JEFFREY'S DIVERGENCE BETWEEN MOVING-AVERAGE AND AUTOREGRESSIVE MODELS

Léo Legrand[1],[2] *and Éric Grivel*[2]

[1]Thales Systèmes Aéroportés S.A., Pessac, FRANCE
[2]University of Bordeaux - Bordeaux INP - IMS - UMR CNRS 5218, Talence, FRANCE

ABSTRACT

This paper deals with model comparison based on the Jeffrey's divergence (JD). More particularly, after providing the JD between the joint distributions of k consecutive values of a white noise and the ones of a real moving-average or autoregressive model, the JD between real 1st-order MA and real 1st-order AR models is studied. Except when the 1st MA parameter is equal to 1, we show that, after a transient period, the JD between both models is incremented by a constant value that depends on the model parameters while k is incremented by 1. The JD is hence characterized by this increment and it is not necessary to consider a lot of samples.

Index Terms— Jeffrey's divergence, moving-average, autoregressive, model comparison

1. INTRODUCTION

Model comparison can be useful, more particularly in applications when classification is required. For instance, in target tracking based on Bayesian approaches, the motion model is usually *a priori* defined and selected among models such as the constant-velocity model, Singer model, *etc*. As one model may not be sufficient, multiple-model based methods can be used where a set of two or three dissimilar motion models are considered [1]. Therefore, classifying motion models in a preliminary step is of interest. In biomedical applications, the data can be represented by stochastic models that are then compared to distinguish healthy patients from people having a pathology. Concerning other applications such as the internet of things (IoT) where various sensors record different data, it can be of interest to compare the models representing the data to analyze if there are some dissimilarities or not.

Thus, when dealing with autoregressive (AR) models, model comparison can be based on the Itakura measure especially used in speech processing, the 2-norm of the difference between the two AR-parameter vectors or the spectral distance such as the log-spectral distance or the Itakura-Saito divergence. Divergences measuring the similarity between sample distributions can be also considered. In [2], Jeffrey's divergence (JD), which is the symmetric Kullback-Leibler (KL) divergence, is computed between the distributions of the successive samples of two time-varying AR (TVAR) models. The approach has been also extended to classify more than two AR models [3] or motion models [4] in various model subsets. As an alternative, metrics in the information geometry can be seen as dissimilarity measures. The reader may refer to Barbaresco's work *et al.* where the information geometry of AR-model covariance matrices [5] [6] is studied. As for the AR case, moving-average (MA) models can be compared by using the 2-norm between the MA-parameter vectors or the spectral distances. In [7], we give the exact analytical expressions of the JD between 1st-order MA models, for any MA parameter and any number of samples. Moreover, the MA models can be real or complex, noise-free or disturbed by additive white Gaussian noises.

To our knowledge, there is no work dealing with the comparison between 1st-order MA and AR models. As an alternative to a spectral distance or the model-parameter vector comparison[1], we propose to analyze the JD between a real 1st-order MA model and a real 1st-order AR model. Analytical expressions, properties, comments and various examples are given in this paper.

The remainder of this paper is organized as follows: in section 2, definitions and properties about real 1st-order MA and AR models are recalled. In section 3, JD between a white noise and a MA or AR model are first presented. They correspond to specific cases. Then, the JD between MA and AR models is studied. Simulation results are then presented and illustrate the theoretical part.

In the following, I_k is the identity matrix of size k, J_1 the $k \times k$ upper shift matrix and e_i a vector of size $k \times 1$ full of zeros except for the i^{th} component which is equal to 1. $(Q)_{g,h}$ is the element of the matrix Q at the g^{th} row and the h^{th} column. Tr denotes the trace of a matrix while the upper-script ^{T} is the transpose. det defines the determinant and $x_{k_1:k_2}$ is the collection of samples from time k_1 to k_2.

