie infinitesimally close on the probabilistic manifold, the square of the Rao distance is approximatively twice the JD between the distributions [11]:

$$\frac{d_{mn}^2(k)}{JD_{mn}(k)} \approx 2. \quad (42)$$

In the next section, we are going to analyze by simulations if (25) and (40) are confirmed and when the above theoretical property (42) is not satisfied.

III. Comparison of Dissimilarity Measures

A. A specific case

Firstly, let us confirm the way the JD and the Rao distance evolve when k increases. Fig. 1 illustrates the equations (25) and (40) when the 1st MA parameters are equal to 0.5 and 0.7.

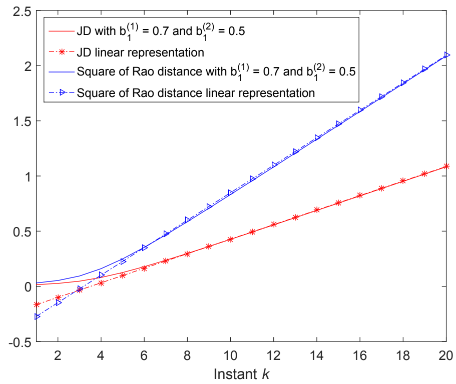


Fig. 1. JD, square of Rao distance and their respective increments

B. General case

Let us now consider the 1st MA parameters of both models vary in the interval $[-3, 3]$, $\sigma_{u(1)}^2 = \sigma_{u(2)}^2 = 3$ and $k = 25$. The Rao distance and the JD are firstly compared with the Itakura-Saito divergence (IS) defined as follows:

$$D_{IS21} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\sigma_{u(2)}^2 |H^{(2)}(z)|^2_{|z=e^{j\theta}}}{\sigma_{u(1)}^2 |H^{(1)}(z)|^2_{|z=e^{j\theta}}} - \ln\left(\frac{\sigma_{u(2)}^2 |H^{(2)}(z)|^2_{|z=e^{j\theta}}}{\sigma_{u(1)}^2 |H^{(1)}(z)|^2_{|z=e^{j\theta}}}\right) - 1\right] d\theta. \quad (43)$$

An approximation of the symmetric IS, based on the rectangle method, is computed on the Fig. 2.

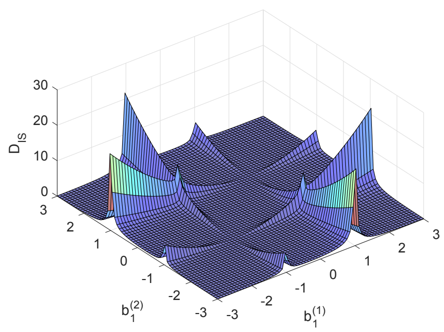


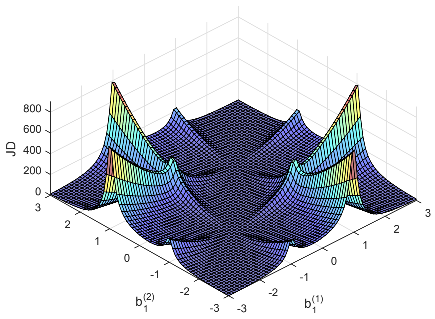
Fig. 2. Symmetric Itakura-Saito Divergence with $b_1^{(l)} \in [-3,\ 3]$, $l = 1, 2$

By looking at Fig. 3 (a) and (b) and also comparing them, some comments can be drawn:

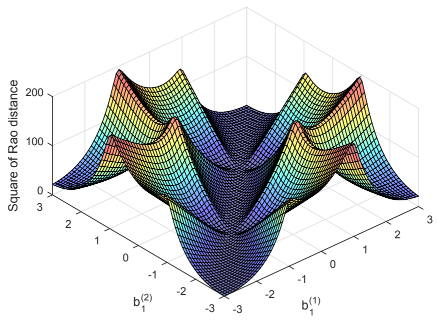
1) The remarks 1 and 4 are confirmed because a center of symmetry $(0, 0)$ can be observed in the figures. In addition, as the JD and the Rao distance are symmetric, the axis $b_1^{(1)} = b_1^{(2)}$ is a symmetric axis on both figures. As a consequence, a fourth of the figure defined by $|b_1^{(1)}| \leq b_1^{(2)}$ and $0 \leq b_1^{(2)} \leq 3$ is enough to retrieve the whole representation.

2) It can be shown from (20) and (35) that if $k > 1$:

- when $|b_1^{(l)}| \leq 1$, for $l = 1$ or 2, the maxima of both the JD and the Rao distance are obtained for $(-1, 1)$ and $(1, -1)$.
- when the absolute value of one of the 1st MA parameters is greater



(a) JD when $k = 25$

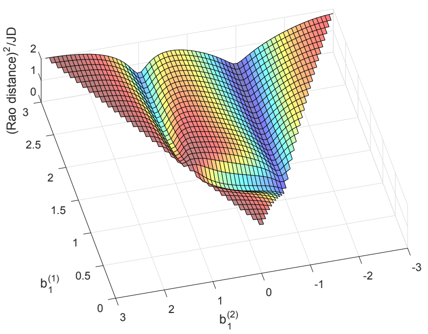


(b) Square of the Rao distance when $k = 25$

Fig. 3. JD and square of Rao distance when $\{b_1^{(l)}\}_{l=1,2}$ vary in $[-3,\ 3]$

than 1, the maximum and the local maxima are reached when the absolute value of the other 1st MA parameter is equal to 1.

