Kullback-Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison

Published Version:
Grivel, E., Diversi, R., Merchan, F. (2021). Kullback-Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison. DIGITAL SIGNAL PROCESSING, 116, 1-21 [10.1016/j.dsp.2021.103089].

Availability:
This version is available at: https://hdl.handle.net/11585/855188 since: 2024-03-01

Published:
DOI: http://doi.org/10.1016/j.dsp.2021.103089

Kullback-Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison

Eric Grivel[a], Roberto Diversi[b], Fernando Merchan[c]

[a]*Bordeaux University - INP Bordeaux - IMS - UMR CNRS 5218, FRANCE*
[b]*Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione, University of Bologna, ITALY*
[c]*Facultad de Ingeniería Eléctrica, Universidad Tecnológica de Panamá, PANAMA*

Abstract

In signal processing, ARMA processes are widely used to model short-memory processes. In various applications, comparing or classifying ARMA processes is required. In this paper, our purpose is to provide analytical expressions of the divergence rates of the Kullback-Leibler divergence, the Rényi divergence (RD) of order α and their symmetric versions for two Gaussian ARMA processes, by taking advantage of results such as the Yule-Walker equations and notions such as inverse filtering. The divergence rates can be interpreted as the sum of different quantities: power of one ARMA process filtered by the inverse filter associated with the second ARMA process, cepstrum, etc. Finally, illustrations show that the ranges of values taken by the divergence rates of the RD is sensitive to α, especially when the latter is close to 1.

Keywords: Kullback-Leibler divergence, Rényi divergence, ARMA process.

Preprint submitted to Digital Signal Processing *March 21, 2021*

1. Introduction

AutoRegressive Moving Average (ARMA) models are used to model short-memory processes for different purposes such as spectral estimation, change detection and feature extraction for signal coding or classification in different domains such as speech processing, biomedical applications or radar processing [1, 19, 23, 36, 45]. In some of the above-mentioned applications, the comparison of the ARMA models is often crucial. For instance, in system identification, the performance of an estimation method is usually evaluated by comparing the identified model with the true one, starting from synthetic data [5, 12]. In change detection and fault diagnosis, the estimated model is continuously updated and compared with a reference model in order to check if a change or a fault has occurred [1, 21].

There are various standard ways to compare two ARMA models: one can consider the 2-norm and/or the infinity norm of the vector storing the ARMA-parameter differences or one can look at the poles and zeros in the z-plane of the transfer function associated with the ARMA model. Starting from the parameters of the ARMA models, it is also possible to determine the corresponding power spectral densities and then compute the log-spectral distance (LSD), the Itakura-Saito distance, that takes its origins in probability theory, and the symmetric Itakura-Saito distance [48].

In this paper, the model comparison issue is addressed by analyzing the behavior of the Kullback-Leibler (KL) divergence and the Rényi divergence between the probability density functions (pdfs) of k consecutive samples of two Gaussian wide-sense stationary (w.s.s.) ARMA processes, when k increases. Their symmetric versions, known as Jeffreys divergence and the symmetric Rényi divergence, are also investigated. More particularly, our purpose is to study the difference between the divergences computed for $k+1$ and k variates and to derive an analytic expression of its limit. This asymptotic increment corresponds to what is called the divergence rate in information theory. Indeed in this field, information measures, such as the entropy, the mutual information and the divergence are often extended to information rates. Thus, different authors have focused their attentions on the divergence rates for stationary Gaussian processes: in [14], Gil first recalls the expression of the KL divergence rate for zero-mean Gaussian processes, initially presented in [20], before giving the Rényi divergence rate. The rates are expressed in terms of integrals depending on functions of the power spectral densities of the processes. Moreover, to obtain their results, the author takes advantage of the theory of Toeplitz matrices and the properties of the asymptotic distribution of the eigenvalues of Toeplitz forms.

In this paper, we get a closed form expression of the divergence rates by exploiting a different approach. Starting from the definition of the divergences, we combine different results on ARMA processes such as the interpretation of the ARMA process as the filtering of a white noise, the Yule-Walker equations [36] and the link between the determinant of the covariance matrix with the AR parameters or the reflection coefficients in order to get the expression of the divergence rates. Among the results we obtain, we will see that the KL divergence rate corresponds to the Itakura–Saito distance, up to a multiplicative factor. In other words, using the KL divergence rate amounts to using the Itakura–Saito distance[1]. As the KL divergence corresponds to the Rényi divergence of order α when α tends to 1, looking at the Réyni divergence rate can hence offer a degree of freedom for ARMA-process comparison. In addition, our results are consistent with the ones Gil obtained for zero-mean Gaussian random processes in [14]. Although this is not visible at the first glance, we will then show that our expression of the divergence rate of the Rényi divergence converges to the divergence rate of the KL when the order α tends to 1. Once again, the proof is mainly based on the properties of the ARMA processes. Finally, we will give

[1] It is of interest to point out this connection as some authors developed some signal processing approches in which they suggest minimizing the Itakura-Saito distance instead of the Kullback-Leibler divergence. This is for instance the case when dealing with the non-negative matrix factorization [29].

some illustrations either based on synthetic data or real data. More particularly the application deals with the characterization of a period of stress by using the divergence rate.

It should be noted that this work is also the opportunity to correct some results we gave in [34] and is a complementary study to the work done on the divergence rate of the Jeffreys divergence between probability density functions of consecutive samples of autoregressive (AR), moving average (MA) or autoregressive moving average (ARMA) processes and autoregressive fractionally integrated moving average (ARFIMA) -noisy or not- in [30, 31, 33, 41].

The remainder of this paper is organized as follows: Section 2 deals with the KL and the Rényi divergence as well as their symmetric versions, their definitions and analytic expressions in the Gaussian case. The expressions of the increments, *i.e.* the differences between the divergences computed for $k+1$ and k variates, are also given. In section 3, various properties of the Gaussian ARMA processes are presented. Even if some results may be well known by some readers, this section is useful as it provides all the necessary information to derive the divergence rates. In section 4, by combining results of sections 2 and 3, the behaviors of the Rényi divergence and the KL divergence are then analyzed for Gaussian ARMA processes. The expressions of their asymptotic increments, *i.e.* their divergence rates, are derived. Connections with Gil's work are then made. It is also shown that the limit of the divergence rate of the Rényi divergence tends to the divergence rate of the KL when the order tends to 1. In section 5, simulations results are provided. An analysis on synthetic data makes it possible to confirm the theoretical analysis. Then, an application on real data is proposed. Finally, two appendices deal with the derivation of the expression of the Rényi divergence in the Gaussian case and the expression of the transfer function associated with the linear combination of two ARMA processes.

2. About Kullback-Leibler and Rényi divergences

2.1. Brief state of the art on the divergences

Divergences are used by several communities in the field of signal and image processing and statistics. On the one hand, with the entropy which can be seen as a measure of information, it is one of the notions that is mainly exploited in information theory. On the other hand, in some applications, divergences can be of interest to compare two time series that can correspond to signals recorded by different sensors. They can also be a relevant tool to detect statistical changes in a signal. In this case, a part of the signal which is known as a reference is compared with another one by using a sliding window. Detecting abrupt changes in time series can be also interest [28]. In image processing, classification can be done thanks to divergences [6, 42]. There is also a need to compare the pdfs for statistical hypothesis tests. Finally, divergences are used in the field of deep learning during the back-propagation step. As a consequence, divergences can be useful in a large number of applications, from econometrics to hydrology, from biomedical signal analysis to radar processing. See for instance [13, 33, 35]. A great deal of interest has been paid to divergences for several decades. The Kullback-Leibler (KL) divergence, also known as KL relative entropy, is one of the most popular divergences [26]. Starting from KL, one can easily define the Jeffreys divergence which is its symmetric version, by interchanging the role of the two distributions to be compared [22]. As for the Jensen-Shannon divergence, it can be deduced by first introducing the distribution mean, then evaluating the KL between each distribution to be compared and the distribution mean and finally computing the mean of both KL. Jensen-Shannon divergence is a particular case of the General Jensen-Shannon divergence, also known as the skewed Jensen-Shannon divergence. In this case, instead of computing the mean of the two KLs, the linear combination of the two KLs is considered. Another generalization of the KL can be considered through α-divergences, which are parameterized by a parameter α. Thus, Basu *et al.* introduced the density power divergences [3]. In this case, it corresponds to the integral of a sum of quantities defined from the pdfs to be compared raised at the value α or $\alpha + 1$. Note that α also appears as a multiplicative factor. Another instance is the Rényi divergence of order α. When α tends to 1, the Rényi divergence tends to KL, whereas $\alpha = \frac{1}{2}$ leads to the Bhattacharyya distance. Moreover, the Rényi divergence of order α is equal to the Chernoff divergence of order α up to a multiplicative constant equal to $\frac{1}{1-\alpha}$. The KL divergence can also be related to an f-divergence, also known as Csiszár divergence. In this case, the divergence is defined as an integral of the product between the second pdf and a convex function f of the ratio of the two pdfs, with $f(1) = 0$. Thus, when the latter is equal to the logarithm, this leads to the KL. The Kay divergence is another example of f-divergences where $f(t) = \ln(\frac{2}{1+t})$ whereas the Tsallis divergence is obtained when $f(t) = \frac{t^{1-\alpha}-1}{\alpha-1}$. It should be noted that the Rényi divergence of order α can be expressed as a function of the logarithm of a f-divergence, provided that $f(t) = t^{1-\alpha} - 1$. Finally, some divergences presented above can be related to Bregman divergences. See [2] for instance.

Currently, there are various types of research activities on divergences that are conducted. As it is difficult to be exhaustive, here are some examples:

1. Some studies deal with closed-form expressions, properties, new types or generalizations of existing divergences [10, 24, 25, 38, 47].

2. Other address the estimation of the divergence from data that can be Gaussian or not [8, 9, 37, 46].

3. Related issues deal with entropy rates, defined as the entropy per unit time [40], [15].

In the next subsection, let us focus our attention on the KL and Rényi divergence as well as symmetric versions.

2.2. Definitions and expression in the Gaussian case

Let $x_{t,1}$ and $x_{t,2}$ be two scalar real Gaussian random processes and $X_{k,1}$ and $X_{k,2}$ the $k \times 1$ column vectors storing k consecutive samples of $x_{t,1}$ and $x_{t,2}$:

$$X_{k,i} = [x_{t,i} \; x_{t-1,i} \; \cdots \; x_{t-k+1,i}]^T \text{ for } i = 1, 2 \tag{1}$$

The pdf of $X_{k,i}$ is given by:

$$p_i(X_{k,i}) = \frac{1}{(\sqrt{2\pi})^k |Q_{k,i}|^{1/2}} \exp\big(-\frac{1}{2}[X_{k,i}-\mu_{k,i}]^T Q_{k,i}^{-1}[X_{k,i}-\mu_{k,i}]\big) \text{ for } i=1,2 \quad (2)$$

with $\mu_{k,i} = E[X_{k,i}]$ the statistical mean, $|Q_{k,i}|$ the determinant of the covariance matrix $Q_{k,i} = E[(X_{k,i}-\mu_{k,i})(X_{k,i}-\mu_{k,i})^T]$ and $E[\cdot]$ the expectation operator. To study the dissimilarities between the random processes, the KL divergence between the joint distributions of k successive values of two random processes can be evaluated [26] and is given by:

$$KL_k^{(1,2)} = \int_{X_k} p_1(X_k) \ln\left(\frac{p_1(X_k)}{p_2(X_k)}\right) \mathrm{d}X_k \quad (3)$$

For Gaussian processes, it can be shown, by substituting $p_1(X_k)$ and $p_2(X_k)$ with the expression (2) and by introducing $\Delta\mu_k = \mu_{k,2} - \mu_{k,1}$ and taking advantages of the properties of the trace of a matrix, that $KL_k^{(1,2)}$ satisfies [39]:

$$KL_k^{(1,2)} = \frac{1}{2}\Big[\mathrm{Tr}\big(Q_{k,2}^{-1}(Q_{k,1} + \Delta\mu_k\Delta\mu_k^T)\big) - k - \ln\frac{|Q_{k,1}|}{|Q_{k,2}|}\Big] \quad (4)$$

where Tr denotes the trace of a matrix.

However, the KL divergence is not symmetric. To address this issue, various approaches can be considered. A first idea is to take the minimum value between $KL_k^{(1,2)}$ and $KL_k^{(2,1)}$. An alternative is to compute the sum $KL_k^{(1,2)} + KL_k^{(2,1)}$. As for Jeffreys divergence, denoted as $JD_k^{(1,2)}$, it aims at computing the mean between $KL_k^{(1,2)}$ and $KL_k^{(2,1)}$. It satisfies:

$$\begin{aligned} JD_k^{(1,2)} &= \frac{1}{2}(KL_k^{(1,2)} + KL_k^{(2,1)}) \\ &= \frac{1}{4}\Big[\mathrm{Tr}\big(Q_{k,2}^{-1}(Q_{k,1} + \Delta\mu_k\Delta\mu_k^T) + Q_{k,1}^{-1}(Q_{k,2} + \Delta\mu_k\Delta\mu_k^T)\big) - 2k\Big] \end{aligned} \quad (5)$$

As mentioned in the introduction, the KL divergence is a specific case of other divergences. Among them, the Rényi divergence (RD) of order α is defined as:

$$RD_k^{(1,2)}(\alpha) = \frac{1}{\alpha-1}\ln\int_{X_k} p_1^{\alpha}(X_k)\, p_2^{1-\alpha}(X_k)\, \mathrm{d}X_k \quad (6)$$

In the above equation, α is a degree of freedom that can be selected by the practitioner. Using L'Hospital rule, one can show that $RD_k^{(1,2)}(\alpha)$ tends to $KL_k^{(1,2)}$ when α tends to 1. $\alpha = \frac{1}{2}$ is the only case when $RD_k^{(1,2)}(\alpha)$ and $RD_k^{(2,1)}(\alpha)$ provide the same value. For $0 < \alpha < 1$, the RD has a skew symmetry since $RD_k^{(1,2)}(\alpha) = \frac{\alpha}{1-\alpha} RD_k^{(2,1)}(1-\alpha)$. Finally, for $0 < \alpha_1 < \alpha_2 < 1$, one has $\frac{\alpha_1}{1-\alpha_1}\frac{1-\alpha_2}{1-\alpha_1}RD_k^{(1,2)}(\alpha_2) \leq RD_k^{(1,2)}(\alpha_1) \leq RD_k^{(1,2)}(\alpha_2)$. In the Gaussian case, combining (2) and (6) and after some mathematical developments, (See Appendix A for proof), one has:

$$RD_k^{(1,2)}(\alpha) = -\frac{1}{2(\alpha-1)}\ln\Big(\frac{|Q_{k,\alpha}|}{|Q_{k,1}|^{1-\alpha}|Q_{k,2}|^{\alpha}}\Big) + \frac{\alpha}{2}\mathrm{Tr}\Big(Q_{k,\alpha}^{-1}\Delta\mu_k\Delta\mu_k^T\Big) \tag{7}$$

where:

$$Q_{k,\alpha} = \alpha Q_{k,2} + (1-\alpha)Q_{k,1} \tag{8}$$

Let us now introduce the symmetric version of the Rényi divergence (SRD). Similarly to the KL divergence, various cases can be considered. In this paper, the definition below is used:

$$\begin{aligned} SRD_k^{(1,2)}(\alpha) &= \frac{1}{2}(RD_k^{(1,2)}(\alpha) + RD_k^{(2,1)}(\alpha)) \\ &= -\frac{1}{4(\alpha-1)}\ln\Big(\frac{|Q_{k,\alpha}||Q_{k,1-\alpha}|}{|Q_{k,1}||Q_{k,2}|}\Big) + \frac{\alpha}{4}\mathrm{Tr}((Q_{k,\alpha}^{-1} + Q_{k,1-\alpha}^{-1})\Delta\mu_k\Delta\mu_k^T) \end{aligned} \tag{9}$$

In the following, let us express the increments of these divergences.