2. ABOUT REAL 1^{ST}-ORDER AR AND MA MODELS

Firstly, let x_k be a 1st-order AR model defined as follows:

$$x_k = -a_1 x_{k-1} + u_k, \tag{1}$$

where a_1 is the AR parameter and the driving process u_k is a zero-mean Gaussian white noise with variance σ_u^2. In this case, the correlation function $r_\tau^{(x)}$ satisfies $r_\tau^{(x)} = \frac{(-a_1)^{|\tau|}}{1-a_1^2}\sigma_u^2$. For $k \geq 2$, the $k \times k$ correlation matrix is defined by:

This work is part of the common research activities between Bordeaux campus and Thales (GIS Albatros).

[1]A 1st-order AR model is an infinte-order MA model (and conversely). If a_1 is the AR parameter, the corresponding infinite-order MA model is defined by the infinite length MA parameter vector: $[1, -a_1, a_1^2, -a_1^3, ...]$. Considering a 1st-order MA model represented by the parameter vector $[1, b_1, 0, ...]$, the difference vector between the two parameter vectors of the MA models is $[0, -a_1 - b_1, a_1^2, -a_1^3, ...]$. The square of the norm of the difference vector is then equal to $(a_1 + b_1)^2 + a_1^4/(1 - a_1^2)$.

$$Q_k^{(x)} = \frac{\sigma_u^2}{1-a_1^2}\begin{bmatrix} 1 & -a_1 & \dots & (-a_1)^{k-1} \\ -a_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -a_1 \\ (-a_1)^{k-1} & \dots & -a_1 & 1 \end{bmatrix}, \quad (2)$$

and its inverse is [8]:

$$\left(Q_k^{(x)}\right)^{-1} = \frac{1}{\sigma_u^2}\left[(1+a_1^2)I_k + a_1(J_1 + J_{-1}) - (1+a_1^2)[e_{1,k}e_{1,k}^T + e_{k,k}e_{k,k}^T]\right]. \quad (3)$$

Secondly, let y_k be a 1st-order MA model defined by:

$$y_k = v_k + b_1 v_{k-1}, \quad (4)$$

where b_1 is the MA parameter. The driving process v_k is a zero-mean Gaussian white noise with variance σ_v^2, uncorrelated with u_k. The correlation function $r_\tau^{(y)}$ satisfies:

$$r_\tau^{(y)} = \begin{cases} (1+b_1^2)\sigma_v^2 & \text{for } \tau = 0, \\ b_1\sigma_v^2 & \text{for } \tau = \pm 1, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

For $k \geq 2$, the $k \times k$ correlation matrix is defined by:

$$Q_k^{(y)} = \sigma_v^2\left[(1+b_1^2)I_k + b_1(J_1 + J_{-1})\right]. \quad (6)$$

The element at the i^{th} row and the j^{th} column of the inverse of $Q_k^{(y)}$ is given by [9]:

- for $|b_1| \neq 1$:

$$\left(Q_k^{(y)}\right)^{-1}_{ij} = \frac{1+b_1^2}{r_0^{(y)}(1-b_1^2)}\Bigg[(-b_1)^{|i-j|} - (-b_1)^{2k-i-j+2} - \frac{(-b_1)^{i+j}(1-b_1^{2k-2i+2})(1-b_1^{2k-2j+2})}{1-b_1^{2k+2}}\Bigg]. \quad (7)$$

- for $|b_1| = 1$:

$$\left(Q_k^{(y)}\right)^{-1}_{ij} = \frac{(-b_1)^{j-i}}{\sigma_v^2}\frac{i(k+1-j)}{k+1}. \quad (8)$$

