Jeffrey's Divergence Between Autoregressive Moving-Average Processes

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

***Abstract*—Various works have been carried out about the Jeffrey's divergence (JD) which is the symmetric version of the Kullback-Leibler (KL) divergence. An expression of the JD for Gaussian processes can be deduced from the definition of the KL divergence and the expression of the Gaussian-multivariate distributions of k-dimensional random vectors. It depends on the $k \times k$ Toeplitz covariance matrices of the stationary processes. However, the resulting computational cost may be high as these matrices must be inverted and it is all the higher as k increases. To circumvent this problem, a recursive expression can be obtained for real 1^{st}-order autoregressive (AR) processes. When they are disturbed by additive uncorrelated white noises, we showed that when k becomes large, the derivative of the JD with respect to k tends to be constant. This constant is sufficient to compare the noisy AR processes. In this paper, we propose to extend our work to AR moving-average (MA) processes with one AR term and one MA term. Some examples illustrate the theoretical analysis.**

***Index Terms*—Jeffrey's divergence, Kullback-Leibler divergence, autoregressive moving-average processes.**

I. Introduction

Comparing stochastic processes such as autoregressive moving-average (ARMA) processes can be useful in many applications, especially when classification is required and when change detection is of interest. One intuitive approach would be to compare the process-parameter vectors by computing the 2-norm or the ∞-norm of the vector difference. Cepstral distance can be also useful. It makes it possible to distinguish EEG signals recorded a few minutes after an asphyxic cardiac arrest injury [1]. Another way is to compare the power spectra of the data. Thus, the spectral distortion between the natural and the synthetic speech can be based on the COSH distance which is defined from the power spectra of the data to be compared [2]. The log-spectral distance and the Itakura-Saito distance are also widely computed, especially in speech processing. As an alternative, divergences that aim at measuring the similarity between sample distributions can be also considered. In recent papers such as [3], the Jeffrey's divergence (JD), which is the symmetric Kullback-Leibler (KL) divergence, is recursively computed for the distributions of successive samples of two time-varying AR (TVAR) processes. The approach has been then extended to classify more than two AR processes in various subsets [4]. Concerning the moving-average (MA) processes, the analytical expression of the JD between 1st-order MA processes, that can be real or complex, noise-free or disturbed by additive white Gaussian noises, is given in [5].

In this paper, our purpose is to compare two real autoregressive moving-average (ARMA) processes with one AR term and one MA term by using the JD between the distributions of k successive samples of each process. As the correlation matrix of an ARMA process is similar to the one of a noisy AR process, we suggest comparing the ARMA processes by using the results we recently obtained about the JD between noisy AR processes. In the ARMA case, and *a fortiori* in the general case, interpreting a value of a JD is not so easy. It is true that the smaller the value is, the less dissimilar the processes are. Among the questions that can arise, the practitioner may wonder which value of k is relevant.

For these reasons, we propose to give an expression of the JD depending on the ARMA parameters. We study how the JD evolves when k increases and we show that after a transient period, the JD is incremented by the same value between two consecutive values of k. This phenomenon is always true except when the MA parameter of one or the two processes is equal to 1 or -1. This analysis can help the practitioner see the influences of the ARMA parameters on the JD.

This paper is organized as follows: in sections II and III, we briefly recall the main properties of the ARMA processes and the JD between 1st-order AR processes as well as the JD between two noisy AR processes. In section IV, the JD between ARMA processes is addressed. In section V, theoretical results are illustrated by some examples.

In the following, I_k is the identity matrix of size k. $(Q)_{g,h}$ denotes the element of the matrix Q which is at the g^{th} row and the h^{th} column. Tr is the trace of a matrix. The upperscript ^{T} denotes the transpose.

II. About real autoregressive moving-average processes

A. Definition and properties of an ARMA process

Let us define the ARMA process, $\text{ARMA}_{1,1}(a, b, \sigma^2)$, as follows:

$$x_k = -ax_{k-1} + u_k + bu_{k-1}, \tag{1}$$

where a and b are the process parameters whereas u is the zero-mean driving process with variance σ^2. It should be noted that a, b and σ^2 are used for the sake of simplicity instead of a_1, b_1 and σ_u^2. When $|b| = 1$, the power spectral density of the ARMA process is equal to 0 either at the normalized angular frequency $\theta = 0$ or $\theta = \pm\pi$. Moreover, $0 < |a| < 1$ to guarantee stability. If $a = 0$, the resulting process is a 1st-order MA process. If b is equal to 0, it reduces to a 1st-order AR process, denoted as $\text{AR}_1(a, \sigma^2)$, and the correlation function is equal to $r_{AR(a,\sigma^2)}(\tau) = \frac{(-a)^{|\tau|}}{1-a^2}\sigma^2$ with τ the time lag. When both a and b are equal to 0, x_k reduces to a white noise.