2.3. Analysis of the divergence increments in the Gaussian case

2.3.1. Case of the Kullback-Leibler and Jeffreys divergences

Let us first compute the k^{th} increment of the KL divergence defined as:

$$\Delta KL_k^{(1,2)} = KL_{k+1}^{(1,2)} - KL_k^{(1,2)} \tag{10}$$

Given the expression (4), one has:

$$\begin{aligned} \Delta KL_k^{(1,2)} &= -\frac{1}{2}\ln\Big(\frac{|Q_{k+1,1}|}{|Q_{k,1}|}\frac{|Q_{k,2}|}{|Q_{k+1,2}|}\Big) \\ &+ \frac{1}{2}\Big(\mathrm{Tr}(Q_{k+1,2}^{-1}(Q_{k+1,1} + \Delta\mu_{k+1}\Delta\mu_{k+1}^T)) - \mathrm{Tr}(Q_{k,2}^{-1}(Q_{k,1} + \Delta\mu_k\Delta\mu_k^T) - 1\Big) \end{aligned} \tag{11}$$

In (11), two terms can be analyzed separately: the difference of traces and the logarithm of determinants of covariance matrices.
Let us now look at the increment of the JD. Using (5) and (11), one has:

$$\begin{aligned} \Delta JD_k^{(1,2)} &= \frac{1}{4}\Big[\mathrm{Tr}(Q_{k+1,2}^{-1}Q_{k+1,1} + Q_{k+1,1}^{-1}Q_{k+1,2}) - \mathrm{Tr}(Q_{k,2}^{-1}Q_{k,1} + Q_{k,1}^{-1}Q_{k,2}) \\ &- 2 + \mathrm{Tr}\big((Q_{k+1,2}^{-1} + Q_{k+1,1}^{-1})\Delta\mu_{k+1}\Delta\mu_{k+1}^T\big) - \mathrm{Tr}\big((Q_{k,2}^{-1} + Q_{k,1}^{-1})\Delta\mu_k\Delta\mu_k^T\big)\Big] \\ &= \frac{1}{2}(\Delta KL_k^{(1,2)} + \Delta KL_k^{(2,1)}) \end{aligned} \tag{12}$$

2.3.2. Case of the Rényi divergence and its symmetric version

Let us address the case of the RD and its symmetric version studying the difference between $RD_{k+1}^{(1,2)}(\alpha)$ and $RD_{k}^{(1,2)}(\alpha)$ and the difference between $SRD_{k+1}^{(1,2)}(\alpha)$ and $SRD_{k}^{(1,2)}(\alpha)$ respectively. Given (7) and (9), they are equal to:

$$\Delta RD_k^{(1,2)}(\alpha) = \frac{\alpha}{2}\Big(\mathrm{Tr}\big(Q_{k+1,\alpha}^{-1}\Delta\mu_{k+1}\Delta\mu_{k+1}^T\big) - \mathrm{Tr}\big(Q_{k,\alpha}^{-1}\Delta\mu_k\Delta\mu_k^T\big)\Big) - \frac{1}{2(\alpha-1)}\ln\Big(\frac{|Q_{k,2}|^\alpha}{|Q_{k+1,2}|^\alpha}\frac{|Q_{k+1,\alpha}|}{|Q_{k,\alpha}|}\frac{|Q_{k,1}|^{1-\alpha}}{|Q_{k+1,1}|^{1-\alpha}}\Big) \tag{13}$$

and

$$\Delta SRD_k^{(1,2)}(\alpha) = \frac{\alpha}{4}\Bigg(\mathrm{Tr}\Big((Q_{k+1,\alpha}^{-1}+Q_{k+1,1-\alpha}^{-1})\Delta\mu_{k+1}\Delta\mu_{k+1}^T\Big) - \mathrm{Tr}\Big((Q_{k,\alpha}^{-1}+Q_{k,1-\alpha}^{-1})\Delta\mu_k\Delta\mu_k^T\Big)\Bigg) - \frac{1}{4(\alpha-1)}\ln\Big(\frac{|Q_{k+1,\alpha}||Q_{k+1,1-\alpha}||Q_{k,1}||Q_{k,2}|}{|Q_{k,\alpha}||Q_{k,1-\alpha}||Q_{k+1,1}||Q_{k+1,2}|}\Big) \tag{14}$$

In the following, we suggest analyzing the way these increments evolve when k increases and becomes larger and larger[2] when comparing w.s.s. Gaussian ARMA processes. For this purpose, in section 3, we present the properties of the ARMA processes, especially those dealing with their covariance matrices. Given (11), (12)-(14), two questions have to be addressed: Which interpretation can be given to the difference of traces of matrices which are pre-multiplied by the inverse of a covariance matrix? What is the limit of the logarithm of the ratio between covariance-matrix determinants?

3. About w.s.s. Gaussian ARMA processes

Let us first recall that a real w.s.s. ARMA(p,q) process is described by[3]:

$$x_t = -\sum_{i=1}^{p} a_i x_{t-i} + u_t + \sum_{j=1}^{q} b_j u_{t-j} \tag{15}$$

where $\{a_i\}_{i=1,\dots,p}$ and $\{b_j\}_{j=1,\dots,q}$ are the ARMA parameters and the driving process u_t is a zero-mean w.s.s. Gaussian white process with variance σ_u^2. In this case, x_t is zero-mean. If the w.s.s. ARMA process has a mean equal to μ_x,

[2]In the following, this will be denoted by: $\lim_{k\to+\infty}(.)$. Nevertheless, k remains finite.

[3]For the sake of simplicity in this section, the subscript related to the number (1^{st} or 2^{nd}) of the process is omitted.

two ways can be considered to generate it: either *a posteriori* adding μ_x to x_t or considering a driving process whose mean is equal to $\frac{1+\sum_{i=1}^{p} a_i}{1+\sum_{j=1}^{p} b_j}\mu_x$.
The ARMA process x_t can be seen as the output of an infinite-impulse-response linear filtering whose input is the driving process. The corresponding transfer function $H(z)$ is defined by the poles $\{p_l\}_{l=1,\dots,p}$ assumed to be inside the unit circle in the z-plane to ensure the asymptotic stability and the zeros $\{z_l\}_{l=1,\dots,q}$. Given θ the normalized angular frequency, the corresponding power spectral density (PSD) satisfies:

$$S_x(\theta) = \sigma_u^2 \left|H(e^{j\theta})\right|^2 \tag{16}$$

Let us now recall six properties that will be useful in the rest of the paper.

1. 2^q **ARMA processes of order** (p, q) **have the same pdf:** since the comparison between ARMA processes is based on their pdfs, it is of interest to highlight when two different ARMA processes are undistinguishable by using the divergences presented above *i.e.* when they are characterized by the same pdf. Many possible ARMA models can be associated with the same PSD. This result is known as the spectral factorization theorem [43, 45]. The transfer function of the process is necessarily defined by the same poles. However, as there is no constraint on the zeros, there are 2^q numerators that can be defined depending on whether the zero z_l or its inverse $1/z_l$ is chosen, for $l = 1, ..., q$. Once the numerator and the denominator are defined, the variance of the driving process can be deduced and is equal to $\sigma_u^2 \prod_{l=1}^{q} K_l$ with $K_l = 1$ when the zero z_l is inside the unit circle in the z-plane and $K_l = |z_l|^2$ when it is outside the unit-circle.

2. **Minimum-phase ARMA model:** if there is no zero on the unit disc, it is always possible to determine within the set of ARMA processes described above a minimum-phase ARMA model defined by zeros are inside the unit disc in the z-plane [43]. Therefore, an arbitrary rational PSD can be represented by an ARMA process whose transfer function $H_{min}(z)$ is asymptotically stable and minimum phase. For every non-minimum phase ARMA model, it is possible to consider the equivalent minimum-phase spectral factor, obtained with the following operations: Let $\{z_l\}_{l=1,\dots,m\leq q}$ be the zeros such that $|z_l| > 1$ with $l = 1, \dots, m$. Replace z_l with $1/z_l^*$ for $l = 1, \dots, m$ to get $H_{min}(z)$. Replace the

variance σ_u^2 of the original ARMA model with $\sigma_{u,min}^2 = \sigma_u^2 \prod_{l=1}^{q} K_l$.

3. **Properties of the AR parameters and the variance of the driving process:** a minimum-phase ARMA process is invertible so that it can be represented by an infinite-order AR process [7] as follows:

$$x_t = -\sum_{i=1}^{+\infty} \alpha_i x_{t-i} + u_t. \tag{17}$$

By truncating the above summation, the ARMA process can be approximated by an AR model of finite-order $\tau > \max\,(p, q)$:

$$x_t \approx -\sum_{i=1}^{\tau} \alpha_{i,\tau} x_{t-i} + u_{t,\tau} \tag{18}$$

Let us now briefly recall how to estimate the AR parameters. To this end, let us first define the covariance function r_k of the real ARMA process and introduce X_k the vector collecting k consecutive samples of the ARMA process as well as its corresponding $k \times k$ symmetric Toeplitz covariance matrix Q_k where the element located at the i^{th} row and j^{th} column is given by the covariance function $r_{j-i} = r_{i-j}$ since the process is assumed to be real. When dealing with a w.s.s. ARMA process, $\sum_\tau r_\tau$ is absolutely summable. As a corollary, the Toeplitz correlation Q_k belongs to the Wiener class Toeplitz matrices. According to [17], the Toeplitz covariance matrix is non singular even if the PSD of the process is equal to zero at some frequencies. It is hence invertible. Nevertheless, the infinite-size Toeplitz covariance matrix is no longer invertible when the corresponding transfer function of the ARMA process has unit roots. The covariance function satisfies:

$$\begin{cases} r_0 & = -\sum_{i=1}^{\tau} \alpha_{i,\tau} r_{-i} + \sigma_{u,\tau}^2 \\ r_k & = -\sum_{i=1}^{\tau} \alpha_{i,\tau} r_{k-i}, \quad k = 1, \ldots, \tau \end{cases} \tag{19}$$

where $\sigma_{u,\tau}^2$ is the variance of the driving process $u_{t,\tau}$ of the τ^{th}-order AR model. The second equation of (19) leads to the Yule-Walker equations:

$$\Theta_\tau = -Q_\tau^{-1} \underline{r}_\tau \tag{20}$$

where Θ_τ is the column vector storing the AR parameters $\{\alpha_{i,\tau}\}_{i=1,\ldots,\tau}$ and $\underline{r}_\tau$ is the covariance vector storing the values of the covariance function for lags equal to $1, ..., \tau+1$. As for the variance of the driving process, by using the first equation of (19), one obtains:

$$\sigma_{u,\tau}^2 = r_0 - \underline{r}_\tau^T (Q_\tau^{-1})^T \underline{r}_\tau = r_0 - \underline{r}_\tau^T Q_\tau^{-1} \underline{r}_\tau \quad (21)$$

Remark: The purpose of the Levison-Durbin algorithm is to compute, in a recursive way the coefficients of increasing-order AR models [45]. In this case, the parameters $\{\alpha_{i,\tau}\}_{i=1,...,\tau}$ can be computed from the knowledge of $\{\alpha_{i,\tau-1}\}_{i=1,,...,\tau-1}$. The Levinson-Durbin algorithm is initialized with $\sigma_{u,0}^2 = r_0$ and the variance of the driving process is updated as follows:

$$\sigma_{u,\tau}^2 = (1 - \alpha_{\tau,\tau}^2)\, \sigma_{u,\tau-1}^2 = r_0 \prod_{n=1}^{\tau} (1 - \alpha_{n,n}^2) \quad (22)$$

where the AR parameters appearing in (17) can be considered as the limits of these parameters when the model order τ tends to infinity:

$$\begin{cases} \alpha_i = \lim_{\tau\to+\infty} \alpha_{i,\tau} \qquad \text{for } i = 1, 2, \ldots . \\ \lim_{\tau\to+\infty} \sigma_{u,\tau}^2 = \lim_{\tau\to+\infty} r_0 \prod_{n=1}^{\tau} (1 - \alpha_{n,n}^2) = \sigma_{u,min}^2 \end{cases} \quad (23)$$

The parameter $\alpha_{n,n}$ is known as the n^{th} reflection coefficient and its modulus is such that $0 < |\alpha_{n,n}| < 1$. Its square corresponds to the square of the the partial autocorrelation function[4] (PACF). Therefore, one can have the same reasoning with these quantities.