<u>Remark</u>: the MA model (4) can be seen as the filtering of a white noise defined by its transfer function $H(z) = 1 + b_1 z^{-1}$. Let us now consider a second MA model defined by $\sigma^2_{v^{(\dagger)}}$, the MA parameter $1/b_1$ and the transfer function $H^{(\dagger)}(z) = 1 + \frac{1}{b_1}z^{-1}$. Then, considering $|b_1| < 1$, $H(z)$ can be rewritten as follows:

$$H(z) = \frac{1}{|b_1|}G(b_1, z^{-1})H^{(\dagger)}(z), \quad (9)$$

where $G(b_1, z^{-1}) = \frac{|b_1|}{-b_1}\frac{-b_1 - z^{-1}}{1+b_1 z^{-1}}$ is a Blaschke product [10] which can be seen as the transfer function of an all-pass filter. If θ denotes the normalized angular frequency, the corresponding power spectral densities (PSD) satisfy:

$$S_{yy}(\theta) = S_{yy}^{(\dagger)}(\theta)\frac{1}{{b_1}^2}\frac{\sigma^2_{v^{(\dagger)}}}{\sigma_v^2}. \quad (10)$$

The above PSDs are hence equal if:

$$\frac{\sigma^2_{v^{(\dagger)}}}{{b_1}^2} = \sigma_v^2. \quad (11)$$

<u>Remark</u>: when $|b_1| = 1$, the PSD is equal to 0 either at $\theta = 0$ or $\theta = \pm\pi$ whereas this property cannot be obtained with a 1st-order AR model.
We will see that this particular case is clearly identified when the JD is used. Given the above definitions and properties, let us address the study of the JD between the AR and MA models in the next section.

3. JEFFREY'S DIVERGENCE BETWEEN STOCHASTIC MODELS

We suggest analyzing the dissimilarities between the real AR and MA models by means of the JD between the joint distributions of k successive values of these models, denoted $p(x_{1:k})$ and $p(y_{1:k})$ respectively.
For this purpose, let us recall the KL divergence between two multivariate normal densities with means $\underline{\mu}_k^{(x)}$, $\underline{\mu}_k^{(y)}$ and covariance matrices $Q_k^{(y)}$ and $Q_k^{(y)}$ [11]:

$$KL_k{}^{(x,y)} = \frac{1}{2}\Bigg[\text{Tr}({Q_k^{(y)}}^{-1}Q_k^{(x)}) - k - \ln\frac{\det Q_k^{(x)}}{\det Q_k^{(y)}} + (\underline{\mu}_k^{(y)} - \underline{\mu}_k^{(x)})^T {Q_k^{(y)}}^{-1}(\underline{\mu}_k^{(y)} - \underline{\mu}_k^{(x)})\Bigg]. \quad (12)$$

The JD between $p(x_{1:k})$ and $p(y_{1:k})$ is deduced by symmetrizing the KL expression (12). For zero-mean Gaussian models, the JD becomes:

$$JD_k{}^{(x,y)} = -k + \frac{1}{2}\left[Tr({Q_k^{(y)}}^{-1}Q_k^{(x)}) + Tr({Q_k^{(x)}}^{-1}Q_k^{(y)})\right]. \quad (13)$$

In the following, as the AR and/or MA parameter, *i.e.* a_1 and b_1, can be equal to 0, our analysis starts by the trivial case corresponding to the JD between white noises. Then, the JDs between a MA or AR model and a white noise are presented. The last part deals with the JD between AR and MA models. Our work is hence complementary to the studies presented in [2] and [12].

3.1. Jeffrey's divergence for k samples of white noises

Let us first assume that $a_1 = b_1 = 0$. It is well-known that the above equation (13) becomes:

$$JD_k{}^{(WN,WN)} = k\left[-1 + \frac{1}{2}\left[\frac{\sigma_u^2}{\sigma_v^2} + \frac{\sigma_v^2}{\sigma_u^2}\right]\right]. \quad (14)$$

The JD depends on the ratio between the noise variances $\frac{\sigma_u^2}{\sigma_v^2}$ or equivalently the noise DSPs. The more dissimilar the variances are, the higher the JD is. It confirms what we could intuitively expect. In addition, as the JD is a linear function of the number of samples k, increasing the number of samples does not bring in anything to compare both models.