The correlation function of the ARMA process, denoted as $r_{xx}(\tau)$, satisfies:

$$\begin{cases} r_{xx}(0) = r_{AR(a,\sigma^2)}(0) + \frac{b}{1-a^2}(b-2a)\sigma^2, \\ r_{xx}(1) = r_{xx}(-1) = -ar_{AR(a,\sigma^2)}(0) + \frac{b}{1-a^2}(1+a^2-ab)\sigma^2, \\ r_{xx}(\tau) = r_{xx}(-\tau) = -ar_{xx}(\tau-1), \text{ otherwise.} \end{cases} \tag{2}$$

In the following, the Toeplitz correlation matrix of the vector storing k consecutive values of the $\text{ARMA}_{1,1}(a, b, \sigma^2)$ process is denoted as $Q_{ARMA(a,b,\sigma^2),k}$ whereas the one related to the process $\text{AR}_1(a, \sigma^2)$ is denoted as $Q_{AR(a,\sigma^2),k}$. If the process $\text{AR}_1(a, \sigma^2)$ is disturbed by an additive white zero-mean Gaussian noise n_k with variance σ_n^2 and uncorrelated with the driving process u_k, the Toeplitz correlation matrix of the noisy AR process is $Q_{NAR(a,\sigma^2,\sigma_n^2),k}$. Since there is no straightforward expression of the inverse of $Q_{ARMA(a,b,\sigma^2),k}$, it

ISBN 978-0-9928626-7-1 © EURASIP 2017

can be shown that it can take the form of the correlation matrix of a "noisy" AR as follows:

$$Q_{ARMA(a,b,\sigma^2),k} = \left[\frac{a-(1+a^2)b+ab^2}{a}\right] Q_{AR(a,\sigma^2),k} + \frac{b}{a}\sigma^2 I_k. \quad (3)$$

When $a = b$ or $a = \frac{1}{b}$, the gain $\left[\frac{a-(1+a^2)b+ab^2}{a}\right]$ is equal to 0. In this case, $Q_{ARMA(a,b,\sigma^2),k}$ reduces to the diagonal matrix $\frac{b}{a}\sigma^2 I_k$. Besides, when $\frac{b}{a}\sigma^2 > 0$, this perturbation corresponds to a zero-mean white noise sequence with variance $\frac{b}{a}\sigma^2$. More generally, this disturbance can be seen in the frequency domain as an offset equal to $\frac{b}{a}\sigma^2$ which can be positive or negative because b and a can take any value except zero (otherwise, this is no longer an ARMA process). In the following, we suggest rewriting (3) as follows:

$$Q_{ARMA(a,b,\sigma_u^2),k} = Q_{AR(a,\Psi),k} + \Upsilon I_k = Q_{NAR(a,\Psi,\Upsilon),k}, \quad (4)$$

where $\Psi = \left[\frac{a-(1+a^2)b+ab^2}{a}\right]\sigma^2$ and $\Upsilon = \frac{b}{a}\sigma^2$. It is true that they are not necessarily positive and hence cannot be considered properly as variances of white sequences. It should be noted that the AR parameter a is unchanged after this transformation.

B. Parameter estimation methods

In the noise-free or noisy case, both the AR parameters and the MA parameters as well as the driving-process variances can be predefined or estimated from a set of data. In this latter case and for the noise-free case, the AR parameters can be estimated in many ways: Yule-Walker (YW) equations, correlation method, adaptive filtering, etc. Concerning the MA parameter estimations, one can use the Durbin algorithm, the approach combining the inverse Fourier transform of the inverse of the MA power spectral density (PSD) and the YW equations [6], the "vocariance" methods [7] [8], the one based on higher-order statistics [9], the covariance fitting approaches [10] and the spectral factorization based on the estimation of the outer factor in the PSD such as [11].

In the noisy case, adaptive filters such as the γ-LMS [12] and the ρ-LMS [13] and extended Kalman filter can estimate the AR parameters from noisy observations. As an alternative, higher-order YW equations, iterative bias compensation schemes such as those proposed by [14], Davila's method [15] or errors-in-variables (EIV) approaches [16] [17] can be considered. In [18] and [19], the estimations of the ARMA parameters are addressed from noisy observations. More particularly, in [18], Fattah *et. al.* suggest combining approaches that were initially proposed by Davila [15] for the AR-parameter estimations and Stoica [10] for the MA-parameter estimations.

III. Definition of Jeffrey's divergence between stochastic processes

To analyze the dissimilarities between two random processes x and y, the JD between the joint distributions of k successive values of both processes can be computed. It is deduced by symmetrizing the KL expression as follows:

$$JD_k^{(x,y)} = \frac{1}{2}(KL_k^{(x,y)} + KL_k^{(y,x)}), \quad (5)$$

where $KL_k^{(x,y)}$ denotes the KL divergence between the multivariate densities p and q.

