Jeffrey's Divergence Between Complex-Valued Sinusoidal Processes

Eric Grivel*, Mahdi Saleh †* and Samir-Mohamad Omar†
*Bordeaux university - INP Bordeaux ENSEIRB-MATMECA - IMS - UMR CNRS 5218, Talence, France
†Department of Physics and Electronics, Faculty of Sciences I, Lebanese University Hadath, Beirut, Lebanon

***Abstract*—Like other divergences, Jeffrey's divergence (JD) is used for change detection, for model comparison, etc. Recently, a great deal of interest has been paid to this symmetric version of the Kullback-Leibler (KL) divergence. This led to analytical expressions of the JD between autoregressive (AR) processes, moving-average (MA) processes, either noise-free or disturbed by additive white noises, as well as ARMA processes. In this paper, we propose to study the JD between processes that are defined as sums of complex-valued sinusoidal processes disturbed by additive white noises. We show that the JD tends to a stationary behavior when the number of variates becomes large. The derivative of the JD becomes a constant that depends on the parameters defining the processes. The convergence speed towards this stationary regime depends on the differences between the normalized angular frequencies. The smaller the difference, the slower the convergence. This result can be obtained by interpreting some steps to compute the JD as orthogonal projections. Some examples illustrate the theoretical analysis.**

***Index Terms*—Jeffrey's divergence, Kullback-Leibler divergence, change detection, model comparison.**

I. Introduction

In various applications from seismic to biomedical signal processing, change detection algorithms can be useful. The problem is often to decide whether a change has occurred in a set of data. For this purpose, a reference set of data is compared with a second one which is defined by using a sliding window. In other cases, it can be of interest to compare different sets of data that are not necessarily recorded in the same situations, i.e. at the same time and/or by the same sensors and/or for the same patients in biomedical applications, etc. This is for instance the case to detect pathology from EEG or ECG and to analyze the properties of sea clutter in different areas when dealing with radar processing.

Among the solutions that can be considered, spectrum comparison can be done, based on the log spectral distance or the Itakura-Saito distance. In these cases, the periodogram or the pseudo-spectra deduced from Capon's method or subspace methods such as MUSIC can be used.

As an alternative, Kullback-Leibler (KL) divergence, Pearson divergence (PD) or the relative Pearson's divergence consist in comparing the distributions of the set of data. Several authors have analyzed these divergences for stochastic processes and have used them in various applications. See for instance [1], [2] and [3]. The estimations of the KL and the PD between two probability distributions that are not necessarily Gaussian from sets of data have been also studied. Instead of estimating the densities from the sets of data, the density ratio is directly estimated. For the KL case, this leads to the KL importance estimation procedure (KLIEP) and the Gaussian mixture KLIEP (GM-KLIEP) [4], [5], as well as the M-estimator-based approach [6]. The symmetric Kullback-Leibler (KL) divergence, known as Jeffrey's divergence (JD), has been computed between the distributions of the successive samples of two time-varying AR (TVAR) processes [7]. This method has been also used to classify more than two AR processes in different subsets [8]. The analytical expressions of the JD between 1st-order MA processes, that can be real or complex, noise-free or disturbed by additive white Gaussian noises, has also been studied in [9]. Finally, comparing AR and MA processes using the JD has been proposed in [10].

In this paper, we present a complementary study that aims at comparing two processes defined as sums of complex exponentials disturbed by uncorrelated additive white noises. Our purpose is to analyze the influence of the process parameters on the JD and to give some clues for the interpretation. We will see that the JD computed for a set of k successive samples is not necessarily enough to conclude. Its evolution with respect to k as well as its asymptotic increment are useful.

This paper is organized as follows: in sections II, we briefly recall the definition and the expression of the JD. Then, the JD between the sum of complex-valued sinusoidal processes disturbed by additive white noises is addressed. In section III, theoretical results are illustrated by some examples.

In the following, I_k is the identity matrix of size k and Tr the trace of a matrix. The upperscripts ^{T} and ^{H} denote the transpose and the hermitian. $x_{k_1:k_2} = (x_{k_1}, ..., x_{k_2})$ is the collection of samples from time k_1 to k_2.