3.2. JD for k samples of a MA model and a white noise

Let us now assume that $a_1 = 0$ and $b_1 \neq 0$. Using (6)-(8) and (13), it can be shown that:

$$JD_k{}^{(MA,WN)} = -k + \frac{1}{2}\left[k(1+b_1^2)\frac{\sigma_v^2}{\sigma_u^2} + \frac{\sigma_u^2}{\sigma_v^2}\frac{A_k}{1+b_1^2}\right], \quad (15)$$

where for $k \geq 1$:

$$A_k = \begin{cases} \frac{(1+b_1^2)}{(1-b_1^2)}\frac{k(1+b_1^{2k+2}) - \frac{2b_1^2(1-b_1^{2k})}{1-b_1^2}}{1-b_1^{2k+2}} & \text{for } |b_1| \neq 1, \\ \frac{k(k+2)}{3} & \text{for } |b_1| = 1. \end{cases} \quad (16)$$

Given (16) and when k increases, the difference between two consecutive values of A_k reduces to:

$$A_{k+1} - A_k \approx \begin{cases} \frac{1+b_1^2}{|1-b_1^2|} & \text{for } |b_1| \neq 1, \\ \frac{2k+3}{3} & \text{for } |b_1| = 1. \end{cases} \tag{17}$$

Therefore, combining (15) and (17), one can deduce $\lim_{k\to+\infty}\left(JD_{k+1}{}^{(MA,WN)} - JD_k{}^{(MA,WN)}\right)$:

$$\lim_{k\to+\infty} \Delta JD_k{}^{(MA,WN)} = \tag{18}$$
$$\begin{cases} -1 + \frac{1}{2}\left[(1+b_1^2)\frac{\sigma_v^2}{\sigma_u^2} + \frac{\sigma_u^2}{\sigma_v^2}\frac{1}{|1-b_1^2|}\right] & \text{for } |b_1| \neq 1, \\ -1 + \frac{1}{2}\left[(1+b_1^2)\frac{\sigma_v^2}{\sigma_u^2} + \frac{\sigma_u^2}{\sigma_v^2}\frac{2k+3}{3(1+b_1^2)}\right] & \text{for } |b_1| = 1. \end{cases}$$

The difference between two successive JD tends to be a constant after a transient period when $|b_1| \neq 1$ whereas it depends on k if $|b_1| = 1$. In addition, the MA model defined by the MA parameter $1/b_1$ which satisfies (11) leads to the same result (18). In other words, if the PSDs of the MA models are the same, ΔJD_k is the same. Similarly, the spectral distance between the white noise and the MA model y_k or the MA model defined by $1/b_1$ and (11) would be the same.

3.3. JD between a white noise and an AR model

In this subsection, $b_1 = 0$ whereas $a_1 \neq 0$. Using (2), (3) and (13), it can be easily shown that:

$$JD_k{}^{(AR,WN)} = -k + \frac{1}{2}\left[\frac{\sigma_u^2}{\sigma_v^2}\frac{k}{(1-a_1^2)} + \frac{\sigma_v^2}{\sigma_u^2}\left[k(1+a_1^2) - 2a_1^2\right]\right]. \tag{19}$$

Therefore for any $k \geq 1$, the difference between two consecutive JD is a constant equal to:

$$\begin{aligned} \Delta JD_k{}^{(AR,WN)} &= JD_{k+1}{}^{(AR,WN)} - JD_k{}^{(AR,WN)} \\ &= -1 + \frac{1}{2}\left[\frac{\sigma_u^2}{\sigma_v^2}\frac{1}{(1-a_1^2)} + \frac{\sigma_v^2}{\sigma_u^2}(1+a_1^2)\right]. \end{aligned} \tag{20}$$

In this case, there is no transient period. The JD between both models can be directly characterized by the increment (20). The latter depends on the noise variance ratio and the AR parameter.

3.4. JD between an AR model and a MA model

Let us now assume that $b_1 \neq 0$ and $a_1 \neq 0$. Given (13), two traces must be computed. In the next two subsections, we analyze how they evolve when k increases.