When dealing with Gaussian processes, the KL divergence between two real multivariate Gaussian densities p and q with means $\mu_{x,k}$ and $\mu_{y,k}$ and covariance matrices $Q_{x,k}$ and $Q_{y,k}$ [20] can be obtained by combining the definition of the KL and the expressions of the Gaussian distributions as follows:

$$KL_k{}^{(x,y)} = \frac{1}{2}\Big[\mathrm{Tr}(Q_{y,k}{}^{-1}Q_{x,k}) - k - \ln\frac{\det Q_{x,k}}{\det Q_{y,k}} \quad (6)$$

$$+ (\mu_{y,k} - \mu_{x,k})^T Q_{y,k}{}^{-1}(\mu_{y,k} - \mu_{x,k})\Big]. \quad (7)$$

Given (5) and (6), the JD becomes for zero-mean processes:

$$JD_k{}^{(x,y)} = -k + \frac{1}{2}\left[Tr(Q_{y,k}{}^{-1}Q_{x,k}) + Tr(Q_{x,k}{}^{-1}Q_{y,k})\right]. \quad (8)$$

In this case, the JD is only defined from the $k \times k$ covariance matrices of the processes. Nevertheless, these matrices must be inverted. When dealing with AR or MA processes, the resulting correlation matrix have specific structures. Various works have been done about the expressions of their inverses. See [21] and [22]. As a consequence, analytical expressions of the JD can be obtained. They depend on the process parameters, but it still implies multiplication of matrices of size $k \times k$. As an alternative, in [3] and [4], the JD between joint densities of k samples of AR processes, namely $\mathrm{AR}_1(a_{(1)}, \sigma^2_{(1)})$ and $\mathrm{AR}_2(a_{(2)}, \sigma^2_{(2)})$, can be deduced recursively as follows:

$$JD_{AR,k+1} = JD_{AR,k} + A + B, \quad (9)$$

with:

$$A = -1 + \frac{1}{2}(R_{ar} + \frac{1}{R_{ar}}), \quad (10)$$

and:

$$B = \frac{(a_{(2)} - a_{(1)})^2}{2}\left[\frac{1}{1-(a_{(1)})^2}\frac{1}{R_{ar}} + \frac{1}{1-(a_{(2)})^2}R_{ar}\right], \quad (11)$$

where the driving-process variance ratio is equal to $R_{ar} = \frac{\sigma^2_{(2)}}{\sigma^2_{(1)}}$ and A is the JD between the distributions of one sample of each driving process. The difference between two consecutive JD is hence a constant equal to $A + B$ and can be defined as follows:

$$\Delta JD_{AR} = JD_{AR,k+1} - JD_{AR,k} = A + B \quad (12)$$

This increment ΔJD_{AR} is sufficient to compare the AR processes. In a recent work, we suggested studying if the JD follows the same type of property when dealing with noisy AR processes, namely $\mathrm{NAR}_1(a_{(1)}, \sigma^2_{(1)}, \sigma^2_{n(1)})$ and $\mathrm{NAR}_2(a_{(2)}, \sigma^2_{(2)}, \sigma^2_{n(2)})$. For this reason, we proposed to deduce this JD denoted as $JD^{(1,2)}_{NAR,k}$ from the one computed when the processes are noise-free. More particularly, using the inversion matrix lemma [23] we showed that:

$$JD^{(1,2)}_{NAR,k} = JD^{(1,2)}_{AR,k} + \frac{C_k}{2}, \quad (13)$$

where the offset C_k due to the white additive noises is the sum of six terms:

$$D_{k,1} = \sigma^2_{n(2)}\,\mathrm{Tr}\left[Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k}\right], \quad (14)$$

$$E_{k,1} = -\sigma^2_{n(1)}\sigma^2_{n(2)}\,\mathrm{Tr}\Big[Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k} \Big(I_k + \sigma^2_{n(1)}Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k}\Big)^{-1} Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k}\Big], \quad (15)$$

$$F_{k,1} = -\sigma^2_{n(1)}\,\mathrm{Tr}\Big[Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k}\Big(I_k + \sigma^2_{n(1)}Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k}\Big)^{-1} Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k}Q_{AR(a_{(2)},\sigma^2_{(2)}),k}\Big], \quad (16)$$

and three other terms, denoted as $D_{k,2}$, $E_{k,2}$ and $F_{k,2}$ similarly defined as $D_{k,1}$, $E_{k,1}$ and $F_{k,1}$ but where the subscripts are switched. We showed that there is a transient period before the derivatives of these terms with respect to the number of variates k become constant.

- For $k \geq 2$, $\Delta D_1 = D_{k+1,1} - D_{k,1}$ satisfies:

$$\Delta D_1 = \frac{\sigma^2_{n(2)}}{\sigma^2_{(1)}}\left(1 + a_{(1)}{}^2\right). \quad (17)$$

ISBN 978-0-9928626-7-1

• When k becomes large, $\Delta E_1 = E_{k+1,1} - E_{k,1}$ is equal to:

$$\Delta E_1 = -\frac{\sigma^2_{n^{(2)}}}{\sigma^2_{(1)}} \frac{Z}{a_{(1)}\left(Z^2-1\right)} \Big[\left(1 + 4{a_{(1)}}^2 + {a_{(1)}}^4\right) + \left(4a_{(1)}\left(1+{a_{(1)}}^2\right)\right) Z + 2{a_{(1)}}^2 Z^2 \Big] \text{ for } 0 < a_{(1)} < 1, \quad (18)$$

and:

