# ARTICLE IN PRESS

[Digital Signal Processing](https://doi.org/10.1016/j.dsp.2018.01.006) ••• (••••) •••-•••



1 Contents lists available at [ScienceDirect](http://www.ScienceDirect.com/) <mark> Digita</mark>l <sup>67</sup>

 $\sum_{i=1}^{3}$  699  $\sum_{i=1}^{69}$  599  $\sum_{i=$ 4 70



[www.elsevier.com/locate/dsp](http://www.elsevier.com/locate/dsp)

### $11$  Interpretian the asymptotic increment of lefts  $\alpha$ 's divergence between  $17$  $\frac{11}{12}$  Interpreting the asymptotic increment of Jeffrey's divergence between  $\frac{77}{78}$  $13$  some random processes  $79$

### <sup>15</sup> Bric Grivel<sup>a</sup>, Mahdi Saleh<sup>a, b,∗</sup>, Samir-Mohamad Omar <sup>b</sup> and a strong  $16$  82

17 83 <sup>a</sup> *Bordeaux University, Bordeaux INP ENSEIRB-MATMECA, IMS, UMR CNRS 5218, Talence, France* 18 84 <sup>b</sup> *Department of Physics and Electronics, Faculty of Sciences I, Lebanese University Hadath, Beirut, Lebanon*

### 20 and the contract of the con 21 ANIILLE INFO ADSINALI A R T I C L E I N F O A B S T R A C T

22 88 *Article history:* Available online xxxx

*Keywords:* Jeffrey's divergence Kullback–Leibler divergence Model comparison ARMA process Asymptotic analysis Physical meaning

23 Article history:<br>23 Australia and image processing, Jeffrey's divergence (JD) is used in many applications for classification,  $24$   $24$  autoregressive (AR) and/or moving average (MA) processes state that the asymptotic JD increment, which  $\frac{1}{2}$ 25 91 is the difference between two JDs based on *k* and *(k* − 1*)*-dimensional random vectors when *k* becomes 26 1ettrey's avergence<br>
Entrepreneur and the second with the second high, tends to a constant value, except JDs which involve a 1st-order MA process whose power spectral 27 Model comparison exceptive that the density (PSD) is null for one frequency. In this paper, our contribution is threefold. We first propose an <sup>93</sup> <sup>28</sup> ARMA process and the asymptotic JD increment for ergodic WSS ARMA processes: it consists in calculating <sup>94</sup> 29 Asymptotic analysis **1953** the power of the first process filtered by the inverse filter associated with the second process and 95 30 Physical meaning the same that the conversely. This explains the atypical cases identified in previous works and generalizes them to any the sequention of the sequention of the sequention of the sequention of the sequen 31 97 ergodic WSS ARMA process of any order whose PSD is null for one or more frequencies. Then, we suggest <sub>32</sub> somparing other random processes such as noisy sums of complex exponentials (NSCE) by using the JD. 33 33 33 33 **1 1** 13 m this case, the asymptotic JD increment and the convergence speed towards the asymptotic JD are 34 100 useful to compare the processes. Finally, NSCE and *p*th-order AR processes are compared. The parameters 35 35 35 101 of the processes, especially the powers of the processes, have a strong influence on the asymptotic JD change detection, etc. The previous studies done on the JD between ergodic wide-sense stationary (WSS) increment.

36 102 © 2018 Elsevier Inc. All rights reserved.

## **1. Introduction**

56 122 different SSRs, can be considered. This leads to multiple-model ap-In the field of signal and image processing or even in the field of control, models or processes are often compared. This is, for instance, the case when dealing with identification issues where the

Corresponding author at: Bordeaux University, Bordeaux INP ENSEIRB-MATMECA, IMS, UMR CNRS 5218, Talence, France.

*E-mail addresses:* [eric.grivel@ims-bordeaux.fr](mailto:eric.grivel@ims-bordeaux.fr) (E. Grivel),

[mahdi.saleh@u-bordeaux.fr](mailto:mahdi.saleh@u-bordeaux.fr) (M. Saleh), [samir.omar@liu.edu.lb](mailto:samir.omar@liu.edu.lb) (S.-M. Omar).