3.4.1. Studying $T_k{}^{(MA,AR)} = Tr(Q_k^{(y)^{-1}} Q_k^{(x)})$

For $k = 1$, it can be easily shown that:

$$T_1{}^{(MA,AR)} = Q_1^{(y)^{-1}} Q_1^{(x)} = \frac{\sigma_u^2}{\sigma_v^2}\frac{1}{(1+b_1^2)(1-a_1^2)}. \tag{21}$$

For $k \geq 2$ and given the definitions (2) and (7) - (8) of $Q_k^{(x)}$ and $Q_k^{(y)^{-1}}$ respectively, $T_k{}^{(MA,AR)}$ satisfies:

$$\frac{\sigma_u^2}{(1-a_1^2)}\left(\sum_{i=1}^{k}(Q_k^{(y)})^{-1}_{i,i} + 2\sum_{l=1}^{k-1}(-a_1)^l\sum_{i=1}^{k-l}(Q_k^{(y)})^{-1}_{i,i+l}\right). \tag{22}$$

In the following, let us express (22) for different values of b_1:

Case 1: $|b_1| \neq 1$, the expression (22) of the trace $T_k{}^{(MA,AR)}$ becomes:

$$\begin{aligned} &\frac{\sigma_u^2}{\sigma_v^2(1-a_1^2)}\Bigg[\frac{k(1-b_1^{2k+4}) + (k+1)b_1^2(b_1^{2k}-1)}{(1-b_1^2)^2(1-b_1^{2k+2})} \\ &+ 2\sum_{l=1}^{k-1}\frac{(a_1b_1)^2}{(1-b_1^2)^2(1-b_1^{2k+2})}\big[(k-l)(1-b_1^{2k-2l+4}) \\ &+ (k-l+2)b_1^2(b_1^{2k-2l}-1)\big]\Bigg]. \end{aligned} \tag{23}$$

Given the above equation, let us now deduce $\lim_{k\to+\infty}\Delta T_k{}^{(MA,AR)}$ where $\Delta T_k{}^{(MA,AR)} = T_{k+1}{}^{(MA,AR)} - T_k{}^{(MA,AR)}$. Two cases must be considered:

- $|b_1| < 1$. Given (23) and after development and simplification, one has:

$$\begin{aligned} &\lim_{k\to+\infty}\Delta T_k{}^{(MA,AR)} = \\ &\frac{\sigma_u^2}{\sigma_v^2(1-a_1^2)}\left[\frac{1}{(1-b_1^2)} + 2\frac{a_1b_1}{(1-a_1b_1)(1-b_1^2)}\right]. \end{aligned} \tag{24}$$

- $|b_1| > 1$.

$$\begin{aligned} &\lim_{k\to+\infty}\Delta T_k{}^{(MA,AR)} = \\ &\frac{\sigma_u^2}{\sigma_v^2(1-a_1^2)}\left[\frac{1}{(b_1^2-1)} + 2\frac{a_1/b_1}{(1-a_1/b_1)(1-b_1^2)}\right]. \end{aligned} \tag{25}$$

Given (24) and (25), it should be noted that the MA model (4) and the one defined by the MA parameter $1/b_1$ which satisfies (11) lead to the same value for the asymptotic increment $\lim_{k\to+\infty}\Delta T_k^{(MA,AR)}$.

Case 2: $|b_1| = 1$, given (23), the increment is equal to:

$$\begin{aligned} \Delta T_k{}^{(MA,AR)} &= \frac{\sigma_u^2}{\sigma_v^2(1-a_1^2)}\Bigg[\frac{2k+3}{6} + \frac{2(a_1b_1)^k}{k+1} \\ &+ \frac{1}{3}\sum_{l=1}^{k}\frac{(a_1b_1)^l(k+1-l)(k+2-l)(2k+3+l)}{(k+1)(k+2)}\Bigg]. \end{aligned} \tag{26}$$

3.4.2. Studying $T_k{}^{(AR,MA)} = Tr(Q_k^{(x)^{-1}} Q_k^{(y)})$

For $k = 1$, it can be easily shown that:

$$Q_1^{(x)^{-1}} Q_1^{(y)} = \frac{\sigma_v^2}{\sigma_u^2}(1-a_1^2)(1+b_1^2). \tag{27}$$