$$\Delta E_1 = \frac{\sigma^2_{n^{(2)}}}{\sigma^2_{(1)}} \frac{Z}{a_{(1)}\left(Z^2-1\right)} \Big[\left(1 + 4{a_{(1)}}^2 + {a_{(1)}}^4\right) + \left(4a_{(1)}\left(1+{a_{(1)}}^2\right)\right) Z^{-1} + 2{a_{(1)}}^2 Z^{-2} \Big] \text{ for } -1 < a_{(1)} < 0, \quad (19)$$

where[1]:

$$Z = \frac{1}{2}\left[-\left(\frac{1}{a_{(1)}}\left(1+\rho_{(1)}\right) + a_{(1)}\right) + \sqrt{\left(\frac{1}{a_{(1)}}\left(1+\rho_{(1)}\right) + a_{(1)}\right)^2 - 4}\right], \quad (20)$$

and where $\rho_{(1)}$ is the ratio between the variance of the driving process and the variance of the additive white noise:

$$\rho_{(1)} = \frac{\sigma^2_{(1)}}{\sigma^2_{n^{(1)}}} > 0. \quad (21)$$

• We can also deduce ΔF_1 in the asymptotical case. For this purpose, let us first introduce $S^{(1)}_k = (I_k + \sigma^2_{n^{(1)}} Q_{AR(a_{(1)},\sigma^2_{(1)}),k})$. One can show that:

$$\left(S^{(1)}_k\right)_{g,h} = -\frac{\sigma^2_{u^{(1)}}}{\sigma^2_{n^{(1)}}} \frac{Z^{h+1-g}}{a_{(1)}\left(Z^2-1\right)} \left(\left(Z^{2g}-1\right) + a_{(1)} Z\left(Z^{2g-2}-1\right)\right) \times \frac{\left(\left(Z^{2k+2-2h}-1\right) + a_{(1)} Z\left(Z^{2k-2h}-1\right)\right)}{Z^{2k+2} - 1 + 2a_{(1)} Z\left(Z^{2k}-1\right) + {a_{(1)}}^2 Z^2\left(Z^{2k-2}-1\right)}, \quad (22)$$

for $g \leq h$ and $\left(S^{(1)}_k\right)_{g,h} = \left(S^{(1)}_k\right)_{h,g}$ when $g \geq h$.
According to (22), the range of values taken by Z varies and depends on the value of the AR parameter $a_{(1)}$. This leads to:

$$\left(S^{(1)}_{lim}\right)_{g,h} = \lim_{k\to+\infty} \left(S^{(1)}_k\right)_{g,h} \quad (23)$$
$$= \begin{cases} -\frac{\sigma^2_{(1)}}{\sigma^2_{n^{(1)}}} \frac{Z^{h+1-g}}{a_{(1)}\left(Z^2-1\right)} \frac{\left(\left(Z^{2g}-1\right)+a_{(1)}Z\left(Z^{2g-2}-1\right)\right)}{\left(1+a_{(1)}Z\right)} \\ \text{when } -1 < Z < 0, \text{ i.e. for } 0 < a_{(1)} < 1, \\ -\frac{\sigma^2_{(1)}}{\sigma^2_{n^{(1)}}} \frac{Z^{-h+1-g}}{a_{(1)}\left(Z^2-1\right)} \frac{\left(\left(Z^{2g}-1\right)+a_{(1)}Z\left(Z^{2g-2}-1\right)\right)}{\left(1+a_{(1)}Z^{-1}\right)} \\ \text{when } Z > 1, \text{ i.e. for } -1 < a_{(1)} < 0. \end{cases}$$

It should be noted that $\left(Z^2-1\right)$ is necessarily different from 0 as well as $\left(1+a_{(1)}Z\right)$ or $\left(1+a_{(1)}Z^{-1}\right)$.
Now, let us consider $T^{(1)}_k = Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k} S^{(1)}_k Q^{-1}_{AR(a_{(1)},\sigma^2_{(1)}),k}$, where for $g = 2, ..., k-1$ and $h = 2, ..., k-1$, it can be shown that $\lim_{k\to+\infty} \left(T^{(1)}_k\right)_{g,h} = \left(T^{(1)}_{lim}\right)_{g,h}$:

$$\left(T^{(1)}_{lim}\right)_{g,h} = \frac{1}{\sigma^4_{u^{(1)}}} \Bigg[\left(1 + a^2_{(1)}\right)^2 \left(S^{(1)}_{lim}\right)_{g,h} + a_{(1)}\left(1+a^2_{(1)}\right) \Bigg(\left(S^{(1)}_{lim}\right)_{g+1,h} + \left(S^{(1)}_{lim}\right)_{g-1,h} + \left(S^{(1)}_{lim}\right)_{g,h+1} + \left(S^{(1)}_{lim}\right)_{g,h-1} \Bigg) + a^2_{(1)} \Bigg(\left(S^{(1)}_{lim}\right)_{g-1,h-1} + \left(S^{(1)}_{lim}\right)_{g+1,h+1} + \left(S^{(1)}_{lim}\right)_{g+1,h-1} + \left(S^{(1)}_{lim}\right)_{g-1,h+1} \Bigg) \Bigg], \quad (24)$$