4. **Properties of the determinant of the covariance matrix:** The LDL factorization of Q_k involves the product between a lower unit triangular matrix L_k and a diagonal matrix D_k defined from the AR parameters and the variance of the driving process of the AR processes whose order varies from 0 to $k-1$. Indeed, it can be obtained by expressing each element of X_k as an AR process

[4] After expressing the process x at times t and $t-n$ as linear combinations of the n values $x_{t-1}, ..., x_{t-n+1}$ and their corresponding residuals, the PACF is defined as the correlation coefficient computed between both residuals.

with a different order using (18). In other words, one has:

$$\begin{cases} L_k^{-1} X_k = U_k \\ L_k^{-1} = \begin{bmatrix} 1 & \alpha_{1,k-1} & \alpha_{2,k-1} & \dots & \alpha_{k-1,k-1} \\ 0 & 1 & \alpha_{1,k-2} & \dots & \alpha_{k-2,k-2} \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \alpha_{1,1} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \\ U_k = [u_{t,k-1} \quad u_{t-1,k-2} \quad \dots \quad u_{t-k+1,0}]^T \end{cases} \tag{24}$$

Therefore, the following equality holds:

$$L_k^{-1} Q_k (L_k^H)^{-1} = D_k = diag(\sigma_{u,k-1}^2 \dots \sigma_{u,0}^2) \tag{25}$$

with $diag(x)$ the diagonal matrix whose main diagonal is x.
After taking the determinant of (25) and showing that $|L_k^H| = |L_k| = 1$ by carrying out the expansion of the determinant with respect to the first column or row, and then doing that again for each smaller determinants, one obtains:

$$|Q_k| = |D_k| = \prod_{n=0}^{k-1} \sigma_{u,n}^2 \tag{26}$$

and consequently by taking into account (22):

$$\frac{|Q_k|}{|Q_{k-1}|} = \sigma_{u,k-1}^2 \tag{27}$$

Note that considering the two remarks in property (iii), the above ratio could be also expressed from the reflection coefficients or the PACFs. Given (23), one has for a minimum-phase ARMA process:

$$\lim_{k \to +\infty} \frac{|Q_k|}{|Q_{k-1}|} = \sigma_{u,min}^2 \tag{28}$$

5. **Premultiplication by the inverse of the covariance matrix:** Let us express the covariance matrices of the two vectors $X_{k,1}$ and $X_{k,2}$ storing k consecutive values of two minimum-phase ARMA processes $x_{t,1}$ and $x_{t,2}$, by using their eigenvalues and eigenvectors:

$$Q_{k,i} = E\left[(X_{k,i} - E[X_{k,i}])(X_{k,i} - E[X_{k,i}])^T\right] = P_{k,i} D_{k,i} P_{k,i}^T \tag{29}$$

where the subscript $i = 1, 2$ defines the i^{th} process under study, $P_{k,i}$ denotes the unitary matrix storing the k eigenvectors of $Q_{k,i}$ and $D_{k,i}$ is the diagonal matrix defined with the k non-null real positive eigenvalues.

Pre-multiplying $X_{k,1} - E[X_{k,1}]$ by $D_{k,1}^{-1/2} P_{k,1}^T$ consists in whitening the process vector. As the process is assumed to be w.s.s. and when k tends to infinity, this amounts to filtering all the samples stored in $X_{k,1} - E[X_{k,1}]$ by the inverse filter defined by the transfer function $H_{1,min}^{-1}(z)$. Similarly, pre-multiplying $X_{k,1} - E[X_{k,1}]$ by $D_{k,2}^{-1/2} P_{k,2}^T$ amounts to filtering the vector $X_{k,1} - E[X_{k,1}]$ by the inverse filter $H_{2,min}^{-1}(z)$. Therefore, one has:

$$\lim_{k \to +\infty} \mathrm{Tr}(Q_{k+1,2}^{-1} Q_{k+1,1}) - \mathrm{Tr}(Q_{k,2}^{-1} Q_{k,1}) = P^{(1,2)} \tag{30}$$

where $P^{(1,2)}$ is the power of the 1^{st} zero-mean ARMA process filtered by $H_{2,min}^{-1}(z)$ associated with the inverse filter associated with the second zero-mean minimum-phase ARMA process.

For the same reason, if the w.s.s ARMA processes are not zero-mean, their means are constant. The vectors $\Delta\mu_{k+1}$ and $\Delta\mu_k$ respectively store $k+1$ and k times the same value $\mu_1 - \mu_2$. Therefore, one has:

$$\begin{aligned} &\lim_{k \to +\infty} \frac{1}{2}\Big(Tr\big(Q_{k+1,2}^{-1} \Delta\mu_{k+1} \Delta\mu_{k+1}^T\big) - Tr\big(Q_{k,2}^{-1} \Delta\mu_k \Delta\mu_k^T\big)\Big) \\ &= \frac{(\mu_1 - \mu_2)^2}{2} |H_{2,min}^{-1}(z)|_{z=1}^2 \end{aligned} \tag{31}$$

6. **Sum of processes:** The sum of two stationary independent ARMA processes $x_{t,1}$ and $x_{t,2}$, respectively of orders (p_1, q_1) and (p_2, q_2), is an ARMA(p, q) process with $p \leq p_1 + p_2$, $q \leq max(p_1 + q_2, p_2 + q_1)$ [16]. As a consequence, the same result holds for the linear combination $x_{t,\alpha} = \sqrt{\alpha}\, x_{t,2} + \sqrt{1-\alpha}\, x_{t,1}$. Multiplying $x_{t,2}$ and $x_{t,1}$ by $\sqrt{\alpha}$ and $\sqrt{1-\alpha}$ affects only the variances of the corresponding driving processes that are multiplied by α and $1 - \alpha$.

In the next section, given the above properties, let us deduce the divergence rate when dealing with Gaussian w.s.s. ARMA processes.

4. Asymptotic analysis of the increment of the divergence for w.s.s. Gaussian ARMA processes

Now we propose to analyze the way the increments of the divergences evolve when k increases and tends to infinity. The results for the KL divergence and its symmetric version are first addressed. Then, we focus our attention on the Rényi divergence and its symmetric version. In each case, the expression of the so-called divergence rate is presented for non-zero mean ARMA processes. As comparing two ARMA processes by means of their pdfs amounts to comparing the corresponding minimum-phase ARMA processes, the properties presented in the section 2 are used. This explains why the divergence rates will depend on the transfer functions of the inverse filters and the variances of the driving processes of the corresponding minimum-phase ARMA processes, *i.e.* for $i = 1, 2$, $H_{i,min}^{-1}(z)$ and $\sigma_{u,i,min}^2 = \sigma_{u,i}^2 \prod_{l=1}^{q_i} K_{l,i}$ with $K_{l,i} = 1$ when the zero of the i^{th} ARMA process $z_{l,i}$ is inside the unit circle and $K_{l,i} = |z_{l,i}|^2$ when it is outside the unit-circle in the z-plane.

4.1. Divergence rate of the Kullback-Leibler divergence

Given (11), (28), (30) and (31), the asymptotic KL increment satisfies:

$$\begin{aligned} \Delta KL^{(1,2)} &= \lim_{k\to+\infty} \Delta KL_k^{(1,2)} \\ &= \frac{1}{2}(P^{(1,2)} - 1) + \frac{(\mu_1 - \mu_2)^2}{2} |H_{2,min}^{-1}(z)|_{z=1}^2 - \frac{1}{2} \ln \frac{\sigma_{u,1,min}^2}{\sigma_{u,2,min}^2} \end{aligned} \tag{32}$$

The divergence rate depends on three terms: the first one is related to $P^{(1,2)}$ the power of the first process filtered by the inverse filter associated with the second process. The second term takes into account the difference of the continuous parts of both processes when it is filtered by the inverse filter associated with the second process. The last term deals with the variances of the driving processes associated with the minimum-phase ARMA processes.

When both processes have the same mean, the term $\frac{(\mu_1-\mu_2)^2}{2}|H_{2,min}^{-1}(z)|_{z=1}^2$ vanishes. When the variances $\sigma_{u,1,min}^2$ and $\sigma_{u,2,min}^2$ are equal, the logarithm is equal to 0.

When the processes are zero-mean and if $S_1(\theta)$ and $S_2(\theta)$ denote the PSD of the first and the second process to be compared, the expression of the KL divergence rate for Gaussian processes[5] given in [20] and [14] is:

$$\Delta KL^{(1,2)} = \frac{1}{4\pi}\int_{-\pi}^{\pi}\Big(\frac{S_1(\theta)}{S_2(\theta)} - 1 - \ln\frac{S_1(\theta)}{S_2(\theta)}\Big)d\theta \tag{33}$$

Firstly, we can notice that this expression also corresponds to the Itakura–Saito distance up to a multiplicative factor equal to $\frac{1}{2}$. Moreover, using the notation $P^{(1,2)}$, the KL divergence rate (33) can be rewritten and decomposed as follows:

$$\Delta KL^{(1,2)} = \frac{1}{2}(P^{(1,2)} - 1) - \frac{1}{2}\big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln S_1(\theta)d\theta - \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln S_2(\theta)d\theta\big) \tag{34}$$

In (34), the ARMA-process PSD can be expressed from its transfer function and its driving process, but also from the transfer function and the driving process of the corresponding minimum-phase ARMA process. This leads to:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln S_i(\theta)d\theta = \ln\sigma^2_{u,i,min} + \frac{1}{\pi}\int_{-\pi}^{\pi}\ln|H_{i,min}(\theta)|d\theta \text{ for } i = 1,2 \tag{35}$$

In (35), the quantity $\frac{2}{\pi}\int_{-\pi}^{\pi}\ln|H_{i,min}(\theta)|d\theta$ can be interpreted as the cepstrum $c_i(0) = \frac{2}{\pi}\int_{-\pi}^{\pi}\ln|H_{i,min}(\theta)|e^{jn\theta}d\theta|_{n=0}$ of the impulse response associated with the minimum-phase ARMA process. As the cepstrum $c_i(n) = \frac{\hat{h}_i(n)+\hat{h}_i(-n)}{2}$ can be expressed from the complex cepstrum $\hat{h}_i(n)$ of the impulse response, (35) becomes[6]:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln S_i(\theta)d\theta = \ln\sigma^2_{u,i,min} + \hat{h}_i(0) + \hat{h}_i(0) = \ln\sigma^2_{u,i,min} \tag{36}$$

Substituting (36) into (34), one retrieves the result we obtain in (32) by combining different properties of the Gaussian ARMA processes.

4.2. Divergence rate of the Jeffreys divergence

The divergence rate can be obtained in two manners: either by looking at the limit of (12) when k tends to infinity by using (28) and (30), or by combining

[5] With no *a priori* made on the correlation properties.

[6] When $H_i(z) = \frac{\prod_{l=1}^{q_i}(1-z_{l,i}z^{-1})}{\prod_{l=1}^{p_i}(1-p_{l,i}z)}$ with $|z_l| < 1$ and $|p_l| < 1$, the complex cepstrum is:

$$\begin{cases} \hat{h}_i(n) = -\sum_{l=1}^{q_i}\frac{z_{l,i}^n}{n} + \sum_{l=1}^{p_i}\frac{p_{l,i}^n}{n} \text{ for } n > 0 \\ \hat{h}_i(n) = 0 \text{ for } n \leq 0 \end{cases}$$

(12) and (32):

$$\Delta JD^{(1,2)} = \lim_{k\to+\infty} \Delta JD_k^{(1,2)} = \frac{1}{2}(\Delta KL^{(1,2)} + \Delta KL^{(2,1)}) \tag{37}$$
$$= \frac{1}{4}(P^{(1,2)} + P^{(2,1)} - 2) + \frac{(\mu_1-\mu_2)^2}{4}(|H_{2,min}^{-1}(z)|_{z=1}^2 + |H_{1,min}^{-1}(z)|_{z=1}^2)$$

When both processes have the same mean, the term $\frac{(\mu_1-\mu_2)^2}{4}(|H_{2,min}^{-1}(z)|_{z=1}^2 + |H_{1,min}^{-1}(z)|_{z=1}^2)$ vanishes. Once again, the results provided in (37) is consistent with the expression we would obtain by using the KL divergence rate in (34) for zero-mean stationary Gaussian processes and the definition of the Jeffreys divergence rate.

4.3. Divergence rate of the Rényi divergence

Given (13) and the results presented in section 3, the divergence rate of the Rényi divergence of order α is equal to:

$$\Delta RD^{(1,2)}(\alpha) = \lim_{k\to+\infty} \Delta RD_k^{(1,2)}(\alpha) \tag{38}$$
$$= \alpha\frac{(\mu_1-\mu_2)^2}{2}|H_{\alpha,min}^{-1}(z)|_{z=1}^2 - \frac{1}{2(\alpha-1)}\ln\Big(\frac{\sigma_{u,\alpha}^2\prod_{l=1}^{q_\alpha}K_{l,\alpha}}{(\sigma_{u,2}^2\prod_{l=1}^{q_2}K_{l,2})^\alpha(\sigma_{u,1}^2\prod_{l=1}^{q_1}K_{l,1})^{1-\alpha}}\Big)$$
$$= \alpha\frac{(\mu_1-\mu_2)^2}{2}|H_{\alpha,min}^{-1}(z)|_{z=1}^2 - \frac{1}{2(\alpha-1)}\ln\Big(\frac{\sigma_{u,\alpha,min}^2}{(\sigma_{u,2,min}^2)^\alpha(\sigma_{u,1,min}^2)^{1-\alpha}}\Big)$$

where $\sigma_{u,\alpha,min}^2$ and $H_{\alpha,min}^{-1}(z)$ are the variance of the driving process and the inverse filter associated with the minimum-phase ARMA process x_k^α defined as the linear combination of the two ARMA processes: $x_t^\alpha = \sqrt{\alpha}\, x_{t,2} + \sqrt{1-\alpha}\, x_{t,1}$ whose order is q_α. $K_{l,\alpha}$ is similarly defined as K_l but it is related to x_t^α.
When both processes have the same mean, the term $\alpha\frac{(\mu_1-\mu_2)^2}{2}|H_\alpha^{-1}(z)|_{z=1}^2$ vanishes. In this case, the divergence rate depends on $\sigma_{u,1,min}^2$, $\sigma_{u,2,min}^2$ and $\sigma_{u,\alpha,min}^2$. One could think that this divergence rate is independent of the dynamical properties of the ARMA processes, such as their ARMA parameters or their poles and zeros, and hence would not be a good measure to evaluate the dissimilarity between two ARMA. However, $\sigma_{u,1,min}^2$ and $\sigma_{u,2,min}^2$ may depend on the zeros of the ARMA process when the zeros have their moduli larger than 1. Moreover, and this is the most important reason, $\sigma_{u,\alpha,min}^2$ depend on the transfer functions and the variances of the driving processes of the two processes to be compared, as mentioned in the appendix B.