For $k = 2$, one has:

$$Tr\left((Q_2^{(x)})^{-1}Q_2^{(y)}\right) = 2\frac{\sigma_v^2}{\sigma_u^2}(1+a_1b_1+b_1^2). \tag{28}$$

For $k > 2$, using the expressions (3) and (6) of $(Q_k^{(x)})^{-1}$ and $Q_k^{(y)}$, one has:

$$\begin{aligned} T_k{}^{(AR,MA)} &= Tr\left((Q_k^{(x)})^{-1}Q_k^{(y)}\right) \\ &= \frac{\sigma_v^2}{\sigma_u^2}\Big[2(k-1)a_1b_1 + k(1+a_1^2)(1+b_1^2) - 2a_1^2(1+b_1^2)\Big]. \end{aligned} \tag{29}$$

Therefore, for $k > 2$, the difference between two consecutive terms ${\Delta T_k}^{(AR,MA)}$ is equal to:

$$ {\Delta T_k}^{(AR,MA)} = \frac{\sigma_v^2}{\sigma_u^2}\left[2a_1 b_1 + (1+a_1^2)(1+b_1^2)\right]. \quad (30) $$

Given (30), the MA model (4) and the one defined by the MA parameter $1/b_1$ which satisfies (11) lead to the same value once again.

3.4.3. *Deducing the JD evolution when k increases*

To deduce the way the JD evolves, *i.e.* $\Delta JD_k^{(AR,MA)}$ which is equal to $JD_{k+1}^{(AR,MA)} - JD_k^{(AR,MA)}$, one first has to combine the results (30) with (25) or (26) obtained in the two above subsections. Indeed, due to (13), one has:

$$ {\Delta JD_k}^{(AR,MA)} = -1 + \frac{1}{2}\left[{\Delta T_k}^{(MA,AR)} + {\Delta T_k}^{(AR,MA)}\right]. \quad (31) $$

We can conclude that except when $|b_1| = 1$, the increment of the JD tends to remain contant when k increases and depends on the MA and AR parameters as well as the driving-process variances. In addition, given an AR model, the MA model (4) and the one defined by the MA parameter $1/b_1$ which satisfies (11) lead to the same JD increment and the same Itakura-Saito distance.

4. SIMULATION RESULTS

4.1. 1st protocol

Let us first illustrate the way the JD evolves for $a_1 = 0.3$, $b_1 = 1.1$, $\sigma_u^2 = 2$ and $\sigma_v^2 = 0.9$. This confirms the theoretical result obtained in section 3.4.

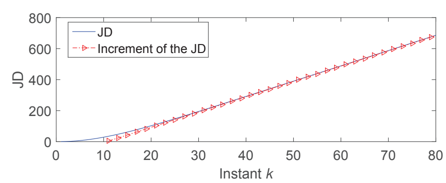

Fig. 1. Evolution of the JD over k when $a_1 = 0.3$, $\sigma_u^2 = 2, b_1 = 1.1, \sigma_v^2 = 0.9$.

4.2. 2nd protocol

Then, let us compute the JD for $k = 20$, $\sigma_u^2 = 2$ and $\sigma_v^2 = 1.1$. The AR parameter varies in the interval $]-1, 1[$ whereas the MA parameter lies in the interval $[-3, 3]$.

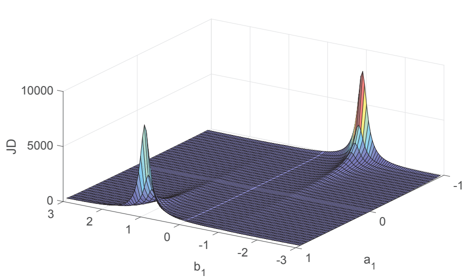

Fig. 2. Evolution of the JD when $k = 20$, $\sigma_u^2 = 2$, $\sigma_v^2 = 0.9$, a_1 varies in $]-1, 1[$ and b_1 varies in $[-3, 3]$.