When k becomes large, $\Delta F_1 = F_{k+1,1} - F_{k,1}$ is equal to:

$$\Delta F_1 = -\sigma^2_{n^{(1)}} \left[\frac{\sigma^2_{(2)}}{1 - {a_{(2)}}^2} \left(\left(T^{(1)}_{lim}\right)_{\frac{k}{2},\frac{k}{2}} + 2\sum_{i=1}^{\frac{k}{2}-1} \left(T^{(1)}_{lim}\right)_{\frac{k}{2},i} \left(-a_{(2)}\right)^{\left(\frac{k}{2}-i\right)} \right)\right], \quad (25)$$

By combining (17), (18) and (25), after a transient period, the difference between C_{k+1} and C_k does not depend on k and becomes a constant. Therefore, using (13) and when k becomes large, one has:

$$\lim_{k\to+\infty} JD_{NAR,k+1} - JD_{NAR,k} = \Delta JD_{NAR} = \Delta JD_{AR} + \frac{\Delta D_1 + \Delta E_1 + \Delta F_1 + \Delta D_2 + \Delta E_2 + \Delta F_2}{2} = Cte, \quad (26)$$

where the variations ΔD_2, ΔE_2 and ΔF_2 are similarly defined as ΔD_1, ΔE_1 and ΔF_1. In this case, the increment ΔJD_{NAR} was shown to be sufficient to compare noisy AR processes.
Finally, in [24], the MA case was also addressed. We showed that the JD increment tends to be a constant except when the MA-parameter modulus was equal to 1.
In the next section, let us study the JD between two ARMA processes.

IV. JD BETWEEN ARMA PROCESSES BASED ON THE ANALYSIS OF THE JD BETWEEN NOISY AR PROCESSES

In this section, we suggest addressing the analysis of the JD between ARMA processes by using the results we obtained for noisy AR processes.
The JD between the process ARMA$(a_{(1)}, b_{(1)}, \sigma^2_{(1)})$ and the process ARMA$(a_{(2)}, b_{(2)}, \sigma^2_{(2)})$ amounts to studying the JD between the process NAR$(a_{(1)},, \Psi_{(1)}, \Upsilon_{(1)})$ and the process NAR$(a_{(2)},, \Psi_{(2)}, \Upsilon_{(2)})$ where $\Psi_{(l)}$ and $\Upsilon_{(l)}$ with $l = 1, 2$ are defined as in (4). In the following, our purpose is to express the increment as we did in (26):

$$\lim_{k\to+\infty} JD_{ARMA,k+1} - JD_{ARMA,k} = \Delta JD_{ARMA} = \Delta JD_{AR} + \frac{\Delta D_1 + \Delta E_1 + \Delta F_1 + \Delta D_2 + \Delta E_2 + \Delta F_2}{2}. \quad (27)$$

In (4), the AR parameter remains unchanged in the mapping between the ARMA and the NAR process but the driving-process variances change. ΔJD_{AR} is hence equal to $A + B$ according to (9)-(11), but R_{ar} is now defined as the ratio between $\Upsilon_{(2)}$ and $\Upsilon_{(1)}$. In addition, due to (17), ΔD_1 now satisfies:

$$\Delta D_1 = D_{k+1,1} - D_{k,1} = \frac{\Upsilon_{(2)}}{\Psi_{(1)}}\left(1 + {a_{(1)}}^2\right). \quad (28)$$

Let us now analyze ΔE_1 and ΔF_1. The expressions (18) and (25)

[1]The notation should be $Z^{(1)}$, but we have decided to omit the superscript $^{(1)}$ in the section in order to simplify the presentations of the equations.

ISBN 978-0-9928626-7-1

must be adjusted. Indeed, in the JD between noisy AR processes, $\rho_{(1)}$ defined in (21) was always positive. As Z could be greater than 1 or between -1 and 0, two expressions of $\left(S_{lim}^{(1)}\right)_{g,h}$ were proposed in (23). For the ARMA case, $\rho_{(1)}$ is no longer necessary positive. Therefore, we have to analyze how the properties of Z evolve in the next subsections.

A. About the properties of $\rho_{(1)}$ in the ARMA case

When dealing with ARMA processes, according to (4), $\rho_{(1)}$ becomes equal to:

$$\rho_{(1)} = \frac{\Psi_{(1)}}{\Upsilon_{(1)}} = \frac{a_{(1)} - (1+a_{(1)}^2)b_{(1)} + a_{(1)}b_{(1)}^2}{b_{(1)}}. \quad (29)$$

This quantity is hence positive or negative depending on the value of the ARMA parameters. It should be also noted that the same value of $\rho_{(1)}$ can be obtained by the couples $(a_{(1)}, b_{(1)})$ and $(a_{(1)}, \frac{1}{b_{(1)}})$.