When the processes are zero-mean, the expression of the Rényi divergence rate that Gil [14] obtained for zero-mean Gaussian processes by using a theorem related to the asymptotic distribution of the eigenvalues of Toeplitz form is the following:

$$\Delta RD^{(1,2)} = -\frac{1}{2(\alpha-1)}\Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\big((1-\alpha)+\alpha\frac{S_2(\theta)}{S_1(\theta)}\big)d\theta - \frac{\alpha}{2\pi}\int_{-\pi}^{\pi}\ln\frac{S_2(\theta)}{S_1(\theta)}d\theta\Big) \quad (39)$$

The latter can be re-expressed as follows:

$$\begin{aligned}\Delta RD^{(1,2)} &= -\frac{1}{2(\alpha-1)}\frac{1}{2\pi}\Big(\int_{-\pi}^{\pi}\ln\big(\frac{(1-\alpha)S_1(\theta)+\alpha S_2(\theta)}{S_2^{\alpha}(\theta)S_1^{1-\alpha}(\theta)}\big)d\theta\Big) \quad (40)\\ &= -\frac{1}{2(\alpha-1)}\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\big(\frac{S_\alpha(\theta)}{S_2^{\alpha}(\theta)S_1^{1-\alpha}(\theta)}\big)d\theta\end{aligned}$$

Following a similar reasoning as the one we made for the KL divergence rate based on the cepstrum, one has:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\big(\frac{S_\alpha(\theta)}{S_2^{\alpha}(\theta)S_1^{1-\alpha}(\theta)}\big)d\theta = \ln\Big(\frac{\sigma_{u,\alpha}^2\prod_{l=1}^{q_\alpha}K_{l,\alpha}}{(\sigma_{u,2}^2\prod_{l=1}^{q_2}K_{l,2})^{\alpha}(\sigma_{u,1}^2\prod_{l=1}^{q_1}K_{l,1})^{1-\alpha}}\Big) \quad (41)$$

Our result is hence consistent with the one obtained by Gil while we do not address the problem in the same way.

4.4. Divergence rate of the symmetric version of the Rényi Divergence

Using (14), (30), (31), one obtains:

$$\Delta SRD^{(1,2)}(\alpha) = \lim_{k\to+\infty}\Delta SRD_k^{(1,2)}(\alpha) = \frac{1}{2}\big(\Delta RD^{(1,2)}(\alpha)+\Delta RD^{(2,1)}(\alpha)\big) \quad (42)$$

$$= \alpha\frac{(\mu_1-\mu_2)^2}{4}(|H_{\alpha,min}^{-1}(z)|_{z=1}^2 + |H_{1-\alpha,min}^{-1}(z)|_{z=1}^2) - \frac{1}{4(\alpha-1)}\ln\Big(\frac{\sigma_{u,\alpha,min}^2\sigma_{u,1-\alpha,min}^2}{\sigma_{u,1,min}^2\sigma_{u,2,min}^2}\Big)$$

When both processes have the same mean, the term $\alpha\frac{(\mu_1-\mu_2)^2}{4}(|H_{\alpha,min}^{-1}(z)|_{z=1}^2 + |H_{1-\alpha,min}^{-1}(z)|_{z=1}^2)$ vanishes.

5. Link between the KL divergence rate and the RD divergence rate

The divergence rate of the KL is expected to be the limit of the divergence of the RD when α tends to 1. When looking at (32) and (38) at the first glance, it does not seem to be the case. In this section, we propose to give a proof by taking advantage of the properties related to ARMA processes. Note that Gil [14] did it for the expressions (33) and (39) he obtained for Gaussian processes.

First of all, let us look at the limit of the first term in (38). Since $x_{t,\alpha} = \sqrt{\alpha}\, x_{t,2} + \sqrt{1-\alpha}\, x_{t,1}$, the inverse filter $H^{-1}_{\alpha,min}(z)$ tends to $H^{-1}_{2,min}(z)$ when α tends to 1. Therefore, one has:

$$\lim_{\alpha\to 1} \alpha \frac{(\mu_1-\mu_2)^2}{2} |H^{-1}_{\alpha,min}(z)|^2_{z=1} = \frac{(\mu_1-\mu_2)^2}{2} |H^{-1}_{2,min}(z)|^2_{z=1} \tag{43}$$

Concerning the second term in (38), *i.e.* $-\frac{1}{2(\alpha-1)} \ln\Big(\frac{\sigma^2_{u,\alpha,min}}{(\sigma^2_{u,2,min})^{\alpha}(\sigma^2_{u,1,min})^{1-\alpha}}\Big)$, we suggest using L'Hospital rule[7] as done in the proof to show that the Rényi divergence tends to the KL divergence when the order α tends to 1. For this purpose, let us introduce the two following functions:

$$g(\alpha) = -2(\alpha-1) \text{ and } f(\alpha) = \ln\Big(\frac{\sigma^2_{u,\alpha}\prod_{l=1}^{q_\alpha} K_{l,\alpha}}{(\sigma^2_{u,2}\prod_{l=1}^{q_2} K_{l,2})^{\alpha}(\sigma^2_{u,1}\prod_{l=1}^{q_1} K_{l,1})^{1-\alpha}}\Big) \tag{44}$$

Their ratio is hence equal to the quantity to be analyzed. In this case, their derivates with respect to α are equal to:

$$\frac{d}{d\alpha} g(\alpha) = -2 \text{ and } \frac{d}{d\alpha} f(\alpha) = \frac{d\ln\Big(\sigma^2_{u,\alpha}\prod_{l=1}^{q_\alpha} K_{l,\alpha}\Big)}{d\alpha} + \ln\Big(\frac{\sigma^2_{u,1}\prod_{l=1}^{q_1} K_{l,1}}{\sigma^2_{u,2}\prod_{l=1}^{q_2} K_{l,2}}\Big) \tag{45}$$

Let us focus our attention on the first term $\frac{d}{d\alpha}\ln\Big(\sigma^2_{u,\alpha}\prod_{l=1}^{q_\alpha} K_{l,\alpha}\Big)$. In the following, we assume that the second process has no zero on the unit-circle [8]. In this case, the minimum-phase ARMA process whose covariance matrix is equal to $Q_{k,\alpha} = \alpha Q_{k,2} + (1-\alpha) Q_{k,1}$ can be expressed as an infinite-order AR process whose driving process variance and AR parameters satisfy the Yule-Walker equations. By applying the properties of the AR processes presented in section 3 to $x_{t,\alpha}$, where $\Theta_{k,\alpha} = -Q^{-1}_{k,\alpha}\underline{r}_{k,\alpha}$ is the column vector storing the AR parameters of $x_{t,\alpha}$ and $\underline{r}_{k,\alpha}$ is the column vector storing the value of the covariance function from 1 to k, one has:

$$\sigma^2_{u,\alpha,min} = \lim_{k\to+\infty} (r_{0,\alpha} + \underline{r}^T_{k,\alpha}\Theta_{k,\alpha}) = r_{0,\alpha} - \lim_{k\to+\infty} \underline{r}^T_{k,\alpha} Q^{-1}_{k,\alpha}\underline{r}_{k,\alpha} \tag{46}$$

---

[7] This rule states that if the functions f and g are differentiable on an open interval except possibly at a point c, if the following three properties are satisfied: 1/ $\lim_{\alpha\to c} f(\alpha) = 0$ and $\lim_{\alpha\to c} g(\alpha) = 0$ 2/ $g'(\alpha) \neq 0$ for any α in the open interval except c 3/ if $\lim_{\alpha\to c} \frac{f'(\alpha)}{g'(\alpha)}$ exists, then $\lim_{\alpha\to c} \frac{f(\alpha)}{g(\alpha)} = \lim_{\alpha\to c} \frac{f'(\alpha)}{g'(\alpha)}$.

[8] This necessarily means that $x_{t,\alpha}$ and $x_{t,1-\alpha}$ correspond to an ARMA processes whose transfer functions have no zero on the unit-circle. Indeed, if the second process has no zero on the unit circle, the corresponding PSD is never equal to 0. Therefore, the PSDs of $x_{t,\alpha}$ and $x_{t,16\alpha}$ cannot be null at a frequency.

In the above equation (46), one can replace $r_{0,\alpha}$ by $\alpha r_{0,2} + (1-\alpha) r_{0,1}$ and take advantage of the matrix inversion lemma[9] to express $Q_{k,\alpha}^{-1}$ as follows:

$$(\alpha Q_{k,2} + (1-\alpha) Q_{k,1})^{-1} = \frac{1}{\alpha} Q_{k,2}^{-1} - \frac{1-\alpha}{\alpha} Q_{k,2}^{-1} (\alpha Q_{k,1}^{-1} + (1-\alpha) Q_{k,2}^{-1})^{-1} Q_{k,2}^{-1} \quad (47)$$

Finally, one has also:

$$\underline{r}_{k,\alpha} = \alpha \underline{r}_{k,2} + (1-\alpha) \underline{r}_{k,1} \quad (48)$$

In (46), let us now rewrite the variance of the driving process by taking into account (47) and (48):

$$\sigma_{u,\alpha,min}^2 = \alpha r_{0,2} + (1-\alpha) r_{0,1} - \lim_{k \to +\infty} \Big(\big(\underline{r}_{k,2} + (1-\alpha) \underline{r}_{k,1} \big)^T \times \quad (49)$$
$$\Big(\frac{1}{\alpha} Q_{k,2}^{-1} - \frac{1-\alpha}{\alpha} Q_{k,2}^{-1} (\alpha Q_{k,1}^{-1} + (1-\alpha) Q_{k,2}^{-1})^{-1} Q_{k,2}^{-1} \Big) \big(\underline{r}_{k,2} + (1-\alpha) \underline{r}_{k,1} \big) \Big)$$

Our goal is to obtain an analytic expression of the following limit:

$$\lim_{\alpha \to 1} \frac{d}{d\alpha} \ln \sigma_{u,\alpha,min}^2 = \lim_{\alpha \to 1} \frac{1}{\sigma_{u,\alpha,min}^2} \frac{d}{d\alpha} \sigma_{u,\alpha,min}^2 \quad (50)$$

Given (50), let us first look at the limit of $\sigma_{u,\alpha,min}^2$ when α tends to 1. One has:

$$\lim_{\alpha \to 1} \sigma_{u,\alpha,min}^2 = \lim_{k \to +\infty} (r_{0,2} - \underline{r}_{k,2}^T Q_{k,2}^{-1} \underline{r}_{k,2}) = \sigma_{u,2,min}^2 \quad (51)$$

Let us now calculate $\frac{d}{d\alpha} \sigma_{u,\alpha,min}^2$. When the right-hand side of the equality in (49) is developed, it consists of ten terms that depend on α. They are listed in the Table 1 as well as their derivatives with respect to α. Therefore, by taking into account that $\sigma_{u,2,min}^2 = \lim_{k \to +\infty} r_{0,2} - \underline{r}_{k,2}^T Q_{k,2}^{-1} \underline{r}_{k,2}$, the derivate of $\sigma_{u,\alpha,min}^2$ with respect to α when α tends to 1 can be expressed as follows:

$$\lim_{\alpha \to 1} \frac{d}{d\alpha} \sigma_{u,\alpha,min}^2 = \lim_{k \to +\infty} \Big(r_{0,2} - r_{0,1} - \underline{r}_{k,2}^T Q_{k,2}^{-1} \underline{r}_{k,2} - \underline{r}_{k,2}^T Q_{k,2}^{-1} Q_{k,1} Q_{k,2}^{-1} \underline{r}_{k,2} \quad (52)$$
$$+ \underline{r}_{k,2}^T Q_{k,2}^{-1} \underline{r}_{k,1} + \underline{r}_{k,1}^T Q_{k,2}^{-1} \underline{r}_{k,2} \Big)$$
$$= \lim_{k \to +\infty} \Big(\sigma_{u,2,min}^2 - r_{0,1} - \underline{r}_{k,2}^T Q_{k,2}^{-1} Q_{k,1} Q_{k,2}^{-1} \underline{r}_{k,2} + \underline{r}_{k,2}^T Q_{k,2}^{-1} \underline{r}_{k,1} + \underline{r}_{k,1}^T Q_{k,2}^{-1} \underline{r}_{k,2} \Big)$$

Moreover, given the Yule-Walker equation applied to the second process, *i.e.* $\Theta_{k,2} = -Q_{k,2}^{-1} \underline{r}_{x,2}$, this leads to:

$$\lim_{\alpha \to 1} \frac{d}{d\alpha} \ln \sigma_{u,\alpha,min}^2 = 1 - \lim_{k \to +\infty} \frac{1}{\sigma_{u,2,min}^2} \Big(r_{0,1} + \Theta_{k,2}^T Q_{k,1} \Theta_{k,2} + \Theta_{k,2}^T \underline{r}_{k,1} + \underline{r}_{k,1}^T \Theta_{k,2} \Big)$$
$$= 1 - P^{(1,2)}$$

[9] $(U+V)^{-1} = U^{-1} - U^{-1}(V^{-1} + U^{-1})^{-1} U^{-1}$.