II. Jeffrey's divergence between sums of complex exponentials disturbed by additive noises

A. Definition and advantages of the Jeffrey's divergence

The Kullback-Leibler (KL) divergence between the joint distributions of k successive values of two random processes, denoted as $p_1(x_{1:k})$ and $p_2(x_{1:k})$, can be evaluated to study the dissimilarities between the processes [11].

$$KL_k{}^{(1,2)} = \int_{x_{1:k}} p_1(x_{1:k}) ln \left(\frac{p_1(x_{1:k})}{p_2(x_{1:k})} \right) \mathrm{d}x_{1:k} \qquad (1)$$

When the processes are both Gaussian and real with means $\mu_{1,k}$ and $\mu_{2,k}$ and covariance matrices $Q_{1,k}$ and $Q_{2,k}$, it can be easily shown, by substituting $p_1(x_{1:k})$ and $p_2(x_{1:k})$ with

ISBN 978-0-9928626-7-1 © EURASIP 2017

the expressions of Gaussian multivariate distributions, that the KL satisfies [12]:

$$KL_k{}^{(1,2)} = \frac{1}{2}\Big[\mathrm{Tr}(Q_{2,k}{}^{-1}Q_{1,k}) - k - \ln\frac{\det Q_{1,k}}{\det Q_{2,k}} + (\mu_{2,k}-\mu_{1,k})^T Q_{2,k}{}^{-1}(\mu_{2,k}-\mu_{1,k})\Big]. \quad (2)$$

When dealing with zero-mean processes, (2) reduces[1] to:

$$KL_k{}^{(1,2)} \propto \frac{1}{2}\Big[\mathrm{Tr}(Q_{2,k}{}^{-1}Q_{1,k}) - k - \ln\frac{\det Q_{1,k}}{\det Q_{2,k}}\Big]. \quad (3)$$

However, the KL is not symmetric. Therefore, the Jeffrey divergence can be preferred. It satisfies:

$$JD_k^{(1,2)} = \frac{1}{2}(KL_k^{(1,2)} + KL_k^{(2,1)}), \quad (4)$$

Using (3) and (4) leads to the following expression of the JD:

$$JD_k{}^{(1,2)} \propto -k + \frac{1}{2}\left[Tr(Q_{2,k}{}^{-1}Q_{1,k}) + Tr(Q_{1,k}{}^{-1}Q_{2,k})\right]. \quad (5)$$

This no longer involves a logarithm but depends on the $k \times k$ covariance matrices of the processes. For instance, the JD between two real Gaussian zero-mean noises with variances $\sigma_l{}^2$ with $l = 1, 2$ is equal to:

$$JD_k{}^{(1,2)} = -k + \frac{k}{2}\left(\frac{\sigma_1{}^2}{\sigma_2{}^2} + \frac{\sigma_2{}^2}{\sigma_1{}^2}\right). \quad (6)$$

The increment of the JD is hence equal to $-1 + \frac{1}{2}\left(\frac{\sigma_1{}^2}{\sigma_2{}^2} + \frac{\sigma_2{}^2}{\sigma_1{}^2}\right)$. In the following, let us study the JD between sums of complex exponentials disturbed by additive white noises.

B. Statistical properties of the sums of sinusoids disturbed by additive white noises

Let us assume that the processes under study are Gaussian zero-mean and defined by their covariance matrices which can be expressed as follows, with $l = 1, 2$:

$$Q_{l,k} = S_{l,k} P_l S_{l,k}^H + \sigma_l{}^2 I_k. \quad (7)$$

where $S_{l,k}$ is a matrix of size $k \times M_l$, with k the number of successive samples that are considered and M_l the number of complex exponentials with normalized angular frequencies $\{\theta_{l,m}\}_{m=1,\ldots,M_l}$. For each process, these latter are in the interval $[-\pi, \pi[$ and are different from each other. However, the two processes can have common normalized angular frequencies. In addition, P_l is a diagonal matrix of size $M_l \times M_l$ whose main diagonal is defined by the variances $\{\gamma_{l,m}\}_{m=1,\ldots,M_l}$ of the random zero-mean magnitudes of the complex exponentials. Finally, one has:

$$S_{l,k} = \begin{bmatrix} S_{l,k}^1 & \ldots & S_{l,k}^{M_l}\end{bmatrix} = \begin{bmatrix} 1 & \ldots & 1 \\ e^{j\theta_{l,1}} & \ldots & e^{j\theta_{l,M_l}} \\ \vdots & & \vdots \\ e^{j(k-1)\theta_{l,1}} & \ldots & e^{j(k-1)\theta_{l,M_l}} \end{bmatrix} \quad (8)$$