B. About the properties of Z in the ARMA case

Let us now analyze the values that Z can take. First of all, due to the new properties of $\rho_{(1)}$ in the ARMA case, one has to check whether the variable Z is real or complex. Due to the definition (20) of Z and the new value of $\rho_{(1)}$ in (29), it depends on the sign of $F(\rho_{(1)}) = \left(\frac{1}{a_{(1)}}\left(1+\rho_{(1)}\right) + a_{(1)}\right)^2 - 4$. If $\rho_{(1)_{min}}$ and $\rho_{(1)_{max}}$ respectively denotes the minimum and the maximum roots of $F(\rho_{(1)})$ among $-\left(1+a_{(1)}\right)^2$ and $-\left(1-a_{(1)}\right)^2$, $F(\rho_{(1)})$ would be negative and Z would be complex for any value of $a_{(1)}$ in the interval $]-1,1[$ provided $\rho_{(1)_{min}} < \rho_{(1)} < \rho_{(1)_{max}}$. This is however not possible for any $a_{(1)}$ and $b_{(1)}$. As a consequence, Z is always real.
Then, it can be easily shown that two cases can happen. Either $Z > 1$ or $-1 < Z < 0$. See Figure 1.

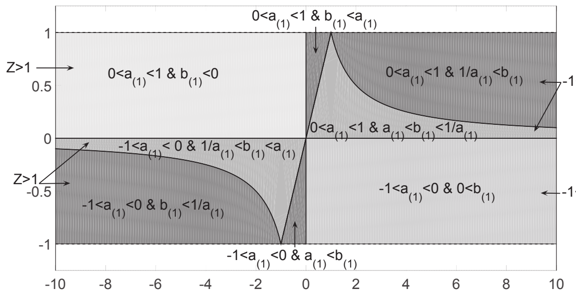


Fig. 1. Property on Z depending on $a_{(1)}$ and $b_{(1)}$.

All the cases have been addressed except $b_{(1)} = 1$ or $b_{(1)} = -1$. In these cases, for any value of $a_{(1)}$, $Z = -1$ or $Z = 1$ meaning that $\left(Z^2 - 1\right) = 0$. This result is hence coherent with [24].

C. About the properties of $\left(S_{lim}^{(1)}\right)_{g,h}$ in the ARMA case

Given the above analysis about Z, (23) now satisfies, except for $|b_{(1)}| = 1$:

$$\left(S_{lim}^{(1)}\right)_{g,h} = \begin{cases} -\frac{\sigma_{u(1)}^2}{\sigma_{n(1)}^2} \frac{Z^{h+1-g}}{a_{(1)}\left(Z^2-1\right)} \frac{\left(\left(Z^{2g}-1\right)+a_{(1)}Z\left(Z^{2g-2}-1\right)\right)}{\left(1+a_{(1)}Z\right)} \\ \text{for} \begin{cases} a_{(1)} > 0 \\ b_{(1)} > 0 \end{cases} \text{or} \begin{cases} a_{(1)} < 0 \\ b_{(1)} > 0 \end{cases}, \\ -\frac{\sigma_{u(1)}^2}{\sigma_{n(1)}^2} \frac{Z^{-h+1-g}}{a_{(1)}\left(Z^2-1\right)} \frac{\left(\left(Z^{2g}-1\right)+a_{(1)}Z\left(Z^{2g-2}-1\right)\right)}{\left(1+a_{(1)}Z^{-1}\right)} \\ \text{for} \begin{cases} a_{(1)} < 0 \\ b_{(1)} < 0 \end{cases} \text{or} \begin{cases} a_{(1)} > 0 \\ b_{(1)} < 0 \end{cases}. \end{cases} \quad (30)$$

D. About the expressions of ΔE_1 and ΔF_1 in the ARMA case

The expression of ΔE_1 initially introduced in (18) must be replaced by:

$$\begin{aligned} \Delta E_1 = & -\frac{\Upsilon_{(2)}}{\Psi_{(1)}} \frac{Z}{a_{(1)}\left(Z^2-1\right)} \Big[\left(1+4{a_{(1)}}^2 + {a_{(1)}}^4\right) \quad (31) \\ & + \left(4a_{(1)}\left(1+{a_{(1)}}^2\right)\right) Z + 2{a_{(1)}}^2 Z^2 \Big] \\ & \text{for} \begin{cases} a_{(1)} > 0 \\ b_{(1)} > 0 \end{cases} \text{or} \begin{cases} a_{(1)} < 0 \\ b_{(1)} > 0 \end{cases}, \\ \Delta E_1 = & \frac{\Upsilon_{(2)}}{\Psi_{(1)}} \frac{Z}{a_{(1)}\left(Z^2-1\right)} \Big[\left(1+4{a_{(1)}}^2 + {a_{(1)}}^4\right) \\ & + \left(4a_{(1)}\left(1+{a_{(1)}}^2\right)\right) Z^{-1} + 2{a_{(1)}}^2 Z^{-2} \Big] \\ & \text{for} \begin{cases} a_{(1)} < 0 \\ b_{(1)} < 0 \end{cases} \text{or} \begin{cases} a_{(1)} > 0 \\ b_{(1)} < 0 \end{cases}. \end{aligned}$$

Finally, the expression of ΔF_1 initially introduced in (25) must be replaced by:

$$\begin{aligned} \Delta F_1 = & -\Upsilon_{(1)} \left[\frac{\Psi_{(2)}}{1-{a_{(2)}}^2} \left(\left(T_{lim}^{(1)}\right)_{\frac{k}{2},\frac{k}{2}} \right.\right. \quad (32) \\ & \left.\left. + 2\sum_{i=1}^{\frac{k}{2}-1} \left(T_{lim}^{(1)}\right)_{\frac{k}{2},i} (-a_{(2)})^{\left(\frac{k}{2}-i\right)} \right) \right]. \end{aligned}$$

ΔJD_{ARMA} is the sum of ΔJD_{AR}, the variations ΔD_1, ΔE_1 and ΔF_1 defined in (28), (31), (32) and ΔD_2, ΔE_2 and ΔF_2 that can be easily deduced by switching the subscripts (1) and (2) in (28), (31), (32).