Terms	$\lim_{\alpha\to 1}\frac{d}{d.\alpha}$
$\alpha r_{0,2}$	$r_{0,2}$
$(1-\alpha)r_{0,1}$	$-r_{0,1}$
$-\alpha \underline{r}_{k,2}^T Q_{k,2}^{-1}\underline{r}_{k,2}$	$-\underline{r}_{x,2}^T Q_{k,2}^{-1}\underline{r}_{k,2}$
$\alpha(1-\alpha)\underline{r}_{k,2}^T\Big(Q_{k,2}^{-1}(\alpha Q_{k,1}^{-1}+(1-\alpha)Q_{k,2}^{-1})^{-1}Q_{k,2}^{-1}\Big)\underline{r}_{k,2}$	$-\underline{r}_{k,2}^T Q_{k,2}^{-1}Q_{k,1}Q_{k,2}^{-1}\underline{r}_{k,2}$
$-(1-\alpha)\underline{r}_{k,2}^T Q_{k,2}^{-1}\underline{r}_{k,1}$	$\underline{r}_{k,2}^T Q_{k,2}^{-1}\underline{r}_{k,1}$
$(1-\alpha)^2\underline{r}_{k,2}^T\Big(Q_{k,2}^{-1}(\alpha Q_{k,1}^{-1}+(1-\alpha)Q_{k,2}^{-1})^{-1}Q_{k,2}^{-1}\Big)\underline{r}_{k,1}$	0
$-(1-\alpha)\underline{r}_{k,1}^T Q_{k,2}^{-1}\underline{r}_{k,2}$	$\underline{r}_{k,1}^T Q_{k,2}^{-1}\underline{r}_{k,2}$
$(1-\alpha)^2\underline{r}_{k,1}^T\Big(Q_{k,2}^{-1}(\alpha Q_{k,1}^{-1}+(1-\alpha)Q_{k,2}^{-1})^{-1}Q_{k,2}^{-1}\Big)\underline{r}_{k,2}$	0
$-\frac{(1-\alpha)^2}{\alpha}\underline{r}_{k,1}^T Q_{k,2}^{-1}\underline{r}_{k,1}$	0
$\frac{(1-\alpha)^3}{\alpha}\underline{r}_{k,1}^T\Big(Q_{k,2}^{-1}(\alpha Q_{k,1}^{-1}+(1-\alpha)Q_{k,2}^{-1})Q_{k,2}^{-1}\Big)\underline{r}_{k,1}$	0

Table 1: Terms and the limits of their derivates with respect to α when α tends to 1

Indeed, the 1^{st} process is filtered by the BIBO-stable minimum phase inverse filter associated with the second process. The latter can be expressed as an infinite-order AR process characterized by its set of AR parameters $\Theta_2 = \lim_{k\to+\infty}\Theta_{k,2}$. The transfer function is equal to $\frac{1}{\sigma_{u,2,min}}\Big(1+\sum_{i=1}^{+\infty}a_{i,2}z^{-i}\Big)$. Let y be the filtered output of the inverse filter associated with the second process when the filter input is the first process x_1. One has:

$$
\begin{aligned}
y_t &= \frac{1}{\sigma_{u,2,min}}\sum_{i=0}^{+\infty}\alpha_{i,2}x_{t-i,1} &\quad (53)\\
&= \frac{1}{\sigma_{u,2,min}}x_{t,1}+\frac{1}{\sigma_{u,2,min}}\sum_{i=1}^{+\infty}\alpha_{i,2}x_{t-i,1} = \frac{1}{\sigma_{u,2,min}}x_{t,1}+\frac{1}{\sigma_{u,2,min}}\lim_{k\to+\infty}\Theta_{k,2}^T X_{k,1}
\end{aligned}
$$

Therefore, the correlation function satisfies for a lag equal to 0 (*i.e.* the power):

$$
\begin{aligned}
P^{(1,2)} = E[y_t^2] &= \frac{1}{\sigma_{u,2,min}^2}E[\lim_{k\to+\infty}(x_{t,1}+\Theta_{k,2}^T X_{k,1})(x_{t,1}+X_{k,1}^T\Theta_{k,2})] &\quad (54)\\
&= \frac{1}{\sigma_{u,2,min}^2}\lim_{k\to+\infty}\Big(r_{0,1}+\Theta_{k,2}^T Q_{k,1}\Theta_{k,2}+\Theta_{k,2}^T\underline{r}_{k,1}+\underline{r}_{k,1}^T\Theta_{k,2}\Big)
\end{aligned}
$$

Using (43) and the above result, one retrieves the KL divergence rate.

6. Additional remarks

In the above section, $H_2(z)$ was assumed to have a minimum phase when dealing with the KL and the RD. The same assumption was made on $H_1(z)$ for the JD and the SRD. In the following, let us look at specific cases.

6.1. *About specific cases of the KL divergence rate and the RD divergence rate*

Let us analyze what happens on the divergence rates when $H_2(z)$ has a zero on the unit circle in the z-plane. To this end, let us define the process $t(k)$ which corresponds to the output of the inverse filter $H_2^{-1}(z)$ whose input was the 1^{st} process. Its power is equal to $P^{(1,2)}$ and its z-transform satisfies:

$$U_1(z)H_1(z)H_2^{-1}(z) = T(z) \tag{55}$$

If $H_1(z)$ has the same unit-zero as $H_2(z)$, then $P^{(1,2)}$ is finite. Otherwise, it will be infinite. Therefore, according to (32) the KL divergence rate will go to infinity. In addition, due to the term $\frac{(\mu_1-\mu_2)^2}{2}|H_2^{-1}(z)|^2_{z=1}$ in (32), the KL divergence rate will go to infinity if $H_2(z)$ has a zero equal to 1 and $\mu_1 - \mu_2 \neq 0$. As for the RD rate, it may go to infinite only when $\mu_1 - \mu_2 \neq 0$ provided that $H_1(z)$ and $H_2(z)$ have a common zero equal to 1. Indeed, this is the only case where $H_\alpha(z)$ has a zero equal to 1.
Therefore, using the RD rate reduces the risk to have an infinite rate.

6.2. *About specific cases of the JD rate and the SRD divergence rate*

Following the same reasoning as above, the JD rate may go to infinite if:
1. the powers $P^{(1,2)}$ and $P^{(2,1)}$ go to infinite. This happens if one process has a zero on the unit-circle that is not shared by the second one.
2. when $\mu_1 - \mu_2 \neq 0$, if $H_1(z)$ and/or $H_2(z)$ has a zero equal to 1. As for the SRD rate, it may go to infinite only when $\mu_1 - \mu_2 \neq 0$ and provided that $H_1(z)$ and $H_2(z)$ have a common zero equal to 1. Indeed, this is the only case where $H_\alpha(z)$ and $H_{1-\alpha}(z)$ have a zero equal to 1. Therefore, the corresponding inverse filters would have an infinite gain at the null frequency.

7. Applications

In this section, our purpose is twofold: first of all, we propose to check if the theoretical results we have presented in the previous sections are confirmed by simulation results. Then, a practical case is presented.

7.1. Simulation results confirming the theory: sensitivity of the estimation of the divergence rate

To confirm the theoretical results, the theoretical divergence rates given by (32), (37), (38) and (42) are compared with the increments of the divergences estimated from different realizations of two wide-sense-stationary real ARMA processes of orders (2,4). To analyze the consistency of the results, the mean and the variance of the normalized error, *i.e.* the absolute value of the difference between the the divergence rate and its estimate which is then divided by the divergence rate, are computed. Two arbitrarily-chosen examples are first given for illustrations. Then, an analysis is based on ARMA processes whose poles and zeros are randomly drawn.

7.1.1. Example 1: comparing ARMA(2,4) processes with very different spectra

Let us consider the two ARMA(2,4) processes $x_{t,1}$ $x_{t,2}$ defined by the following zeros, poles, driving noise variances and means: $z_{1,1} = 3\,exp\,(j\frac{2\pi}{3})$, $z_{2,1} = z_{1,1}^*$, $z_{3,1} = \frac{1}{3}\,exp\,(j\frac{\pi}{3})$, $z_{4,1} = z_{1,3}^*$ and $p_{1,1} = 0.7\,exp(j\frac{\pi}{9})$, $p_{2,1} = p_{1,1}^*$, $\sigma_{u,1}^2 = 1$, $\mu_1 = 1$. $z_{1,2} = 0.85\,exp\,(j\frac{\pi}{6})$, $z_{2,2} = z_{1,2}^*$, $z_{3,2} = 0.8\,exp(j\frac{5\pi}{6})$, $z_{4,2} = z_{3,2}^*$ and $p_{1,2} = 0.4\,exp\,(j\frac{\pi}{4})$, $p_{2,2} = p_{1,2}^*$, $\sigma_{u,2}^2 = 9$. $\mu_2 = 0.5$. Fig. 1 shows the periodograms and the power spectral densities of the two ARMA processes. Due to the zeros $z_{1,1}$ and $z_{2,1}$, the first ARMA process is clearly a non-minimum phase ARMA process. The comparison can be performed by considering the corresponding minimum-phase spectral factor, which is characterized by the zeros $z_{1,1} = \frac{1}{3}\,exp\,(j\frac{2\pi}{3}), z_{2,1} = z_{1,1}^*$.

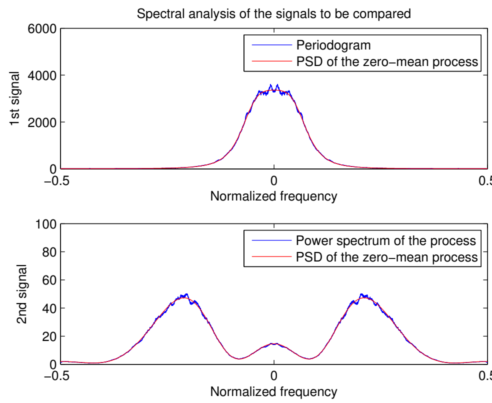

Figure 1: Periodogram of one realization of each ARMA process and the corresponding PSDs, Example 1

For each ARMA model, 20 realizations of 100000 samples have been considered. For each realization, the covariance matrices and the means are first estimated. Then, the estimations of the divergences for consecutive samples are computed by using (4), (5), (7) and (9) for α equal to 0.99, 0.995, 0.999 and 0.9999. Finally, the divergence increments are computed. The evolutions of the JD and SRD divergences and their increments when k increases are presented in Fig. 2. Since KL and RD exhibit similar behaviours, the corresponding figures have not been reported for the sake of space.

One can notice that whatever the realizations of the ARMA processes, the estimations of the increments converge to the theoretical divergence rates.

In Tables 2 and 3, the theoretical divergence rates are compared with the last computed increments ($k = 60$) that are considered as estimates of the divergence rates. The means and variance of the normalized errors are small. This illustrates that the theoretical results are consistent with the experimental ones. As expected, the value of α has an influence on the SRD rate. As α approaches 1, the rates of the RD and the SRD respectively converge to the rates of the KL and the JD. In addition, the larger α, the larger the divergence rate.

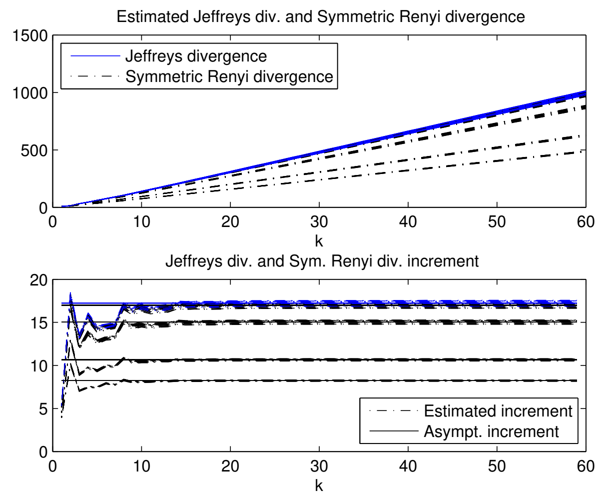

Figure 2: Estimated Jeffreys divergence (blue) and symmetric Rényi divergence (black) and the resulting increments for Example 1 (20 realizations). For the symmetric Rényi divergence, α is equal to 0.99, 0.995, 0.999 and 0.9999

Type of divergence	True	Last estimated increment		Normalized error	
	div. rate	Mean	Variance	Mean	Variance
$KL^{(1,2)}$	33.691	33.812	0.18172	1.12E-02	4.07E-05
$RD^{(1,2)}$ ($\alpha = 0.9999$)	33.167	33.284	0.16987	1.10E-02	3.95E-05
$RD^{(1,2)}$ ($\alpha = 0.999$)	29.314	29.403	0.10171	9.57E-03	3.15E-05
$RD^{(1,2)}$ ($\alpha = 0.995$)	20.543	20.589	2.63E-02	6.99E-03	1.62E-05
$RD^{(1,2)}$ ($\alpha = 0.99$)	15.692	15.723	1.02E-02	5.77E-03	1.07E-05
$KL^{(2,1)}$	0.79343	0.79444	3.95E-06	2.13E-03	3.20E-06
$RD^{(2,1)}$ ($\alpha = 0.9999$)	0.79341	0.79442	3.95E-06	2.13E-03	3.20E-06
$RD^{(2,1)}$ ($\alpha = 0.999$)	0.79329	0.79430	3.95E-06	2.13E-03	3.20E-06
$RD^{(2,1)}$ ($\alpha = 0.995$)	0.79272	0.79373	3.95E-06	2.13E-03	3.20E-06
$RD^{(2,1)}$ ($\alpha = 0.99$)	0.79201	0.79302	3.94E-06	2.14E-03	3.19E-06

Table 2: Statistics on the estimations of divergence rates for different symmetric divergences, based on 20 realizations of the processes of example 1

Type of divergence	True div. rate	Estimated div. rate Mean	Estimated div. rate Variance	Normalized error Mean	Normalized error Variance
JD	17.24	17.30	4.55E-02	1.10E-02	3.91E-05
SRD ($\alpha = 0.9999$)	16.98	17.03	4.26E-02	1.08E-02	3.79E-05
SRD ($\alpha = 0.999$)	15.05	15.10	2.55E-02	9.33E-03	3.02E-05
SRD ($\alpha = 0.995$)	10.66	10.69	6.63E-03	6.78E-03	1.51E-05
SRD ($\alpha = 0.99$)	8.24	8.25	2.60E-03	5.54E-03	9.91E-06

Table 3: Statistics on the estimations of divergence rates for different symmetric divergences, based on 20 realizations of the processes of example 1

7.1.2. Example 2: comparing ARMA(2,4) processes with rather-similar spectra

Let us now consider the zero-mean ARMA(2,4) processes $x_{t,1}$, $x_{t,2}$ defined by the following zeros, poles and driving noise variances: $z_{1,1} = 3\,exp\,(j\frac{2\pi}{3})$, $z_{2,1} = z_{1,1}^*$, $z_{3,1} = \frac{1}{3}\,exp\,(j\frac{\pi}{3})$, $z_{4,1} = z_{1,3}^*$ and $p_{1,1} = 0.7\,exp(j\frac{4\pi}{9})$, $p_{2,1} = p_{1,1}^*$, $\sigma_{u,1}^2 = 1$. $z_{1,2} = 3\,exp\,(j\frac{2\pi}{3})$, $z_{2,1} = z_{1,1}^*$,$z_{3,1} = \frac{1}{3}\,exp\,(j\frac{\pi}{3})$, $z_{4,1} = z_{1,3}^*$, $p_{1,1} = 0.7\,exp(j\frac{3\pi}{9})$, $p_{2,1} = p_{1,1}^*$, $\sigma_{u,1}^2 = 1$. Due to their zeros $z_{1,1}, z_{2,1}$ and $z_{1,2}, z_{2,2}$, both processes are clearly non-minimum phase. The periodograms and PSDs are given in Fig. 3. Unlike example 1, both exhibit resonances at low frequency.