64 130 <https://doi.org/10.1016/j.dsp.2018.01.006>

65 131 1051-2004/© 2018 Elsevier Inc. All rights reserved.66 and the contract of the con

<sup>41</sup> **1. Introduction 1. Introduction case, selecting dissimilar models is suggested by Bar-Shalom in [\[6\].](#page--1-0) <sup>107</sup>** 42 108 Therefore, a way to compare models *a priori* has to be designed. <sup>43</sup> In the field of signal and image processing or even in the field On the other hand, the practitioners may prefer models whose pa-<sup>44</sup> of control, models or processes are often compared. This is, for in-<br> **Example 12.** The estimation may be easier and which lead to SSRs that the 1<sup>10</sup> <sup>45</sup> stance, the case when dealing with identification issues where the can be written in a simple way. Thus, they can propose to use  $111$ <sup>46</sup> estimated model parameters are compared with the true ones in an autoregressive (AR) model instead of a moving average (MA)  $112$ <sup>47</sup> order to analyze the estimation accuracy [\[1\] \[2\].](#page--1-0) Model compar- model. Therefore, comparing AR and MA models can be useful es- $48$  ison also occurs when designing a Bayesian estimation approach pecially when the power spectral density (PSD) is not null at some  $114$ 49 115 based on Kalman filtering, *H*∞ filtering or particle filtering [\[3\].](#page--1-0) In <sup>50</sup> this case, *a* priori modeling the system under study is necessary can also be of interest, especially in the field of image process-<sup>51</sup> and leads to the state space representation (SSR) of the system. ing when textures are compared [7]. In biomedical applications or  $\frac{52}{100}$  However, several problems may arise. On the one hand, the per-<br> $\frac{118}{1000}$  forecast, change detection can be useful. In this latter case, <sup>53</sup> 53 formance of the estimation algorithm depends on how good the the problem is to detect whether the statistical properties of a pro- $^{54}$  SSR fits the system. As it is not necessarily easy to set it properly, cess change over time. In the above situations, statistical properties  $\frac{55}{2}$  estimation approaches combining different models, or equivalently  $\frac{253}{2}$  have to be analyzed an autoregressive (AR) model instead of a moving average (MA) model. Therefore, comparing AR and MA models can be useful especially when the power spectral density (PSD) is not null at some frequencies or does not exhibit resonances. Process comparison can also be of interest, especially in the field of image processing when textures are compared [\[7\].](#page--1-0) In biomedical applications or flood forecast, change detection can be useful. In this latter case, the problem is to detect whether the statistical properties of a process change over time. In the above situations, statistical properties have to be analyzed.

 $\frac{57}{12}$  proaches such as the interactive multiple models  $\frac{1}{2}$  [5]. In this trance measures which include the log-spectral distance (LSD) the  $_{58}$  proacues such as the interactive multiple models [4] [5]. In this tance measures which include the log-spectral distance (LSD), the  $_{124}$ 59 125 Itakura–Saito divergence (ISD), the Itakura divergence (ID), the 60 126 model distance measure proposed by Itakura and their symmetric 61 Controller and the South and Superintensity, boldeaux in Ensemble Versions as well as parametric spectral distances such as the cep-<br>61 MATMECA IMS UMR CNRS 5218 Talence France 62  $_1$  E-mail addresses: eric grivel@ims-bordeaux.fr (E. Grivel).  $\blacksquare$  Stral distance  $[8]$ . In  $[9]$ , a comparative study was recently done 128 63 mahdi.saleh@u-bordeaux.fr (M. Saleh), samir.omar@liu.edu.lb (S.-M. Omar). **between them.** The *l*-norm (with  $l = 1, 2, \infty$ ) between the true 129 To address this issue, one could consider the spectral dis-

Please cite this article in press as: E. Grivel et al., Interpreting the asymptotic increment of Jeffrey's divergence between some random processes, Digit. Signal Process. (2018), https://doi.org/10.1016/j.dsp.2018.01.006