E. Comments

Except when the MA parameter of one or the two ARMA processes is equal to 1 or -1 , the derivative of the JD between ARMA processes with respect to k becomes a constant when k is getting higher. As for the AR and the MA case, the resulting asymptotic increment is sufficient to compare the processes. In the next section, let us illustrate these theoretical results by some examples.

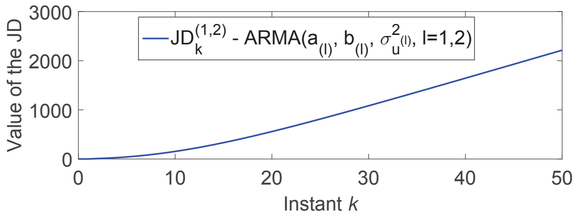
V. Illustrations and comments

Let us first consider two ARMA processes where $a_{(1)} = 0.7$, $b_{(1)} = 2$, $\sigma_{u(1)}^2 = 2$ and $a_{(2)} = -0.9$, $b_{(2)} = 1.1$ and $\sigma_{u(2)}^2 = 3$. The results are presented on Figure 2. It is confirmed that the JD between both ARMA processes tends to be constant after a transient period. In addition, on Figure 2 (b), $JD_{ARMA,k+1} - JD_{ARMA,k}$ is compared with ΔJD_{ARMA}. This particular case illustrates how the difference between the JDs for two successive instants tends to the constant derived in the theoretical part.
Depending on the ARMA parameters, the convergence speed towards the asymptotic increment is not the same. This phenomenon can be seen by comparing Figure 2 (b) and Figure 3, where $a_{(1)} = 0.9, a_{(2)} = 0.7$ - $b_{(1)} = 2.5, b_{(2)} = -0.5$ - $\sigma_{u(1)}^2 = 1, \sigma_{u(2)}^2 = 2$. Note that the value of the increment is not the same also. In addition, we can see the increment as a function of the AR parameter $a_{(2)}$, on Figure 4 where $a_{(1)} = 0.7$, $b_{(1)} = 2$, $\sigma_{u(1)}^2 = 2$, $b_{(2)} = 1.1$, $\sigma_{u(2)}^2 = 3$, k is set at 50 whereas $a_{(2)}$ varies in $[-0.9, 0.9]$. On this example, k is sufficiently large and ΔJD_{ARMA} fits $JD_{ARMA,k+1} - JD_{ARMA,k}$.

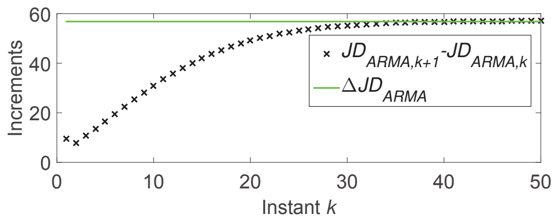
VI. Conclusions and perspectives

In this paper, we have studied the JD between two real ARMA processes with one AR parameter and one MA parameter. As the correlation matrix of the ARMA process has a structure similar to the one of the noisy AR process, we suggested analyzing how to benefit of the study we did for the JD between AR processes disturbed by additive white noises. This study is also useful to select

ISBN 978-0-9928626-7-1



(a) JD evolution over time.



(b) Increment evolution over time.

Fig. 2. JD and its increment evolutions when $a_{(1)} = 0.7, a_{(2)} = -0.9$ - $b_{(1)} = 2, b_{(2)} = 1.1$ - $\sigma^2_{u^{(1)}} = 2, \sigma^2_{u^{(2)}} = 3$

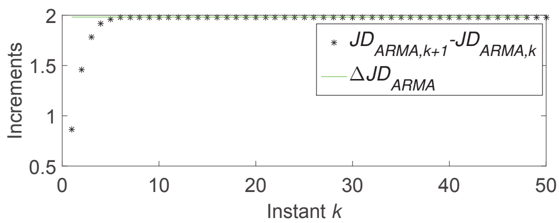
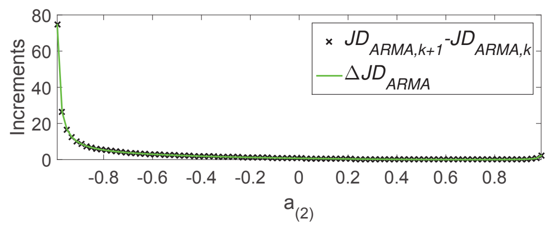


Fig. 3. JD and its increment evolutions when $a_{(1)} = 0.9, a_{(2)} = 0.7$ - $b_{(1)} = 2.5, b_{(2)} = -0.5$ - $\sigma^2_{u^{(1)}} = 1, \sigma^2_{u^{(2)}} = 2$



$a_{(1)} = 0.7, a_{(2)}$ varies in $[-0.9, 0.9]$,
$b_{(1)} = 2, b_{(2)} = 1.1$ - $\sigma^2_{u^{(1)}} = 2, \sigma^2_{u^{(2)}} = 3, k = 50$.