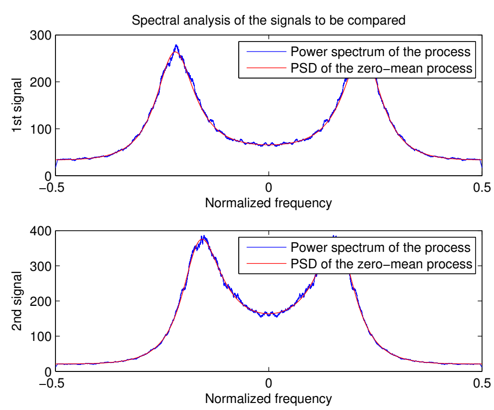

Figure 3: Periodogram of one realization of each ARMA process and the corresponding PSDs, Example 2

The same type of analysis as the one developed in Example 1 has been carried out. The results are summarized in Fig. 4 and Tables 4 and 5.

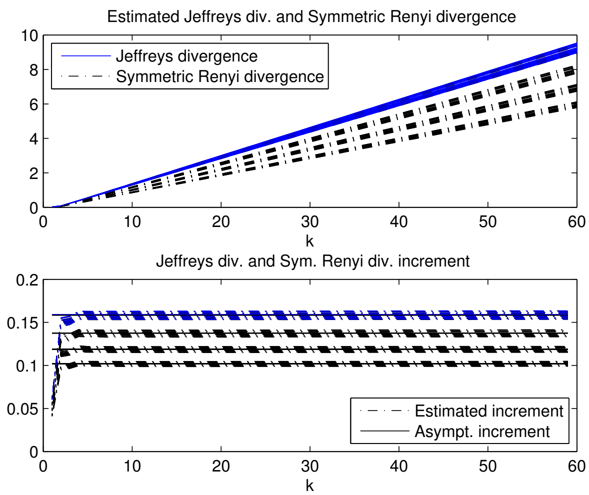

Figure 4: Estimated Jeffreys divergence (blue) and Symmetric Rényi divergence (black) and the resulting increments for Example 2 (20 realizations). For the symmetric Rényi divergence, α is equal to 0.7, 0.8, 0.9 and 0.999

Type of divergence	True div. rate	Last estim. increment Mean	Last estim. increment Variance	Normalized error Mean	Normalized error Variance
$KL^{(1,2)}$	0.1411	0.1416	1.13E-05	1.96E-02	1.71E-04
$RD^{(1,2)}$ $(\alpha = 0.999)$	0.1410	0.1414	1.12E-05	1.96E-02	1.71E-04
$RD^{(1,2)}$ $(\alpha = 0.9)$	0.1258	0.1261	7.79E-06	1.86E-02	1.35E-04
$RD^{(1,2)}$ $(\alpha = 0.8)$	0.1115	0.1117	5.44E-04	1.77E-02	1.10E-04
$RD^{(1,2)}$ $(\alpha = 0.7)$	0.0979	0.0980	3.80E-06	1.70E-02	9.37E-05
$KL^{(2,1)}$	0.1763	0.1760	1.20E-05	1.63E-02	1.07E-04
$RD^{(2,1)}$ $(\alpha = 0.999)$	0.1760	0.1757	1.19E-05	1.63E-02	1.07E-04
$RD^{(2,1)}$ $(\alpha = 0.9)$	0.1491	0.1489	7.80E-06	1.57E-02	9.33E-05
$RD^{(2,1)}$ $(\alpha = 0.8)$	0.1262	0.1260	5.30E-06	1.53E-02	8.76E-05
$RD^{(2,1)}$ $(\alpha = 0.7)$	0.1061	0.1060	3.69E-06	1.53E-02	8.12E-05

Table 4: Statistics on the estimations of divergence rates for different non-symmetric divergences, based on 20 realizations of the processes of example 2

As the processes have a more similar PSD than those in Example 1, the values of the JD and the SRD reach smaller values. As expected, as α approaches 1, the

Type of divergence	True	Last estim. increment		Normalized error	
	div. rate	Mean	Variance	Mean	Variance
JD	0.15868	0.15877	1.01E-05	1.70E-02	9.93E-05
SRD ($\alpha = 0.999$)	0.15845	0.15854	1.01E-05	1.70E-02	9.92E-05
SRD ($\alpha = 0.9$)	0.13746	0.13751	7.14E-06	1.65E-02	9.13E-05
SRD ($\alpha = 0.8$)	0.11885	0.11887	5.11E-06	1.62E-02	8.58E-05
SRD ($\alpha = 0.7$)	0.10204	0.10205	3.66E-06	1.60E-02	8.24E-05

Table 5: Statistics on the estimations of divergence rates for different symmetric divergences, based on 20 realizations of the processes of example 2

SRD approaches the JD. Once again, the normalized error mean and variance tends to be smaller when the value of α decreases.
Remark: these results are consistent with other dissimilarity measures. Indeed, for the processes of Example 1, the LSD [27] is equal to 13.14 whereas it is equal to is 3.3375 for the processes of Example 2.

7.1.3. Example 3: comparing two ARMA(2,2) processes whose poles and zeros are randomly drawn

To show that the obtained results above which confirm the theoretical analysis do not depend on the selected ARMA models, we compare ARMA processes whose poles and zeros are randomly drawn. In particular, modules and arguments of the poles and zeros are randomly selected in the range $[0, 1)$ and $[0, \pi]$, respectively, to generate a pair of complex conjugate poles and a pair of complex conjugate zeros. Once again, the normalized error between the last divergence increment and the divergence rate is computed for the different divergences under study. The results are given in Tables 6 and 7 for 20 realizations of these randomly generated processes. Once again, the simulations confirm the theory.

Type of divergence	Normalized error	
	Mean	Variance
$KL^{(1,2)}$	2.23E-02	7.61E-04
$RD^{(1,2)}$ ($\alpha = 0.999$)	1.83E-02	5.56E-04
$RD^{(1,2)}$ ($\alpha = 0.9$)	9.54E-03	1.45E-04
$RD^{(1,2)}$ ($\alpha = 0.8$)	8.09E-03	9.54E-05
$RD^{(1,2)}$ ($\alpha = 0.7$)	7.62E-03	6.95E-05
$KL^{(2,1)}$	4.38E-02	1.08E-02
$RD^{(2,1)}$ ($\alpha = 0.999$)	1.69E-02	2.91E-04
$RD^{(2,1)}$ ($\alpha = 0.9$)	9.64E-03	5.13E-05
$RD^{(2,1)}$ ($\alpha = 0.8$)	8.67E-03	3.85E-05
$RD^{(2,1)}$ ($\alpha = 0.7$)	8.12E-03	3.54E-05

Table 6: Normalized errors on the divergence rates for non symmetric divergences when comparing two random ARMA(2,2) processes (20 realizations)

Type of divergence	Normalized error	
	Mean	Variance
JD	4.54E-02	9.78E-03
SRD ($\alpha = 0.999$)	1.64E-02	3.21E-04
SRD ($\alpha = 0.9$)	8.86E-03	6.91E-05
SRD ($\alpha = 0.8$)	8.14E-03	5.35E-05
SRD ($\alpha = 0.7$)	7.79E-03	4.72E-05

Table 7: Normalized errors on the divergence rates for symmetric divergences when comparing two random ARMA(2,2) processes (20 realizations)

7.2. How do the divergence rate vary when some poles or zeroes vary?

In this subsection, our goal is to illustrate how the divergence rates may evolve when the poles or the zeros vary. To this end, the ARMA(2,2) process described by the following poles and zeroes:: $z_{1,1} = 0.8\, exp\,(j\frac{2\pi}{3})$, $z_{2,1} = z_{1,1}^*$, $p_{1,1} = 0.9\, exp(j\frac{\pi}{3})$, $p_{2,1} = p_{1,1}^*$, is taken as a reference model. Then, two different experimental setups are considered: i) the reference model is compared with ARMA(2,2) processes having the same zeros but different complex conjugate poles; ii) the reference model is compared with ARMA(2,2) processes having the same poles but different complex conjugate zeros. The obtained divergence rates are presented as a function of the modulus and the argument of one of the varying poles for case i) and of one of the varying zeros for case ii).

7.2.1. SRD rate when the poles vary

In Fig. 5, we present the SRD divergence rate for $\alpha = 0.95$ and 0.999 as a function of the modulus and the argument of the second-process pole.

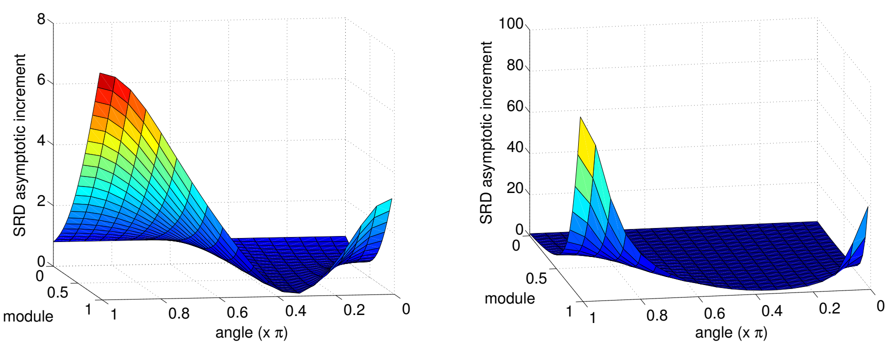

Figure 5: SRD divergence rate in function of the modulus and argument of the second-process pole with $\alpha = 0.95$ (left) and $\alpha = 0.999$ (right)

The minimum value occurs when the poles of the second process coincide with the poles of the first process (for instance, for modulus 0.9 and argument $\pi/3$). When the modulus of the pole is close to 0, the PSD of the second process -and consequently its correlation function- does not change much whatever the argument. This explains why the SRD rate is the same whatever the argument. The divergence rate increases much when the modulus is far from the origin and the argument is far from $\pi/3$. In these conditions the PSDs of the two ARMA processes PSD are very different.

For the SRD, the increment behavior depends on the value of α. For instance, when $\alpha = 0.95$, the SRD rate presents a rather linear behavior, while for $\alpha = 0.999$, the rate increases exponentially in function of the distance of the poles with respect to the first process. This shows that the value of α could be selected according to the sensitivity required by the specific application.

For comparison purposes, Fig. 6 shows the LSD. Even though the minimum value still occurs for modulus and argument close to 0.9 and $\pi/3$, the LSD presents a different behaviour than the SRD rate. For the LSD, the value increases linearly in the region near the minimum, and increases slowly when the poles are more separated.

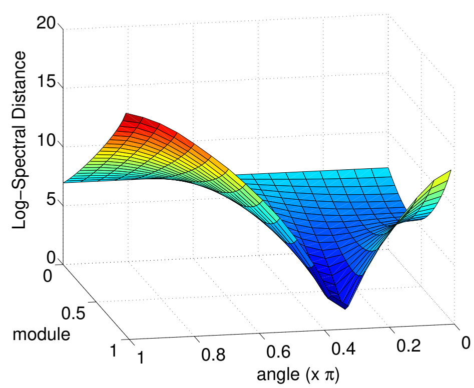

Figure 6: LSD as a function of the modulus and argument of the second-process pole

7.2.2. SRD rate when the zeros vary

Fig. 7 shows the SRD divergence rate for $\alpha = 0.95$ and $\alpha = 0.999$, as a function of the modulus and the argument of the second-process zero. The minimum value occurs when the zeros of the second process coincides with the poles of the first process (for instance, for modulus 0.8 and argument $2\pi/3$). The same type of comments as those made in the previous subsection can be given. When the modulus of the zero is equal to 0, the PSD of the second process is the same whatever the argument, leading to the same divergence rate. Then, the divergence rate increases much when the modulus is far from the origin and the argument is far from $2\pi/3$ (in these conditions the PSDs of the two ARMA processes PSD are very different). Fig. 8 presents the LSD, whose minimum value still occurs with modulus 0.8 and argument $2\pi/3$. We observe that the evolutions of the SRD rate and the LSD in the regions near and far from the minimum exhibit a behavior similar to those described in the varying poles example.

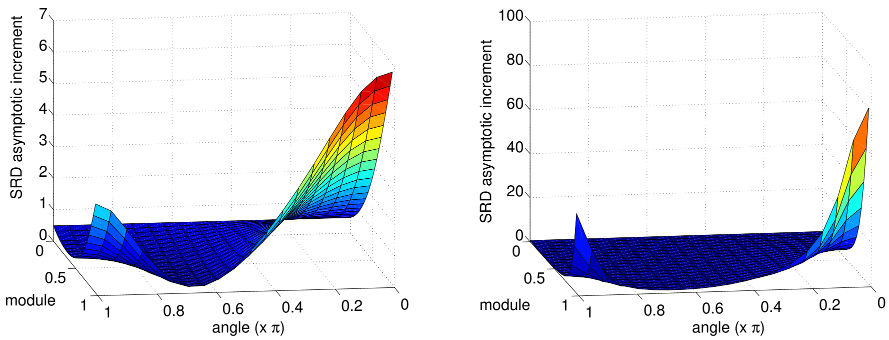

Figure 7: SRD divergence rate in function of the modulus and argument of the second-process zero with $\alpha = 0.95$ (left) and $\alpha = 0.999$ (right)

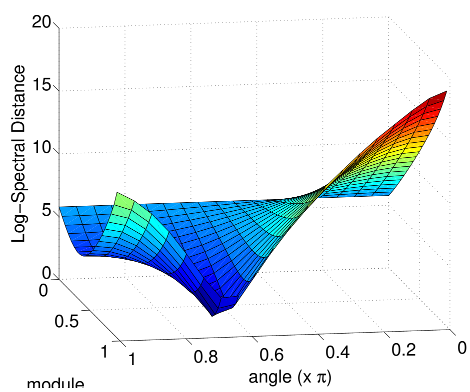

Figure 8: LSD as a function of the modulus and argument of the second-process zero

7.3. Divergence-rate based change detection when dealing with a non-stationary ARMA process

In this part, a time-varying ARMA process (TV-ARMA) is considered. In terms of modeling, it means that its poles and zeros vary over time. One example is presented in this paper for illustration. Its time representation and spectrogram are given in Figure 9. The first 2000 samples of the TV-ARMA process are considered as the frame of reference. Then, a sliding window is used. Detecting and evaluating a statistical change in the ARMA process consists in estimating the divergence rate. To this end, for the frame of reference and the sliding one, the model parameters are estimated by using the PEM method [32]. Given the estimates of the model parameters of the two frames related to $H_1(z)$ and $H_2(z)$ and the variances of the driving processes, the model parameters related to the transfer function $H_\alpha(z)$ and $H_{1-\alpha}(z)$ as well as the associated driving-process variance can be deduced by using the method presented in Appendix B.