<sup>9</sup> known as the relative entropy, generalized the notion of mutual of the AR correlation-matrices [34]. <sup>10</sup> information. In 1952, Chernoff [\[16\]](#page--1-0) introduced another measure **Concerning the above cases, we can summarize** the results we  $76$ 11 77 of divergence called Chernoff distance of order *λ*. In 1961, Rényi <sup>12</sup> [\[17\]](#page--1-0) suggested an extension of the entropy of order  $α$  for discrete 1) Links with Rao distance [35] have been proposed when it was <sup>78</sup> <sup>13</sup> probabilities. Then, the Rényi divergence of order  $\alpha$ , also called possible. It was confirmed that the square of the Rao distance was <sup>79</sup> 14 α-divergence, was introduced. Then, in 1960s, another degree of approximately twice the value of the JD, except when a 1st-order <sup>80</sup> <sup>15</sup> generalization was proposed through the so-called f-divergences MA process is considered whose zero is close to the unit-circle in <sup>81</sup> <sup>16</sup> where the probability density function (pdf) ratio is weighted by a the z-plane. <sup>17</sup> function *f*. They are also known as Csiszar *f*-divergences, Csiszar– 2) The JD tends to have a stationary regime. The difference be- <sup>83</sup> 18 Morimoto divergences or Ali–Silvey distances. Depending on the tween two JDs computed for *k* and (*k* − 1)-dimensional random <sup>84</sup> <sup>19</sup> choice of the function f, one can retrieve specific cases. Finally, vectors tends to a constant when k increases, except for a 1st-order <sup>85</sup> 20 f-dissimilarities have been introduced when more than two pdfs MA process when the zero is on the unit-circle in the z-plane. 86 <sup>21</sup> are considered. The reader may refer to [18] [19] for more details This difference is called asymptotic JD increment when *k* becomes <sup>87</sup> generalization was proposed through the so-called *f* -divergences Morimoto divergences or Ali–Silvey distances. Depending on the choice of the function *f* , one can retrieve specific cases. Finally, *f* -dissimilarities have been introduced when more than two pdfs are considered. The reader may refer to  $[18]$  [19] for more details and for information about recent works.

2 *E. Grivel et al. / Digital Signal Processing* ••• *(*••••*)* •••*–*•••

<sup>23</sup> Among the above measures, the KL divergence remains one of the asymptotic JD increment are provided for AR and/or MA pro- <sup>89</sup> <sup>24</sup> the most popular. Several authors analyzed it in various fields of cesses. They depend on the parameters of the processes. <sup>25</sup> applications, for classification, identification or change detection 3) The asymptotic JD increment can be used to compare the ran- <sup>91</sup> 26 [\[7\],](#page--1-0) [\[20\],](#page--1-0) [\[9\],](#page--1-0) [\[21\]](#page--1-0) and [\[22\].](#page--1-0) Meanwhile, the estimations of the KL alom processes instead of the JD between k successive samples of 32 27 between two pdfs that are not necessarily Gaussian, by using sets the processes. The resulting computation cost is smaller since the the 33 Among the above measures, the KL divergence remains one of of data, were studied. See [\[23\] \[24\].](#page--1-0)

<sup>29</sup> However, when dealing with Gaussian processes, the expression function of the parameters of the processes. Large-size matrix in-<sup>30</sup> of the KL depends on the logarithm of the ratio between the co- versions for different values of k and trace computations are no <sup>96</sup> <sup>31</sup> variance matrix determinants. Secondly, the KL divergence is not longer required. <sup>32</sup> a distance: it is not symmetric and does not satisfy the triangular (4) As the asymptotic JD increment does not depend on k, the se-<sup>33</sup> inequality. For the above reasons, a great deal of interest has been lection of the number of variates k is no longer a problem for the <sup>99</sup> <sup>34</sup> 100 paid to the symmetric KL divergence, known as Jeffrey's divergence 10 practitioner. <sup>35</sup> (JD) [\[25\].](#page--1-0) When dealing with the JD, the symmetry conditions are although various particular cases have been studied by taking <sup>101</sup> <sup>36</sup> satisfied. As the logarithms compensate each other, they no longer advantage of the expressions of the inverses of the correlation ma- <sup>102</sup>  $37$  appear in the expression of the JD for the Gaussian case. Given all trices [\[36\]\[34\],](#page--1-0) the JD between ergodic WSS ARMA processes of  $103$ 38 these considerations, we will focus our attention on the JD in the orders p and q, denoted as ARMA(p,q), has not been addressed yet 104 However, when dealing with Gaussian processes, the expression a distance: it is not symmetric and does not satisfy the triangular satisfied. As the logarithms compensate each other, they no longer appear in the expression of the JD for the Gaussian case. Given all these considerations, we will focus our attention on the JD in the following.