At this stage let us recall the following properties that these vectors satisfy. They will be useful in the remainder of this paper.

$$\begin{cases} \frac{1}{k}\left(S_{l,k}^m\right)^H S_{l,k}^m &= 1 \\ \frac{1}{k}\left(S_{l,k}^m\right)^H S_{l,k}^n &= \frac{1}{k}\sum_{t=0}^{k-1} e^{j(\theta_{l,m}-\theta_{l,n})t} \\ & = \frac{sin(\frac{k(\theta_{l,m}-\theta_{l,n})}{2})}{ksin(\frac{(\theta_{l,m}-\theta_{l,n})}{2})} e^{j\left(\frac{(k-1)(\theta_{l,m}-\theta_{l,n})}{2}\right)} \end{cases} \quad (9)$$

As $\lim_{k\to+\infty} |\frac{sin(\frac{k(\theta_{l,m}-\theta_{l,n})}{2})}{ksin(\frac{(\theta_{l,m}-\theta_{l,n})}{2})})| = 0$ for any set of normalized angular frequencies satisfying $\theta_{l,n} \neq \theta_{l,m}$ in the interval $[-\pi, \pi[$, (9) becomes:

$$\begin{cases} \frac{1}{k}\left(S_{l,k}^m\right)^H S_{l,k}^m &= 1 \\ \lim_{k\to+\infty} \frac{1}{k}\left(S_{l,k}^m\right)^H S_{l,k}^n &= 0 \end{cases} \quad (10)$$

The same properties hold when we consider normalized angular frequencies of both processes. The convergence speed to tend to this limit when k increases depends on the difference between the normalized angular frequencies. The smaller the difference, the slower the convergence.

Given (7) and using the matrix inversion lemma [2], one has for $l = 1, 2$:

$$Q_{l,k}{}^{-1} = \frac{1}{\sigma_l{}^2}\left(I_k - S_{l,k}\left(\sigma_l{}^2 P_l{}^{-1} + S_{l,k}^H S_{l,k}\right)^{-1} S_{l,k}^H\right) \quad (11)$$

In the next section, using the definitions and properties of the JD and the processes under study, we propose to study $Tr(Q_{2,k}{}^{-1}Q_{1,k})$.

C. Expression of the trace $Tr(Q_{2,k}{}^{-1}Q_{1,k})$

Given (7) and (11), after developing, $Tr(Q_{2,k}{}^{-1}Q_{1,k})$ can be expressed as the sum of four terms denoted as $A^{(2,1)}$, $B^{(2,1)}$, $C^{(2,1)}$ and $D^{(2,1)}$. The order of the upperscripts corresponds to the order of the matrices in the trace to be computed. Let us start expressing the first one. Since $\mathrm{Tr}(FG) = \mathrm{Tr}(GF)$ where F and G are non square matrices but FG and GF are square matrices, one has:

$$A^{(2,1)} = \frac{1}{\sigma_2{}^2}Tr\left(S_{1,k}P_1S_{1,k}^H\right) = \frac{1}{\sigma_2{}^2}Tr\left(S_{1,k}^H S_{1,k}P_1\right) = \frac{k}{\sigma_2{}^2}\sum_{m=1}^{M_1}\gamma_{1,m} = \frac{k}{\sigma_2{}^2}Tr(P_1) \quad (12)$$

Then, one has:

$$B^{(2,1)} = \frac{1}{\sigma_2{}^2}Tr\left(\sigma_1{}^2 I_k\right) = \frac{\sigma_1{}^2}{\sigma_2{}^2}k \quad (13)$$

and

$$C^{(2,1)} = -\frac{\sigma_1{}^2}{\sigma_2{}^2}Tr\left(S_{2,k}^H S_{2,k}\left(S_{2,k}^H S_{2,k} + \sigma_2{}^2 P_2{}^{-1}\right)^{-1}\right) \quad (14)$$

[1]In the complex case, ^{T} is replaced by ^{H} and $\frac{1}{2}$ disappears in (2). In the following, $\propto$ is used instead of $=$.