Fig. 4. Evolutions of $JD_{ARMA,k+1} - JD_{ARMA,k}$ and ΔJD_{ARMA} when $a_{(2)}$ varies.

k. Thus, due to the transient regime, it is better to avoid computing the JD for small values of k. Secondly, the derivative of the JD with respect to the number of variates k tends to be a constant. The resulting asymptotic increment is relevant of the behavior of the JD and is sufficient to compare ARMA processes. An expression of this asymptotic increment is given and depends on the ARMA parameters. We are currently investigating the case of noisy ARMA processes.

REFERENCES

[1] K. Assaleh, H. Al-Nashash, and N. Thakor, "Spectral subtraction and cepstral distance for enhancing EEG entropy," *IEEE Proc. of the Engineering in Medicine and Biology*, vol. 3, pp. 2751–2754, 2005.

[2] A. R. Gladston, P. Vijayalakshmi, and N. Thangavelu, "Improving speech intelligibility in Cochlear implants using Vocoder-centric acoustic models," *Recent Trends in Information Technology*, pp. 66–71, 2012.

[3] 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.

[4] C. Magnant, A. Giremus, and E. Grivel, "Jeffreys divergence between state models: Application to target tracking using multiple models," *EUSIPCO*, pp. 1–5, September 2013.

[5] 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.

[6] R. Bhansali, "A simulation study of autoregressive and window estimators of the inverse correlation function," *Journal of the Royal Statistical Society - Applied Statistics*, vol. 32, no. 2, pp. 141–149, 1983.

[7] A. Kaderli and A. Kayhan, "Spectral estimation of ARMA processes using ARMA cepstrum recursion," *IEEE Signal Processing Letter*, vol. 7, no. 9, pp. 259–261, 2000.

[8] C. Byrnes, P. Enqvist, and A. Lindquist, "Cepstral coefficients, covariance lags, and pole-zero models for finite data strings," *IEEE Trans. on Signal Processing*, vol. 49, no. 4, pp. 677–693, 2001.

[9] J. M. Mendel, "Tutorial on higher-order statistics (Spectra) in signal processing and systems theory: Theoretical results and some applications," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 79, no. 3, pp. 278–305, 1991.

[10] P. Stoica, T. McKelvey, and J. Mari, "MA estimation in polynomial time," *IEEE Trans. on Signal Processing*, vol. 48, no. 7, pp. 1999–2012, 2000.

[11] F. Merchan, F. Turcu, E. Grivel, and M. Najim, "Rayleigh fading channel simulator based on inner-outer factorization," *Signal Processing*, vol. 90, no. 1, pp. 24–33, 2010.

[12] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 27, pp. 53–62, 1979.

[13] W. R. Wu and P. C. Chen, "Adaptive AR modeling in white Gaussian noise," *IEEE Trans. on Signal Processing*, vol. 45, pp. 1184–1191, 1997.

[14] W. X. Zheng, "A least-squares based method for autoregressive signal in presence of noise," *IEEE Trans. on Circuits and Systems II : Analog and Digital Signal Processing*, vol. 46, pp. 531–534, 1999.

[15] C. E. Davila, "A subspace approach to estimation of autoregressive parameters from noisy measurements," *IEEE Trans. on Signal Processing*, vol. 46, pp. 531–534, February 1998.

[16] S. Beghelli, R. Guidorzi, and U. Soverini, "The Frisch scheme in dynamic system identification," *Automatica*, vol. 26, pp. 171–176, 1990.

[17] R. Diversi, R. Guidorzi, and U. Soverini, "Identification of autoregressive models in the presence of additive noise," *Int. J. Adapt. Control Signal Process*, vol. 22, pp. 465–481, 2008.

[18] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Identification of autoregressive moving average systems based on noise compensation in the correlation domain," *IET Signal Process.*, vol. 5, no. 3, pp. 292–305, 2011.

[19] M. K. Hasan, N. M. Hosssain, and P. A. Naylor, "Autocorrelation model-based identification method for ARMA systems in noise," *IEE Proc.-Vis. Image Signal Process.*, vol. 152, no. 5, pp. 520–526, 2005.

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

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

[22] 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.

[23] W. W. Hager, "Updating the Inverse of a Matrix," *SIAM Review*, vol. 31, no. 2, pp. 221–239, 1989.

[24] L. Legrand and E. Grivel, "Evaluating dissimilarities between moving-average models: a comparative study between Jeffreys divergence and rao distance," *24th EUSIPCO conference*, 2016.