40 Mhen dealing with k-dimensional Gaussian random vectors of matrices. Therefore, it could be of interest to find an alternative 106 <sup>41</sup> size k, the JD amounts to computing the sum of two traces of approach. In this paper, we suggest giving an interpretation of the <sup>107</sup> 42 matrices, that can be expressed as the  $k \times k$  covariance matrix asymptotic JD increment in order to better understand the influ-  $^{108}$ <sup>43</sup> of the first process pre-multiplied by the inverse of the covari- ence of each process parameter on the JD and to generalize the <sup>109</sup> 44 ance matrix of the second process. When *k* increases, the resulting results we obtained to ARMA(p,q) processes. For this reason, our 110 <sup>45</sup> computational cost of the JD increases because the standard com-<br>first purpose is to provide a new way to derive the asymptotic <sup>111</sup> 46 putational burden of a generic  $k \times k$  matrix inversion is usually ID increment between WSS ARMA processes. It amounts to calcu-<br>112  $47$   $0(k^2)$  [26]. To address this problem with processes that are ergodic is a lating the power of the first process filtered by the inverse filter 113 48 WSS autoregressive (AR) and/or moving average (MA), eigenvalue associated with the second process and conversely. The second that 49 decomposition could be considered *a priori*. Analytical expressions contribution of this paper is to study the JD that involves sums of 115 50 of the eigenvalues and the eigenvectors exist for 1st-order MA complex exponentials that are disturbed by additive white noises 116 51 processes [\[27\].](#page--1-0) Concerning 1st-order AR processes, estimates of constress). Furthermore, at the end of the paper, a comparison be- that the saper and the paper and the saper be-52 eigenvalues have been proposed for a large correlation matrix [\[28\].](#page--1-0) Utween an AR process and an NSCE process is addressed by using 118 53 However, to the best of our knowledge, these estimates do not ex-<br>53 However, to the best of our knowledge, these estimates do not ex-<br> $\frac{1}{2}$  the ID. 54 st for higher-order AR processes. For this reason, for a pth-order this paper is organized as follows: in section [2,](#page--1-0) we briefly re-55 The process, a LDL factorization could be rather used and requires The all the definitions and properties of the processes under study. In the 121 56 the parameters of the AR process where the order varies between section 3, the expression of the ID is introduced and our contri- 122 57 1 and p. Its computational cost is of  $O(\frac{2k^3}{3})$  [\[26\].](#page--1-0) Moreover, al-<br>butions are presented. In section 4, we apply our interpretation of 123 <sup>58</sup> ternative approaches have been proposed. Taking advantage of the the asymptotic JD increment to various random processes. Illustra- <sup>124</sup> 59 Markovian properties of the AR process, the JD between the pdfs of tions are proposed in some cases, especially those based on NSCE 125  $60$  the k successive samples of two pth-order time-varying AR (TVAR) processes. Finally, in section [5](#page--1-0) we conclude our work. It is followed 126 <sup>61</sup> processes or AR processes can be recursively computed [\[29\].](#page--1-0) In this by an appendix that reveals some necessary derivations.  $127$ <sup>62</sup> case, the expression of the JD for *k*-dimensional vectors only de-<br>**In the following,**  $I_k$  **is the identity matrix of size** *k* **and Tr is the 128** <sup>63</sup> pends on matrices of size p, which significantly reduces the com- trace of a matrix. The upper-scripts <sup>1</sup> and <sup>H</sup> denote the transpose <sup>129</sup> <sup>64</sup> putational cost. Then, this method has been used to classify more and the Hermitian of a matrix.  $^{130}$   $^{130}$ <sup>65</sup> than two AR processes in different subsets [\[30\].](#page--1-0) The analytical ex-<br> $x_{k_1:k_2} = (x_{k_1},...,x_{k_2})$  is the collection of samples from time  $k_1$  <sup>131</sup> <sup>66</sup> pression of the JD between ergodic WSS 1st-order MA processes, to  $k_2$ . *l* is the label of the process under study.  $l = 1, 2$ . When dealing with *k*-dimensional Gaussian random vectors of size *k*, the JD amounts to computing the sum of two traces of matrices, that can be expressed as the  $k \times k$  covariance matrix of the first process pre-multiplied by the inverse of the covaricomputational cost of the JD increases because the standard com- $O(k^3)$  [\[26\].](#page--1-0) To address this problem with processes that are ergodic WSS autoregressive (AR) and/or moving average (MA), eigenvalue decomposition could be considered *a priori*. Analytical expressions ist for higher-order AR processes. For this reason, for a *p*th-order AR process, a LDL factorization could be rather used and requires Markovian properties of the AR process, the JD between the pdfs of the *k* successive samples of two *p*th-order time-varying AR (TVAR)