[2]Given the matrices A, U, C and V where A and C are assumed to be invertible, one has: $(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}$

In (14), let us focus our attention on the matrix $\left(S_{2,k}^H S_{2,k} + {\sigma_2}^2 {P_2}^{-1}\right)^{-1}$. By applying again the inversion matrix lemma, one has:

$$(S_{2,k}^H S_{2,k} + {\sigma_2}^2 {P_2}^{-1})^{-1} = (S_{2,k}^H S_{2,k})^{-1} \tag{15}$$
$$-(S_{2,k}^H S_{2,k})^{-1}({\sigma_2}^{-2} P_2 + (S_{2,k}^H S_{2,k})^{-1})^{-1}(S_{2,k}^H S_{2,k})^{-1}$$

Using (14) and (15) leads to:

$$C^{(2,1)} = -\frac{{\sigma_1}^2}{{\sigma_2}^2} Tr(I_{M_2}) \tag{16}$$
$$+ \frac{{\sigma_1}^2}{{\sigma_2}^2} Tr(({\sigma_2}^{-2} P_2 + (S_{2,k}^H S_{2,k})^{-1})^{-1}(S_{2,k}^H S_{2,k})^{-1})$$

When k increases, due to the asymptotic properties (10) of "orthogonality", the second term of $C^{(2,1)}$ in (16) tends to zero. Therefore, the trace $C^{(2,1)}$ tends to the following expression:

$$\lim_{k\to+\infty} C^{(2,1)} = -\frac{{\sigma_1}^2}{{\sigma_2}^2} Tr\left(I_{M_2}\right) = -\frac{{\sigma_1}^2}{{\sigma_2}^2} M_2 \tag{17}$$

Depending on the normalized angular frequencies of the second process, the convergence speed is more or less fast when k increases. The closer the normalized angular frequencies are, the slower the convergence speed is.

Finally, let us study the fourth term. It is defined as follows:

$$D^{(2,1)} = \tag{18}$$
$$-\frac{1}{{\sigma_2}^2} Tr\left(S_{1,k}^H S_{2,k}\left({\sigma_2}^2 {P_2}^{-1} + S_{2,k}^H S_{2,k}\right)^{-1} S_{2,k}^H S_{1,k} P_1\right)$$

In the above equation (18), let us denote:

$$T_{1/2,k} = S_{2,k}\left({\sigma_2}^2 {P_2}^{-1} + S_{2,k}^H S_{2,k}\right)^{-1} S_{2,k}^H S_{1,k} \tag{19}$$

As we already did for $C^{(2,1)}$, (19) can be approximated by using (15) as follows:

$$T_{1/2,k} = S_{2,k}(S_{2,k}^H S_{2,k})^{-1} S_{2,k}^H S_{1,k} - S_{2,k}(S_{2,k}^H S_{2,k})^{-1}\times \tag{20}$$
$$({\sigma_2}^{-2} P_2 + (S_{2,k}^H S_{2,k})^{-1})^{-1}(S_{2,k}^H S_{2,k})^{-1} S_{2,k}^H S_{1,k}$$
$$= S_{1/2,k} + U_{1/2,k}$$

where:

$$\begin{cases} S_{1/2,k} & = S_{2,k}\left(S_{2,k}^H S_{2,k}\right)^{-1} S_{2,k}^H S_{1,k} \\ U_{1/2,k} & = -S_{2,k}(S_{2,k}^H S_{2,k})^{-1}\times \\ & ({\sigma_2}^{-2} P_2 + (S_{2,k}^H S_{2,k})^{-1})^{-1}(S_{2,k}^H S_{2,k})^{-1} S_{2,k}^H S_{1,k} \end{cases} \tag{21}$$

When substituting the expression (20) of $T_{1/2,k}$ into (18), $D^{(2,1)}$ can be expressed as the sum of two traces. In the following, we propose to evaluate both:

1) **The first one is induced by $\mathbf{S_{1/2,k}}$**: The matrix $S_{1/2,k}$ stores the orthogonal projections of the columns of $S_{1,k}$ onto the space spanned by the columns of $S_{2,k}$. Given the "asymptotic"' properties (10) of orthogonality when k tends to infinity, the orthogonal projection of the m^{th} column $S_{1,k}^m$ of $S_{1,k}$ onto $S_{2,k}$ is a null column vector except when the two processes have common normalized angular frequencies. Indeed, if there exists $n \in [\![1, M_2]\!]$ such as $\theta_{1,m} = \theta_{2,n}$, this leads to:

$$\lim_{k\to+\infty} S_{2,k}\left(S_{2,k}^H S_{2,k}\right)^{-1} S_{2,k}^H S_{1,k}^m = S_{1,k}^m \tag{22}$$