In addition, the JD has a greater range of values than the square of the Rao distance and its range grows with k due to (23). Due to (16), (17) and (20), when the 1st MA parameter of one model tends to ± 1, a small variation of this 1st MA parameter leads to a large variation on the JD value whatever the value of the 1st MA parameter of the other model. This sensitivity is all the greater as k grows. Fig. 3 (b) describing the Rao distance for various 1st MA parameters is less spiky than Fig. 3 (a) describing the JD. The above phenomenon explains why the ratio between the square of the Rao distance and the JD is not always close to 2.



$k = 25$

Fig. 4. Ratio between the square of the Rao distance and the JD

Therefore, as depicted in Fig. 4, (42) is not necessarily satisfied. When comparing two 1st-order MA models, the JD is rather of real interest when the absolute value of the 1st MA parameter of one model is close[2] to 1. However, due to its large range of values, the JD could be less attractive in other cases, especially when comparing more than two models. Therefore, the Rao distance would be more suited.

IV. Conclusions and perspectives

In this paper, we have suggested using the JD and the square of the Rao distance to compare two real 1st-order MA models. In most of the cases, both dissimilarity measures can be approximated by affine functions depending on time k. Their slopes depend on the parameters of both MA models. In addition, the ratio between the JD and the square of the Rao distance is often close to 2, but the main difference between both measures appears when the 1st MA parameter of one model is equal to ± 1. In this case, we show that the JD has a different behaviour. The next step is to use the JD and the Rao distance to compare more than two 1st-order MA models by using an approach similar to the one presented in [16]. In addition, we are going to investigate the comparison of complex 1st-order MA models.

References

[1] W. Bobillet, R. Diversi, E. Grivel, R. Guidorzi, M. Najim, and U. Soverini, "Speech enhancement combining optimal smoothing and errors-in-variables identification of noisy AR processes," *IEEE Trans. on Signal Processing*, vol. 55, n12, pp. 5564–5578, December 2007.

[2] C. Magnant, A. Giremus, and E. Grivel, "On computing Jeffrey's divergence between time-varying autoregressive models," *Signal Processing Letters*, vol. 22, issue 7, pp. 915–919, 2014.

[3] C. Magnant, E. Grivel, A. Giremus, B. Joseph, and L. Ratton, "Jeffrey's divergence for state-space model comparison," *Signal Processing*, vol. 114, pp. 61–74, September 2015.

[4] K. Abou-Moustafa and F. P. Ferrie, "A note on metric properties for some divergence measures: the Gaussian case," in *JMLR workshop and conference proceedings*, vol. 25, pp. 1–15, 2012.

[5] J. Lapuyade-Lahorgue and F. Barbaresco, "Radar detection using Siegel distance between autoregressive processes, application to HF and X-band radar," in *IEEE Radar conference*, (Rome, Italy), pp. 1–6, 2008.

[6] L. Yang, M. Arnaudon, and F. Barbaresco, "Riemannian median, geometry of covariance matrices and radar target detection," in *European Radar Conference*, (Paris, France), pp. 415–418, September 2010.

[7] J. Pastore, E. Moler, and V. Ballarin, "Segmentation of brain magnetic resonance images through morphological operators and geodesic distance," *Digital Signal Processing*, vol. 15, No. 2, pp. 153–160, 2005.

[8] Q. Yong and Y. Jie, "Modified kernel functions by geodesic distance," *EURASIP Journal on Applied Signal Processing*, vol. 16, 2004.

[9] P. Formont, J. Ovarlez, F. Pascal, and G. Vasile, "On the extension of the product model in POLSAR processing for unsupervised classification using information geometry of covariance matrices," *IEEE International Geoscience and Remote Sensing Symposium*, pp. 1361–1364, 2011.

[10] C. Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, vol. 37. Bull. Calcutta Math. Soc., 1945.

[11] S. Kullback, *Information Theory and Statistics*. New York: Dover Publications, 1968.

[12] B. C. Sutradhar and P. Kumar, "The inversion of correlation matrix for MA(1) process," *Applied Mathematics Letters*, vol. 16, issue 3, pp. 317–321, April 2003.

[13] R. A. Usmani, "Inversion of Jacobi's tridiagonal matrix," in *Computers and Mathematics with Applications*, vol. 27, No 8, pp. 59–66, 1994.

[14] P. Colwell, *Blaschke Products: Bounded Analytical Functions*. University of Michigan Press, 1985.

[15] S. Noschese, L. Pasquini, and L. Reichel, "Tridiagonal Toeplitz matrices: properties and novel applications," *Numerical linear algebra with applications*, vol. 20, issue 2, pp. 302–326, March 2013.

[16] C. Magnant, E. Grivel, A. Giremus, B. Joseph, and L. Ratton, "Classifying several state models using Jeffrey's divergence: Application to target tracking," *IEEE International Radar Conference*, pp. 1–6, 2014.

[2] *i.e.* where some angular frequencies (0 or π) tend to lack in the spectrum