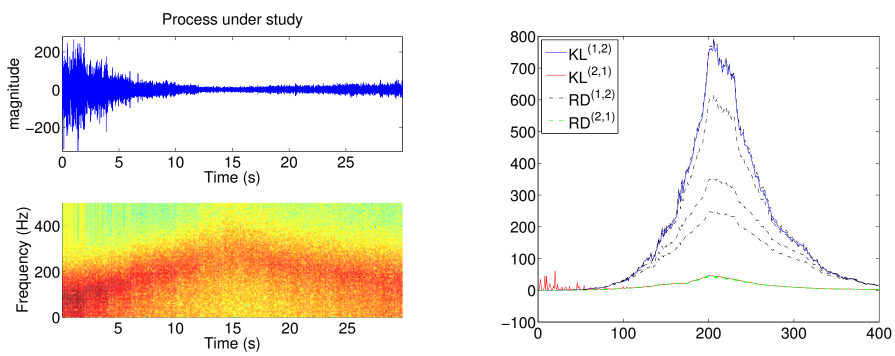

Figure 9: Time-domain representation and spectrogram of the time-varying ARMA process (left) and evolution of the KL (blue and red) and Rényi divergences increment (black and green) for a TV-ARMA process (α equal to 0.999, 0.9995, 0.9999 and 0.99999) (right)

Some results are presented in Fig. 9 when α is set to the values 0.999, 0.9995, 0.9999 and 0.99999. One can notice that the ranges of the divergence rates significantly decreases when α decreases. The divergence rate is sensitive to the selection of α, especially when α is close to 1.

7.4. Comparing the divergence rates to characterize experiment-induced stress

Various studies based on RR intervals, which correspond to the time between two consecutive R-waves of the QRS signal on electrocardiograms, have been recently conducted in order to analyze the interconnections between cardiac regulations and the central nervous system. The reader may refer to [11] and [49] for instance. As an illustration of our work, we study if the divergence rates can be relevant to detect when people who enforce cognitive tasks are under stress. To this end, this last subsection is organized as follows. Some information about the experimental protocol are first given. Then, data analysis is provided.

7.4.1. Population and experimental protocol

With some psychologists and physiologists[10], one of the authors developed a psychological protocol the purpose of which was to induce levels of stress during a cognitive task. 33 healthy volunteers (age: 35.6 ± 13.9 years, 19 women) recruited among students and employees followed this protocol lasting more than 1 hour. They gave their written informed consent to participate in the study. In addition, they all filled out a series of questionnaires such as the Spielberger state anxiety questionnaire[11] (STAI) [44] and the NASA-TLX[12] [18] before and after each test session. Both questionnaires made it possible to measure the impact of stress on anxiety and workload after each situation.
In the current experiment, three consecutive situations are considered:

- 1^{st} **period called reference period and denoted as** Ref**:** the subjects were first seated in front of a computer in a room at 20°C between 10 am and noon in order to limit the effects of chronobiology. They watched an

[10]The authors would like to thank Prof. V. Deschodt-Arsac, Prof. L. Arsac, Dr. V. Lespinet-Najib and Dr. E. Blons. One of the authors has different publications with them for instance in [4].

[11]It consists of 20 questions that evaluate the current state of anxiety by using items that measure subjective feelings of apprehension, tension, nervousness and worry.

[12]A self-assessed measure of workload based on six components: mental demand, physical demand, temporal demand, performance, effort, and frustration level.

emotionally-neutral documentary film.

- 2^{nd} **period called situation of cognitive tasks and denoted as T_c:** Still seated, the subjects had to answer questions of logic, memorization and mental calculation, proposed on the screen of the computer.

- 3^{nd} **period called situation of cognitive tasks and stress, denoted as $T_c + S$:** Still seated, the subjects had to answer questions proposed on the screen of the computer, but different types of disturbances could be considered: people behaving as an attentive and evaluative audience were near the subject, auditory and visual distractions were generated.

During the experiments, the RR intervals were recorded with the Polar H10 belt product connected to an Ipod using Bluetooth. An application was installed on the ipod, making it possible to store the RR intervals. When dealing with 8-minute periods, this corresponds to approximately 500 successive RR intervals, depending on the average individual heart rate.

7.4.2. Data analysis

Usually, various criteria are considered by physiologists to analyze the RR intervals. This can be the power in the frequency bands 0.04-0.15 Hz and 0.15-0.5 Hz respectively in order to evaluate the contributions of the sympathetic and parasympathetic systems of the autonomic nervous system. They are often called low-frequency power (LF) and high frequency power (HF). One can also look at the ratio of powers in low and high frequency (LF/HF). The regularity of the process can be also considered. The reader may refer to [4].

Let us study how the divergence rate varies from the 2^{nd} and the 3^{rd} period with respect to the reference period. In other words, let us compare the distributions, which are *a priori* Gaussian, of the RR processes in the i^{th} period (with $i = 2, 3$) with the first one. It should be noted that we checked if the time-series were globally wide-sense stationary on the periods of analysis. The divergence rates are estimated using the data obtained in each period: in each case, the covariance matrices and the means are estimated by using a maximum-likelihood estimator for different sizes k in an interval k_{min} and k_{max} defined by the practitioner. Then, the increments are computed. Once these differences are smaller than a pre-defined threshold, the divergence increments are averaged to get an estimation of the divergence rate. In the table below, the rates of the Kullback-Leibler divergence, the Jeffreys divergence, the Rényi divergence and its symmetric version are presented.

Estimates of the divergence rates	T_c vs. Ref	T_c+S vs. Ref
$\Delta KL^{(1,2)}$	0.230	0.263
$\Delta RD^{(1,2)}(0,99)$	0.229	0.259
$\Delta RD^{(1,2)}(0,95)$	0.219	0.245
$\Delta RD^{(1,2)}(0,9)$	0.208	0.231
$\Delta RD^{(1,2)}(0,8)$	0.188	0.205
$\Delta RD^{(1,2)}(0,7)$	0.169	0.184
$\Delta JD^{(1,2)}$	0.366	0.421
$\Delta SRD^{(1,2)}(0,99)$	0.351	0.405
$\Delta SRD^{(1,2)}(0,95)$	0.312	0.356
$\Delta SRD^{(1,2)}(0,9)$	0.277	0.314
$\Delta SRD^{(1,2)}(0,8)$	0.228	0.254
$\Delta SRD^{(1,2)}(0,7)$	0.190	0.210

Table 8: Estimates of divergence rates for the different period

Given Table 8, one can first notice that the larger α, the larger the divergence rate whatever the divergence and the period (T_c or $T_c + S$). When α tends to 1, the estimation of the rate of the Rényi divergence and the symmetric Rényi divergence tends to the estimations of the Kullback-Leibler and Jeffreys divergence rates respectively. When computing Anova test, one noticed that the larger α, the smaller the p-value is, guaranteeing a better distinction between the cognitive-task period and the stress one. In conjunction with other features as the ones cited above and whose results are presented in [4], an automatic detection of the stress could probably developed in the future.

8. Conclusions and perspectives

This paper aims at analyzing the divergence rate of different divergences (Kullback-Leibler, Jeffreys and Rényi) when considering Gaussian ARMA processes. Our work, based on the notions such as inverse filtering and the Yule-Walker equation, is complementary to Gil's work dealing with Gaussian processes and taking advantages of results related to the asymptotic distribution of the eigenvalues of Toeplitz form. We show that the expressions we obtained are consistent with the ones he obtained. Illustrations are provided and confirm the theoretical analysis. It highlights that the ranges of values of the divergence rates significantly increases when α increases, especially when the latter is close to 1.

There are numerous perspectives: we would like to analyze how the ARMA parameters influence how quickly the divergence increment converges to the divergence rate. In addition, the analysis for non Gaussian ARMA processes and multivariate ARMA processes could be other topics of interest. Finally, we plan to study other divergence increments for Gaussian ARMA processes such as the one of the Sharma-Mittal divergence.

9. Acknowledgments

The Sistema Nacional de Investigación (SNI), SENACYT-Panama supports research activities by Fernando Merchan.

AppendixA. Derivation of the RD in the Gaussian case

By combining (2) and (6), the RD can be written as follows:

$$RD_k^{(1,2)}(\alpha) = \frac{1}{\alpha-1}\ln\int \frac{1}{(\sqrt{2\pi})^k|Q_{k,1}|^{\alpha/2}|Q_{k,2}|^{(1-\alpha)/2}} exp\big(-\frac{1}{2}A\big) \tag{A.1}$$

where, after some mathematical developments, A can be expressed as:

$$\begin{aligned} A &= \alpha[x_k-\mu_{k,1}]^T Q_{k,1}^{-1}[x_k-\mu_{k,1}] + (1-\alpha)[x_k-\mu_{k,2}]^T Q_{k,2}^{-1}[x_k-\mu_{k,2}] \qquad \text{(A.2)} \\ &= x_k^T\big(\alpha Q_{k,1}^{-1}+(1-\alpha)Q_{k,2}^{-1}\big)x_k - 2x_k^T\big(\alpha Q_{k,1}^{-1}\mu_{k,1}+(1-\alpha)Q_{k,2}^{-1}\mu_{k,2}\big) \\ &+ \alpha\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + (1-\alpha)\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} \end{aligned}$$

Let us now introduce the following two quantities:

$$Q_{k,equ}^{-1} = \alpha Q_{k,1}^{-1} + (1-\alpha) Q_{k,2}^{-1} \tag{A.3}$$

and

$$Q_{k,equ}^{-1}\mu_{k,equ} = \alpha Q_{k,1}^{-1}\mu_{k,1} + (1-\alpha) Q_{k,2}^{-1}\mu_{k,2} \tag{A.4}$$

Using (A.3) and (A.4), it can be shown that (A.2) becomes:

$$\begin{aligned} A &= (X_k - \mu_{k,equ})^T Q_{k,equ}^{-1}(X_k - \mu_{k,equ}) \\ &- \mu_{k,equ}^T Q_{k,equ}^{-1}\mu_{k,equ} + \alpha\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + (1-\alpha)\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} \end{aligned} \tag{A.5}$$

Let us now rewrite the expression (A.1) of the Rényi divergence using (A.5):

$$\begin{aligned} RD_{12,\alpha}(k) &= \frac{1}{\alpha-1}\ln\Big(\frac{|Q_{k,equ}|^{1/2}}{|Q_{k,1}|^{\alpha/2}|Q_{k,2}|^{(1-\alpha)/2}}\times \\ &\exp\big(-\frac{1}{2}(\alpha\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + (1-\alpha)\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} - \mu_{k,equ}^T Q_{k,equ}^{-1}\mu_{k,equ})\big)\Big) \end{aligned} \tag{A.6}$$

because $\int \frac{1}{(\sqrt{2\pi})^k|Q_{k,equ}|^{1/2}} \exp\big(-\frac{1}{2}(X_k - \mu_{k,equ})^T Q_{k,equ}^{-1}(X_k - \mu_{k,equ})\big)\mathrm{d}X_k = 1$. At this stage, let us express the logarithm of the first part of the equation (A.6), *i.e.* $\frac{1}{2(\alpha-1)}\ln\Big(\frac{|Q_{k,equ}|}{|Q_{k,1}|^{\alpha}|Q_{k,2}|^{(1-\alpha)}}\Big)$. By introducing:

$$Q_{k,\alpha} = \alpha Q_{k,2} + (1-\alpha)Q_{k,1} \tag{A.7}$$

one has:

$$\begin{aligned} Q_{k,equ}^{-1} &\underset{(A.3)}{=} \alpha Q_{k,1}^{-1} + (1-\alpha)Q_{k,2}^{-1} = Q_{k,1}^{-1}\big(\alpha I + (1-\alpha)Q_{k,1}Q_{k,2}^{-1}\big) \\ &= Q_{k,1}^{-1}(\alpha Q_{k,2} + (1-\alpha)Q_{k,1})Q_{k,2}^{-1} = Q_{k,1}^{-1}Q_{k,\alpha}Q_{k,2}^{-1} \end{aligned} \tag{A.8}$$

Their determinants hence satisfy:

$$|Q_{k,equ}| = \frac{|Q_{k,1}||Q_{k,2}|}{|Q_{k,\alpha}|} \tag{A.9}$$

Therefore, one has:

$$\frac{1}{2(\alpha-1)}\ln\Big(\frac{|Q_{k,equ}|}{|Q_{k,1}|^{\alpha}|Q_{k,2}|^{(1-\alpha)}|}\Big) = -\frac{1}{2(\alpha-1)}\ln\Big(\frac{|Q_{k,\alpha}|}{|Q_{k,1}|^{1-\alpha}|Q_{k,2}|^{\alpha}}\Big) \tag{A.10}$$

Now, let us express the logarithm of the second part in (A.6) given by:

$$B = -\frac{1}{2(\alpha-1)}\Big(\alpha\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + (1-\alpha)\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} - \mu_{k,equ}^T Q_{k,equ}^{-1}\mu_{k,equ}\Big) \tag{A.11}$$

To this end, let us rewrite $\mu_{k,equ}^T Q_{k,equ}^{-1} \mu_{k,equ}$ by using (A.3) and by considering that that $(Q_{k,equ}^{-1})^T = Q_{k,equ}^{-1}$:

$$\begin{aligned}
\mu_{k,equ}^T Q_{k,equ}^{-1} \mu_{k,equ} &= \mu_{k,equ}^T Q_{k,equ}^{-1} Q_{k,equ} Q_{k,equ}^{-1} \mu_{k,equ} \\
&\underset{(A.3)}{=} \Big(\alpha \mu_{k,1}^T Q_{k,1}^{-1} + (1-\alpha)\mu_{k,2}^T Q_{k,2}^{-1}\Big) Q_{k,equ} \Big(\alpha Q_{k,1}^{-1}\mu_{k,1} + (1-\alpha) Q_{k,2}^{-1}\mu_{k,2}\Big) \\
&= \alpha^2 \mu_{k,1}^T Q_{k,1}^{-1} Q_{k,equ} Q_{k,1}^{-1}\mu_{k,1} + (1-\alpha)^2 \mu_{k,2}^T Q_{k,2}^{-1} Q_{k,equ} Q_{k,2}^{-1}\mu_{k,2} \\
&+ 2\alpha(1-\alpha)\mu_{k,1}^T Q_{k,1}^{-1} Q_{k,equ} Q_{k,2}^{-1}\mu_{k,2}
\end{aligned} \tag{A.12}$$