Therefore, $\lim_{k\to+\infty} S_{1/2,k}$ can be approximated by:

$$S_{1/2} = \lim_{k\to+\infty} S_{1/2,k} \approx \begin{bmatrix} S_{1,k}^1 \delta_{1,2}^1 & \dots & S_{1,k}^{M_1} \delta_{1,2}^{M_1} \end{bmatrix} \tag{23}$$

where $\delta_{1,2}^m = \sum_{n=1}^{M_2} \delta_{\theta_{1,m},\theta_{2,n}}$ with $\delta_{\theta_{1,m},\theta_{2,n}}$ equal to 1 when $\theta_{1,m} = \theta_{2,n}$ and zero otherwise. In other words, $\delta_{1,2}^m = 1$ if both processes share the normalized angular frequency $\theta_{1,m}$. Otherwise it is equal to 0.

Combining (18) and (23), this leads to:

$$\lim_{k\to+\infty} -\frac{1}{{\sigma_2}^2} Tr(S_{1,k}^H S_{2,k}(S_{2,k}^H S_{2,k})^{-1} S_{2,k}^H S_{1,k} P_1) \approx \tag{24}$$

$$-\frac{1}{{\sigma_2}^2} Tr\left(S_{1,k}^H S_{1/2} P_1\right) = \frac{-k}{{\sigma_2}^2} \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \gamma_{1,m} \delta_{\theta_{1,m},\theta_{2,n}}$$

This term is null if the processes do not have at least one common normalized angular frequency.

2) **The second one is induced by $\mathbf{U_{1/2,k}}$**: When k increases, due to the "asymptotic orthogonality properties"' (10), some simplifications can be done. One has:

$$\lim_{k\to+\infty} U_{1/2,k} = \tag{25}$$

$$-\lim_{k\to+\infty} \frac{1}{k^2} S_{2,k}({\sigma_2}^{-2} P_2 + \frac{1}{k} I_{M_2})^{-1} S_{2,k}^H S_{1,k}$$

$$= -\lim_{k\to+\infty} \frac{{\sigma_2}^2}{k^2} S_{2,k} P_2^{-1} S_{2,k}^H S_{1,k}$$

As we aim at calculating the trace of $-\frac{1}{{\sigma_2}^2} Tr\left(S_{1,k}^H U_{1/2,k} P_1\right)$, one has:

$$\lim_{k\to+\infty} \frac{1}{k^2} Tr\left(S_{1,k}^H S_{2,k} {P_2}^{-1} S_{2,k}^H S_{1,k} P_1\right) \tag{26}$$
$$= \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \frac{\gamma_{1,m}}{\gamma_{2,n}} \delta_{\theta_{1,m},\theta_{2,n}}$$

Combining (12), (13), (17), (24) and (26), $Tr({Q_{2,k}}^{-1} Q_{1,k})$ can be expressed as follows when k tends to infinity:

$$Tr({Q_{2,k}}^{-1} Q_{1,k}) \approx \frac{k}{{\sigma_2}^2} \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \gamma_{1,m}(1 - \delta_{\theta_{1,m},\theta_{2,n}}) \tag{27}$$
$$+ (k - M_2)\frac{{\sigma_1}^2}{{\sigma_2}^2} + \sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \frac{\gamma_{1,m}}{\gamma_{2,n}} \delta_{\theta_{1,m},\theta_{2,n}}$$

Remark: When the processes have the same covariance matrix, $\sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \frac{\gamma_{1,m}}{\gamma_{2,n}} \delta_{\theta_{1,m},\theta_{2,n}} = M_1 = M_2$ and

$\sum_{m=1}^{M_1}\sum_{n=1}^{M_2}\gamma_{1,m}(1-\delta_{\theta_{1,m},\theta_{2,n}})=0$. In this case, (27) becomes:

$$Tr({Q_{2,k}}^{-1}Q_{1,k})=(k-M_2)\frac{{\sigma_1}^2}{{\sigma_2}^2}+M_1=k=Tr(I_k) \quad (28)$$