According to Fig. 2, it can be seen that the JD is sensitive to the value of the MA parameter when the modulus of the latter is around 1.

The square of the norm of the difference vector between $[1, b_1, ...]$ and $[1, -a_1, a_1^2, ...]$ given by $(a_1+b_1)^2 + a_1^4/(1-a_1^2)$ is represented on Fig. 3.

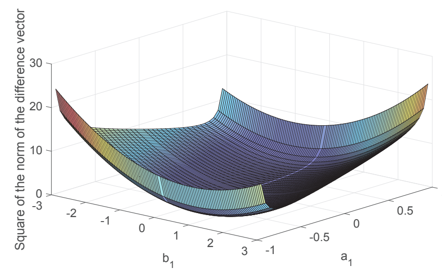

Fig. 3. Evolution of $(a_1 + b_1)^2 + a_1^4/(1 - a_1^2)$, when a_1 varies in $]-1, 1[$ and b_1 varies in $[-3, 3]$.

The log-spectral distance LSD can be approximated by the following discrete calculation made on N points of the power spectrum densities:

$$ LSD^{(AR,MA)} \approx \sqrt{\frac{1}{N}\sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1}\left[10\ln\frac{S_{xx}(2\pi\frac{n}{N})}{S_{yy}(2\pi\frac{n}{N})}\right]^2}. \quad (32) $$

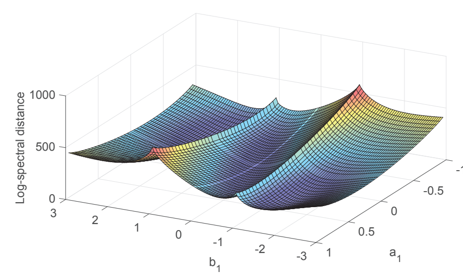

Fig. 4. Evolution of the log-spectral distance when $k = 20$, $\sigma_u^2 = 2$, $\sigma_v^2 = 0.9$, a_1 varies in $]-1, 1[$ and b_1 varies in $[-3, 3]$.

According to Fig. 4 and unlike the square of the model parameter vector difference, the log-spectral distance is also sensitive to the value of the MA parameter when the modulus of the latter is around 1 and is similar to the JD. However, it should be noted that the range of values of the log-spectral distance is far smaller than the range of values of the JD. The phenomenon of increasing sensitivity of the JD for $|b_1| = 1$ is pointed out in the theoretical part.

5. CONCLUSIONS AND PERSPECTIVES

In this paper, the JD between real 1^{st}-order AR and MA models is studied. We show that the derivative of the JD tends to be a constant when the number of samples increases. This phenomenon occurs except when the modulus of MA parameter is equal to 1. The JD has hence the advantage of emphasizing this particular case where the PSD associated with the MA model is null at a frequency whereas all the frequencies appear in the PSD of the AR model. We are currently studying the JD between noisy AR and MA models.

6. REFERENCES

[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications To Tracking and Navigation*, Wiley, Hoboken, NJ, USA, 2001.

[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, A. Giremus, and E. Grivel, "Jeffreys divergence between state models: Application to target tracking using multiple models," *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, pp. 1–5, September 2013.

[4] 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, October 2014.

[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, May 2008, pp. 1–6.

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

[7] L. Legrand and E. Grivel, "Jeffrey's divergence between moving-average models that are real or complex, noise-free or disturbed by additive white noises," *Elsevier Signal Processing*, vol. 131, pp. 350–363, 2017.

[8] B. Cernuschi-Frias, "A Derivation of the Gohberg-Semencul Relation," *IEEE Transactions on Signal Processing*, vol. 39, issue 1, pp. 190–192, 1991.

[9] 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.

[10] P. Colwell, *Blaschke products: bounded analytical functions*, University of Michigan Press, 1985.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, MIT Press, 2006.

[12] L. Legrand and E. Grivel, "Evaluating dissimilarities between two moving-average models: a comparative study between Jeffreys divergence and Rao distance," *European Signal Processing Conference (EUSIPCO)*, 2016.