By combining (A.11) and (A.12) and rearranging the terms, one has:

$$\begin{aligned}
B = -\frac{1}{2(\alpha-1)}\Big(&\alpha\mu_{k,1}^T (Q_{k,1}^{-1} - \alpha Q_{k,1}^{-1} Q_{k,equ} Q_{k,1}^{-1})\mu_{k,1} + \\
&(1-\alpha)\mu_{k,2}^T (Q_{k,2}^{-1} - (1-\alpha) Q_{k,2}^{-1} Q_{k,equ} Q_{k,2}^{-1})\mu_{k,2}\Big) - \alpha\mu_{k,1}^T Q_{k,1}^{-1} Q_{k,equ} Q_{k,2}^{-1}\mu_{k,2}
\end{aligned} \tag{A.13}$$

Given (A.3), $Q_{k,\alpha}^{-1}$ can be expressed using the matrix inversion lemma. Two expressions can be considered:

$$Q_{k,\alpha}^{-1} = \frac{1}{\alpha} Q_{k,2}^{-1} - \frac{1}{\alpha} Q_{k,2}^{-1} (\frac{1}{1-\alpha} Q_{k,1}^{-1} + \frac{1}{\alpha} Q_{k,2}^{-1})^{-1} \frac{1}{\alpha} Q_{k,2}^{-1} \tag{A.14}$$

which can rewritten after some simplifications as follows:

$$\alpha Q_{k,\alpha}^{-1} = Q_{k,2}^{-1} + (\alpha-1) Q_{k,2}^{-1} Q_{k,equ} Q_{k,2}^{-1} \tag{A.15}$$

Similarly, one has:

$$Q_{k,\alpha}^{-1} = \frac{1}{1-\alpha} Q_{k,1}^{-1} - \frac{1}{1-\alpha} Q_{k,1}^{-1} (\frac{1}{\alpha} Q_{k,2}^{-1} + \frac{1}{1-\alpha} Q_{k,1}^{-1})^{-1} \frac{1}{1-\alpha} Q_{k,1}^{-1} \tag{A.16}$$

which can rewritten as follows:

$$(1-\alpha) Q_{k,\alpha}^{-1} = Q_{k,1}^{-1} - \alpha Q_{k,1}^{-1} Q_{k,equ} Q_{k,1}^{-1} \tag{A.17}$$

Let us now combine (A.13), (A.15) and (A.17):

$$\begin{aligned}
B &= \frac{\alpha}{2}\Big(\mu_{k,1}^T Q_{k,\alpha}^{-1}\mu_{k,1} + \mu_{k,2}^T Q_{k,\alpha}^{-1}\mu_{k,2}\Big) - \alpha\mu_{k,1}^T Q_{k,\alpha}^{-1}\mu_{k,2} \\
&= \frac{\alpha}{2}(\mu_{k,1} - \mu_{k,2})^T Q_{k,\alpha}^{-1}(\mu_{k,1} - \mu_{k,2})
\end{aligned} \tag{A.18}$$

Finally, by using (A.10) and (A.18), the expression of the RD is given by:

$$RD_k^{(1,2)}(\alpha) = -\frac{1}{2(\alpha-1)}\ln\Big(\frac{|Q_{k,\alpha}|}{|Q_{k,1}|^{1-\alpha}|Q_{k,2}|^{\alpha}}\Big) + \frac{\alpha}{2}\mathrm{Tr}\Big(Q_{k,\alpha}^{-1}\Delta\mu_k\Delta\mu_k^T\Big) \tag{A.19}$$

AppendixB. Appendix B

Let us assume that the ARMA parameters of order (p_i, q_i) with $i = 1, 2$ are known. As mentioned in the description of the ARMA processes, the parameter b_0 for any ARMA process is equal to 1. The purpose of this appendix is to explain how to retrieve the ARMA parameters related to H_α. The same reasoning can be done for the ARMA parameters related to $H_{1-\alpha}$. The complex spectrum associated with H_α can be expressed in two different ways. On the one hand, one has by definition:

$$S_\alpha(\theta) = \sigma_{u,\alpha}^2 \, |H_\alpha(e^{j\theta})|^2 = \sigma_{u,\alpha}^2 \, \frac{|B_\alpha(e^{j\theta})|^2}{|A_\alpha(e^{j\theta})|^2} \tag{B.1}$$

On the other hand, given the way the process is built, one has:

$$\begin{aligned} S_\alpha(\theta) &= \alpha\sigma_{u,2}^2 \, |H_2(e^{j\theta})|^2 + (1-\alpha)\sigma_{u,1}^2 \, |H_1(e^{j\theta})|^2 \\ &= \frac{\alpha\sigma_{u,2}^2 |B_2(e^{j\theta})|^2 |A_1(e^{j\theta})|^2 + (1-\alpha)\sigma_{u,1}^2 \, |B_1(e^{j\theta})|^2 |A_2(e^{j\theta})|^2}{|A_2(e^{j\theta})|^2 |A_1(e^{j\theta})|^2} \end{aligned} \tag{B.2}$$

Therefore, the poles of the H_α are the poles of H_1 and H_2. Let us now search for the zeros of H_α and the variance $\sigma_{u,\alpha}^2$. Given the above two expressions of $S_\alpha(\theta)$, one has:

$$\sigma_{u,\alpha}^2 |B_\alpha(e^{j\theta})|^2 = \alpha\sigma_{u,2}^2 |B_2(e^{j\theta})|^2 |A_1(e^{j\theta})|^2 + (1-\alpha)\sigma_{u,1}^2 \, |B_1(e^{j\theta})|^2 |A_2(e^{j\theta})|^2 \tag{B.3}$$

By replacing $e^{j\theta}$ by z, this leads to:

$$\begin{aligned} \sigma_{u,\alpha}^2 B_\alpha(z) B_\alpha(z^{-1}) = \alpha\sigma_{u,2}^2 B_2(z) B_2(z^{-1}) A_1(z) A_1(z^{-1}) \\ + (1-\alpha)\sigma_{u,1}^2 B_1(z) B_1(z^{-1}) A_2(z) A_2(z^{-1}) \end{aligned} \tag{B.4}$$

When the AR and MA parameters of the two processes under study are available, the polynomials $\{A_i(z)A_i(z^{-1})\}_{i=1,2}$ and $\{B_i(z)B_i(z^{-1})\}_{i=1,2}$ can be easily derived as well as the right hand side of (B.4). They correspond to a weighted sum of z^l with $l = -p_i, ..., 0, ..., p_i$ for $\{A_i(z)A_i(z^{-1})\}_{i=1,2}$ and with $l = -q_i, ..., 0, ..., q_i$ for $\{B_i(z)B_i(z^{-1})\}_{i=1,2}$. $B_\alpha(z)B_\alpha(z^{-1})$ correspond to a weighted sum of z^l where l varies between $-max(p_1 + q_2, p_2 + q_1), ..., max(p_1 + q_2, p_2 + q_1)$. It should be noted that if z is a root of the right hand side of (B.4), z^*, $\frac{1}{z}$ and $\frac{1}{z^*}$ are also roots. As the roots of the right hand side of (B.4) are the roots of $B_\alpha(z)B_\alpha(z^{-1})$, the zeros of H_α can be easily defined[13]. Consequently the coefficients of the polynomial of $B_\alpha(z)$ can be deduced. Finally, the variance $\sigma^2_{u,\alpha}$ can be obtained by identification by comparing the weights of z^0 in both sides of (B.4).

[1] M. Basseville. Detecting changes in signals and systems – A survey. Automatica, 24 (3):309–326, 1988.

[2] M. Basseville. Divergence measures for statistical data processing. an annotated bibliography. Signal Processing, 93 (4):621–633, 2013.

[3] A. Basu, I. R. Harris, N. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. Biometrika, pages 549–559, 1998.

[4] E. Blons, L. M. Arsac, P. Gilfriche, H. McLeod, V. Lespinet-Najib, E. Grivel, and V. Deschodt-Arsac. Alterations in heart-brain interactions under mild stress during a cognitive task are reflected in entropy of heart rate dynamics. Scientific reports, 9:18190, 2019.

[5] 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, 55:5564–5578, December 2007.

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

[7] G. Box, G. Jenkins, G. Reinsel, and G. Ljung. Time series analysis, forecasting and control. University of Michigan Press, 2015.

[13]For instance, one can select the zeros whose modulus is smaller or equal to 1.

[8] J. E. Contreras-Reyes. Analyzing fish condition factor index through skew-gaussian information theory quantifiers. Fluctuation and Noise Letters, 15 (2):1–16, 2016.

[9] J. E. Contreras-Reyes and R. B. Arellano-Valle. Kullback—Leibler divergence measure for multivariate skew-normal distributions. Entropy, 14 (9):1606–1626, 2012.

[10] D. C. de Souza, R. F. Vigelis, and C. C. Cavalcante. Geometry induced by a generalization of rényi divergence. Entropy, 18, 407, 2015.

[11] D. A. Dimitriev, E. V. Saperova, and A. D. Dimitriev. State anxiety and nonlinear dynamics of heart rate variability in students. Plos one, 2016.

[12] R. Diversi, E. Grivel, and F. Merchan. ARMA model identification from noisy observations based on a two-step errors-in-variables approach. IFAC PapersOnLine, 50:14143–14149, 2017.

[13] T. T. Georgiou and A. Lindquist. A convex optimization approach to ARMA modeling. IEEE Transactions on Automatic Control, 53:1108–1119, 2008.

[14] M. Gil. On Rényi divergence measures for continuous alphabet sources. PhD Thesis, Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada, 2011.

[15] V. Girardin and P. Lhote, L. Regnault. Different closed-form expressions for generalized entropy rates of markov chains. Methodology and Computing in Applied Probability, 21:1431–1452, 2019.

[16] C. W. J. Granger and M. J. Morris. Time series modellig and interpretation. Journal of the Royal Statistical Society A, 139, Part 2:246–257, 1976.

[17] R. M. Gray. Toeplitz and circulant matrices: A review. Foundations and Trends in Communications and Information Theory, 2:155–239, 2006.

[18] S. G. HartLowell and E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. Advances in Psychology, 52:139–183, 1988.

[19] S. Haykin and A. Steinhardt. Adaptive Radar Detection and Estimation. Wiley–Interscience, 1992.

[20] S. Ihara. Information Theory for Continuous Systems. World Scientific, 1993.

[21] R. Isermann. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer, 2006.

[22] H. Jeffreys. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 186:453–461, September 1946.

[23] S. M. Kay. Modern Spectral Estimation. Prentice-hall, 1988.

[24] P. Kluza. On Jensen-Rényi and Jeffreys-Rényi type f-divergences induced by convex functions. Physica A: Statistical Mechanics and its Applications, 548:122527, 2020.

[25] P. Kluza and M. Niezgoda. Generalizations of Crooks and Lin's results on Jeffreys-Csiszár and Jensen-Csiszár f-divergences. Physica A: Statistical Mechanics and its Applications, 463:383–393, 2016.

[26] S. Kullback and R. A. Leibler. On Information and Sufficiency. The Annals of Mathematical Statistics, 22 (1):79–86, 1951.

[27] B.-H. Juang L. R. Rabiner. Fundamentals of speech recognition. PTR Prentice Hall, 1993.

[28] M. Lasta and R. Shumway. Detecting abrupt changes in a piecewise locally stationary time series. Journal of Multivariate Analysis, 99:191–214, 2008.

[29] A. Lefevre, F. Bach, and C. Fevotte. Online algorithms for nonnegative matrix factorization with the itakura-saito divergence. WASPAA 2011, 2011.

[30] 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. Signal Processing, 131:350–363, 2017.

[31] L. Legrand and E. Grivel. Jeffrey's divergence between autoregressive processes disturbed by additive white noises. Signal Processing, 149:162–178, 2018.

[32] L. Ljung. System identification: theory for the user. Prentice-hall, 1999.

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

[34] F. Merchan, E. Grivel, and R. Diversi. Rényi divergence to compare moving-average processes. 2018 IEEE Statistical Signal Processing Workshop (SSP 2018), pages 149—153, 2018.

[35] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis. Touchless monitoring of breathing function. IEEE EMBS, pages 1196–1199, 2004.

[36] M. Najim. Modeling, estimation and optimal filtering in signal processing. Wiley, 2010.

[37] 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, 56, n°11:5847–5861, 2010.

[38] F. Nielsen and R. Nock. A closed-form expression for the sharmamittal entropy of exponential families. Journal of Physics A: Mathematical and Theoretical, 45, 2012.

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

[40] P. Regnault, V. Girardin, and L. Lhote. Weighted closed form expressions based on escort distributions for Rényi entropy rates of markov chains. GSI'17, 2017.

[41] M. Saleh, E. Grivel, and S.-M. Omar. Jeffrey's divergence between ARFIMA processes. Digital Signal Processing, 82:175–186, 2018.

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

[43] T. Soderstrom and P. Stoica. System Identification. Prentice Hall, 1989.

[44] C. D. Spielberger, R. L. Gorsuch, R. Lushene, P. R. Vagg, and G. A. Jacobs. Manual for the state-trait anxiety inventory (STAI). Palo Alto, CA: Mind Garden, 1983.

[45] P. Stoica and R. Moses. Spectral analysis of signals. Prentice Hall, 2005.

[46] M. Sugiyama, S. Liu, M. Christoffel du Plessis, M. Yamanaka, T. Suzuki, and T. Kanamori. Direct divergence approximation between probability distributions and its applications in machine learning. Journal of Computing Science and Engineering, 7, no. 2:99–111, 2013.

[47] R. F. Vigelis, L. H.F. de Andrade, and C. C. Cavalcante. Conditions for the existence of a generalization of Rényi divergence. Physica A, 558:124953, 2020.

[48] B. Wei and J. Gibson. Comparison of distance measures in discrete spectral modeling. 9th Digital Signal Processing Workshop, 2000.

[49] H. Young and D. Benton. We should be using nonlinear indices when relating heart-rate dynamics to cognition and mood. Scientific Reports, 5, 2015.