D. Expression of the JD

As the expression of $Tr({Q_{1,k}}^{-1}Q_{2,k})$ when k tends to infinity is similarly defined when the indices 1 and 2 are switched and given (5) and (27), the JD can be approximated when k tends to infinity as follows:

$$\begin{aligned} {JD_k}^{(1,2)} \propto & -k+\frac{1}{2}[(k-M_2)\frac{{\sigma_1}^2}{{\sigma_2}^2}+(k-M_1)\frac{{\sigma_2}^2}{{\sigma_1}^2} \\ & +k\sum_{m=1}^{M_1}\sum_{n=1}^{M_2}(\frac{\gamma_{1,m}}{{\sigma_2}^2}+\frac{\gamma_{2,n}}{{\sigma_1}^2})(1-\delta_{\theta_{1,m},\theta_{2,n}}) \\ & +\sum_{m=1}^{M_1}\sum_{n=1}^{M_2}\left(\frac{\gamma_{1,m}}{\gamma_{2,n}}+\frac{\gamma_{2,n}}{\gamma_{1,m}}\right)\delta_{\theta_{1,m},\theta_{2,n}}]. \end{aligned} \quad (29)$$

The JD depends on the parameters of the processes, namely the normalized angular frequencies of each process, the variances of the magnitudes of each component as well as the variances of the additive noises. The first term of the expression includes the JD between two zero-mean white noises with variances ${\sigma_1}^2$ and ${\sigma_2}^2$. When there is no complex exponentials in both processes, the processes under study correspond to white noises and (29) reduces to (6).

Note that the two other terms in (29) make it possible to point out the differences between the sets of complex exponentials.

In the following subsection, let us analyze how the JD evolves when k is incremented.

E. Analysis of the increment of the Jeffrey's divergence

Given the expression (29) of the JD we obtained, let us now deduce the asymptotic increment, i.e. the increment of the JD when k tends to infinity:

$$\Delta JD=\lim_{k\to+\infty}{JD_k}^{(1,2)}-{JD_{k-1}}^{(1,2)} \quad (30)$$

One has:

$$\Delta JD \propto -1 +\frac{1}{2}\left[\frac{{\sigma_1}^2}{{\sigma_2}^2}+\frac{{\sigma_2}^2}{{\sigma_1}^2}+\sum_{m=1}^{M_1}\sum_{n=1}^{M_2}(\frac{\gamma_{1,m}}{{\sigma_2}^2}+\frac{\gamma_{2,n}}{{\sigma_1}^2})(1-\delta_{\theta_{1,m},\theta_{2,n}})\right] \quad (31)$$

The ratio of the additive-white-noise variances has an influence on the slop of the JD. In addition, when there are common normalized angular frequencies between two processes, ΔJD is smaller.

In the next section, we suggest illustrating the way the JD evolves when comparing two processes.

III. Illustrations and comments

A. Influence of the additive-noise variances

The parameters are the following in this first simulation: $M_1=1$, $\theta_{1,1}=-\pi/5$, $\gamma_{1,1}=100$ and ${\sigma_1}^2=1$. $M_2=1$, $\theta_{2,1}=-2\pi/5$, $\gamma_{2,1}=40$ and ${\sigma_2}^2=0.5$. Then, a second simulation is done where ${\sigma_2}^2$ is doubled: ${\sigma_2}^2=1$. Finally, a last simulation is done where ${\sigma_2}^2=2$. As illustrated by Fig. 1, the asymptotic increment is modified according to (31).

It always depends on $\frac{{\sigma_1}^2}{{\sigma_2}^2}+\frac{{\sigma_2}^2}{{\sigma_1}^2}$ and may depend on the set of ratios $\{\frac{\gamma_{1,m}}{{\sigma_2}^2}+\frac{\gamma_{2,n}}{{\sigma_1}^2}\}_{m=1,\dots,M_1 \text{and } n=1,\dots,M_2}$.

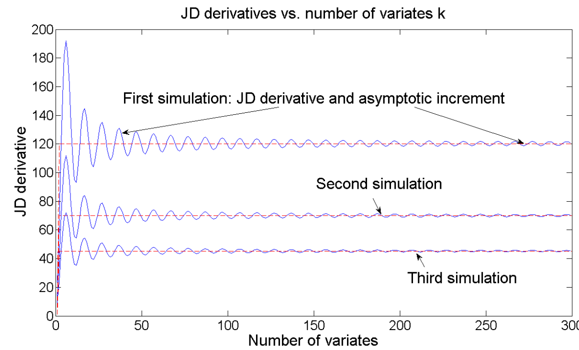

Fig. 1: Asymptotic increment vs increment, 1^{st} example, with three simulations where ${\sigma_2}^2$ is modified.