<sup>1</sup> model parameter vector and the estimated one could be also used, intal can be real or complex, noise-free or disturbed by additive 67  $^2$  as well as the COSH distance [\[10\].](#page--1-0) As an alternative, general dis-white Gaussian noises, has also been studied in [31]. For this pur-  $^{\rm 68}$ <sup>3</sup> tance measures [\[8\]](#page--1-0) can be considered. After the pioneering works pose, the authors use the analytical expression of each element <sup>69</sup> 4 70 of Pearson [\[11\]](#page--1-0) in 1900, Hellinger [\[12\]](#page--1-0) in 1909, Bhattacharyya in <sup>5</sup> 1943 [\[13\]](#page--1-0) and Shannon [\[14\]](#page--1-0) in 1948 where the measure of entropy AR processes, no recursive expression of the JD can be obtained for  $71$ <sup>6</sup> and the mutual information were introduced, several researchers and the-order MA processes. Finally, comparing ergodic WSS 1st-order <sup>72</sup> <sup>7</sup> focused their attentions on quantifying how close two distributions AR and ergodic WSS 1st-order MA processes by using the JD has <sup>73</sup> <sup>8</sup> are from one another: Kullback–Leibler (KL) divergence [\[15\],](#page--1-0) also been proposed in [33]. It is based on the expression of the inverses <sup>74</sup> that can be real or complex, noise-free or disturbed by additive white Gaussian noises, has also been studied in [\[31\].](#page--1-0) For this purpose, the authors use the analytical expression of each element of the tridiagonal-correlation-matrix inverse [\[32\].](#page--1-0) Unlike *p*th-order AR processes, no recursive expression of the JD can be obtained for 1st-order MA processes. Finally, comparing ergodic WSS 1st-order AR and ergodic WSS 1st-order MA processes by using the JD has been proposed in [\[33\].](#page--1-0) It is based on the expression of the inverses of the AR correlation-matrices [\[34\].](#page--1-0)

> Concerning the above cases, we can summarize the results we obtained as follows:

1) Links with Rao distance [\[35\]](#page--1-0) have been proposed when it was possible. It was confirmed that the square of the Rao distance was approximately twice the value of the JD, except when a 1st-order the z-plane.

22 and for information about recent works. The same state of the high. In previous papers [\[29\] \[31\] \[33\],](#page--1-0) analytical expressions of  $\frac{88}{3}$ 2) The JD tends to have a stationary regime. The difference becesses. They depend on the parameters of the processes.

28 94 analytical expression of the asymptotic JD increment is a known 3) The asymptotic JD increment can be used to compare the random processes instead of the JD between *k* successive samples of the processes. The resulting computation cost is smaller since the versions for different values of *k* and trace computations are no longer required.

> lection of the number of variates *k* is no longer a problem for the practitioner.

39 105 as there is no explicit expression for the inverses of the correlation Although various particular cases have been studied by taking results we obtained to ARMA(*p*,*q*) processes. For this reason, our JD increment between WSS ARMA processes. It amounts to calcucomplex exponentials that are disturbed by additive white noises (NSCE). Furthermore, at the end of the paper, a comparison between an AR process and an NSCE process is addressed by using the JD.

> section [3,](#page--1-0) the expression of the JD is introduced and our contributions are presented. In section [4,](#page--1-0) we apply our interpretation of the asymptotic JD increment to various random processes. Illustraby an appendix that reveals some necessary derivations.

> In the following,  $I_k$  is the identity matrix of size  $k$  and  $Tr$  is the trace of a matrix. The upper-scripts *<sup>T</sup>* and *<sup>H</sup>* denote the transpose and the Hermitian of a matrix.

 $x_{k_1:k_2} = (x_{k_1},...,x_{k_2})$  is the collection of samples from time  $k_1$ to  $k_2$ . *l* is the label of the process under study.  $l = 1, 2$ .

Download English Version:

<https://daneshyari.com/en/article/6951821>

Download Persian Version:

<https://daneshyari.com/article/6951821>

[Daneshyari.com](https://daneshyari.com)