B. Convergence speed towards the stationary regime

The parameters are the following: the two processes are defined as in the first example (See III.A) except that ${\sigma_2}^2=0.1$. Then, two other simulations are done with $\theta_{2,1}=-1.5\pi/5$ and finally $\theta_{2,1}=-1.1\pi/5$.

Given Fig. 2, one can notice that the JD computed from (5) and (7) tends to be the same for the three simulations when k increases. The main differences are located when k is small. For this reason, we suggest computing the derivatives of these JD. They are given in Fig. 3 where they are compared with the asymptotic increment given in (31).

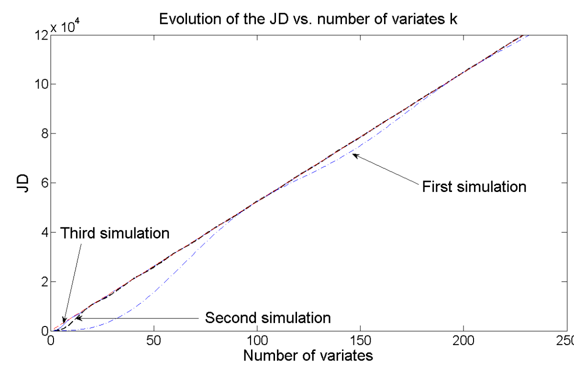

Fig. 2: JD and its approximation, 2^{nd} example, with three simulations where $\theta_{2,1}$ becomes closer and closer to $\theta_{1,1}=-\pi/5$

As presented in (31) in the theoretical analysis in section II, we can see that the asymptotic increment is here the same for all the simulations. The only difference stands in the convergence speed towards the stationary regime. Indeed, given Fig. 3, the convergence of the JD derivative towards the asymptotic increment is faster when the difference between the normalized angular frequencies is large. The fluctuations of the derivatives around the asymptotic increment are mainly due to the products such as $S_{1,k}^H S_{2,k}$ and $S_{2,k}^H S_{1,k}$ that must be computed. They correspond to values of $\frac{sin(\frac{k(\theta_{l,m}-\theta_{l,n})}{2})}{sin(\frac{(\theta_{l,m}-\theta_{l,n})}{2})}$ whose square is periodic with respect to k with period equal to $\frac{2\pi}{|\theta_{1,n}-\theta_{2,m}|}$.

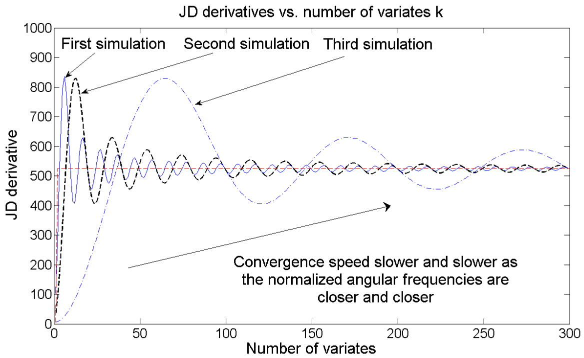

Fig. 3: Asymptotic increment vs increment, 2^{nd} example with three simulations where $\theta_{2,1}$ becomes closer and closer to $\theta_{1,1} = -\pi/5$

C. A more general case

The parameters are the following: $M_1 = 3$, $\theta_{1,1} = \pi/10$, $\theta_{1,2} = \pi/4$, $\theta_{1,3} = \frac{4\pi}{5}$, $\gamma_{1,1} = 100$, $\gamma_{1,2} = 50$, $\gamma_{1,3} = 75$ and ${\sigma_1}^2 = 1$. $M_2 = 3$, $\theta_{2,1} = -\pi/10$, $\theta_{2,2} = \pi/2$,$\theta_{2,3} = 4\pi/5$, $\gamma_{2,1} = 40$, $\gamma_{2,2} = 60$, $\gamma_{2,3} = 80$ and ${\sigma_2}^2 = 0.1$. Then, a second simulation is done where $\theta_{2,2}$ is modified: $\theta_{2,2} = \pi/4$. Finally, a third simulation is presented where the normalized angular frequencies are the same for both processes.

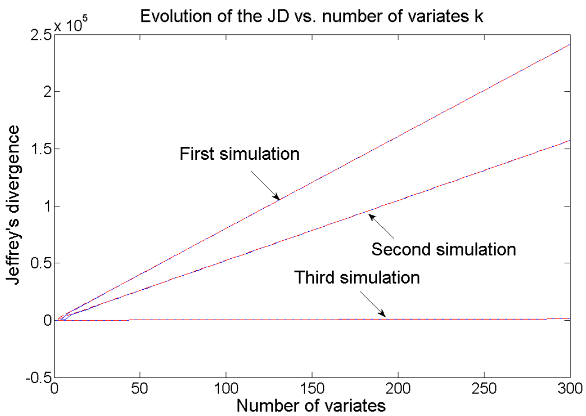

Fig. 4: JD and its approximation, 3^{rd} example with three simulations where the processes share 1, then 2 and then 3 normalized angular frequencies

In Fig. 4 and 5, we present the evolution of the JD and the increment over time obtained from (5) respectively. We compare the increment (i.e. the derivative of the JD with respect to the number of variates k) with the asymptotic increment obtained in (31). It confirms our theoretical analysis. In practical cases, the covariance matrices can be estimated in the maximum likelihood sense from the data. Then, the JD is deduced for various sizes k of the covariance matrices. In order to make the interpretation easier, the derivative of the JD with respect to k must be computed and plotted to analyze the convergence speed towards the stationary regime.

IV. Conclusions and Perspectives

In this paper, we have studied the JD between two sums of complex exponentials disturbed by additive white noises.

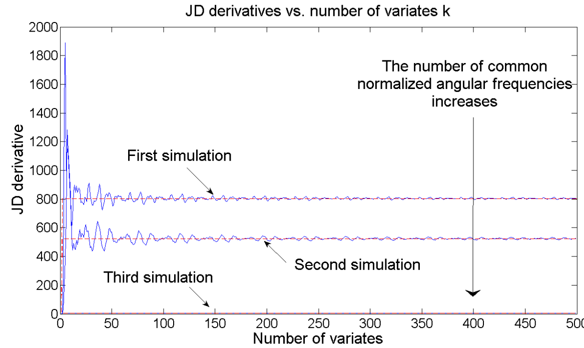

Fig. 5: Asymptotic increment vs increment, 3^{rd} example with three simulations where the processes share 1, then 2 and then 3 normalized angular frequencies

We have shown that the derivative of the JD with respect to the number of variates k tends to be a constant. The resulting asymptotic increment is of interest to compare the processes. It consists of two terms: one which depends on the ratio of the additive-white-noise variances and another which depends on the complex-exponential magnitude variances, the noise variances and the normalized angular frequencies. The analysis of the convergence speed as well as the value of the asymptotic increment have to be taken into account for the interpretation.

We are currently studying a comparison between the JD between complex exponentials disturbed by additive noises and the JD between two AR processes whose power spectral densities exhibit sharp resonances.

References

[1] L. Bombrun, N.-E. Lasmar, Y. Berthoumieu, and G. Verdoolaege, "Multivariate texture retrieval using the SIRV representation and the geodesic distance," *IEEE ICASSP*, pp. 865–868, 2011.

[2] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis, "Touchless monitoring of breathing function," *IEEE EMBS*, pp. 1196–1199, 2004.

[3] A. Schutz, L.Bombrun, Y. Berthoumieu, and M. Najim, "Centroid-based texture classification using the generalized gamma distribution," *EUSIPCO*, pp. 1–5, 2013.

[4] M. Sugiyama, T. Suzuki, S. Nakajima, P. von Bunau, and M. Kawanabe, "Direct importance estimation for covariate shift adapation," *Annals of the Institute*, vol. 60, n4, pp. 699–746, 2008.

[5] M. Yamada and M. Sugiyama, "Direct importance estimation with gaussian mixture models," *IEICE Transactions on Information and Systems*, vol. E92-D, n10, pp. 2159–2162, 2009.

[6] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "Estimating divergence functionals and the likelihood ratio by convex risk minimization," *IEEE Transactions on Information Theory*, vol. 56, n11, pp. 5847–5861, 2010.

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

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

[9] 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.

[10] L. Legrand and E. Grivel, "Jeffrey's divergence between moving-average and autoregressive models," *IEEE ICASSP*, 2017.

[11] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

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