THE BICEPSTRAL DISTANCE BETWEEN RANDOM SIGNALS: A NEW TOOL FOR COMPARISON OF ARMA MODELS IDENTIFICATION METHODS BASED ON HIGHER-ORDER STATISTICS

Jean-Luc Vuattoux and Eric Le Carpentier
Institut de Recherche en Cybernétique de Nantes, UMR CNRS 6597,
1 rue de la Noë, BP 92101, 44321 Nantes Cedex 03, FRANCE
Tel: +33 (0)2 40 37 16 46; fax: +33 (0)2 40 37 25 22
e-mail: Eric.Lecarpentier@lan.ec-nantes.fr

ABSTRACT

This paper deals with distance measures for signal processing and pattern recognition. It proposes a new distance between stationary random signals, called the bicepstral one, which can be easily converted in a distance between ARMA models. This distance is based on higher order statistics, and therefore is not phase blind. Thus, it provides a good tool for comparison of ARMA model identification methods based on higher order statistics.

1 INTRODUCTION

Distance measures between stationary random signals are widely used in signal processing for detection, clustering and pattern recognition (see [2] for a survey). Known distances are based on deviation measurement between power spectral density (cepstral distance, Itakura-Saïto distance...) or between probability laws under gaussian assumption (Kullback divergence...), and therefore are phase-blind.

In this paper, we present a new spectral distance, called the bicepstral distance, which is based on the third-order properties of the signals (section 2). This distance is particularly meaningful when an ARMA representation of observed signals is suitable (section 3). Furthermore, it provides a visual tool for ARMA model identification methods comparison (section 4), based on cumulated histogram of the distances between actual model and estimated ones, easier to interpret than usual analysis of mean and standard deviation of the estimated ARMA models parameters.

2 BICEPSTRAL DISTANCE BETWEEN SIGNALS

Let $\mathbf{y}_j = \{y_j[n]\}_n$, for $j = 1, 2$, two linear stationary non-gaussian signals and $S_{3,y_j}(\omega_1, \omega_2)$, for $j = 1, 2$, their bispectrum. Similarly to the cepstral distance [2], we defined the bicepstral distance between $\mathbf{y}_1$ and $\mathbf{y}_2$ as the L_2 norm of the difference of the logarithms of the bispectra $S_{3,y_1}(\omega_1, \omega_2)$ and $S_{3,y_2}(\omega_1, \omega_2)$:

$$d(\mathbf{y}_1; \mathbf{y}_2) = \\ \|\log S_{3,y_1}(\omega_1, \omega_2) - \log S_{3,y_2}(\omega_1, \omega_2)\|_2 \qquad (1)$$

where for any function $f(\omega_1, \omega_2)$:

$$\|f(\omega_1, \omega_2)\|_2 = \left[\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |f(\omega_1, \omega_2)|^2 \, \frac{d\omega_1 \, d\omega_2}{(2\,\pi)^2}\right]^{\frac{1}{2}} \qquad (2)$$

$d(\mathbf{y}_1; \mathbf{y}_2)$ is a true distance in the usual mathematical sense (triangle inequality, symmetry and positive definiteness) due to the L_2 norm mathematical properties.

It can be expressed in function of the bicepstra of $\mathbf{y}_1$ and $\mathbf{y}_2$, noted $bc_{y_j}(k, l)$ for $j = 1, 2$, where the bicepstrum is defined as the inverse 2-D Fourier transform of the logarithm of the bispectrum [4]. So:

$$\log S_{3,y}(\omega_1, \omega_2) = \sum_{k=-\infty}^{+\infty} \sum_{\ell=-\infty}^{+\infty} bc_y(k, \ell) \; e^{-\jmath\omega_1 k} \; e^{-\jmath\omega_2 \ell} \qquad (3)$$

By means of Parseval formula, we obtain:

$$d^2(\mathbf{y}_1; \mathbf{y}_2) = \sum_{k=-\infty}^{+\infty} \sum_{\ell=-\infty}^{+\infty} |bc_{y_1}(k, \ell) - bc_{y_2}(k, \ell)|^2 \qquad (4)$$

The equation (4) show that the bicepstral distance is nothing but the euclidean distance between the bicepstral coefficients, as the cepstral distance is the euclidean distance between the power cepstral coefficients [2]. Since the bicepstrum contains magnitude and phase information of the processes [4], it is clear that the new distance d may be used to differentiate spectrally equivalent non-gaussian processes with different phases, unlike the cepstral distance.

If the third-order statistics of the processes are not null, we can, as in the cepstral distance case, define the normalized bicepstral distance as the distance between processes having a log bispectral integral equals to zero. This distance, noted $\overline{d}(\mathbf{y}_1;\mathbf{y}_2)$, is defined by:

$$\overline{d}(\mathbf{y}_1;\mathbf{y}_2) = \left\| \log \frac{S_{3,y_1}(\omega_1,\omega_2)}{\lambda_{3,y_1}} - \log \frac{S_{3,y_2}(\omega_1,\omega_2)}{\lambda_{3,y_2}} \right\|_2 \quad (5)$$

where for $j = 1, 2$

$$\log \lambda_{3,y_j} = \frac{1}{(2\,\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log S_{3,y_j}(\omega_1,\omega_2)\; d\omega_1\; d\omega_2 \quad (6)$$

3 BICEPSTRAL DISTANCE BETWEEN ARMA MODELS

Let us suppose that the process $\mathbf{y} = \{y[n]\}_n$ can be modeled as the output of a discrete non-causal stable linear time invariant system of impulse response $\mathbf{h}_\theta$ with zero-mean, stationary, non-gaussian white input $\mathbf{x} = \{x[n]\}_n$ of skewness $\gamma_{3,x}$:

$$\mathbf{y} = \mathbf{h}_\theta * \mathbf{x} \quad (7)$$

where $*$ is the convolution sum. The z-transform H_θ of $\mathbf{h}_\theta$ is written under parametrical form:

$$H_\theta(z) = \frac{1 + \sum_{i=1}^{q} \beta[i]\; z^{-i}}{1 + \sum_{i=1}^{p} \alpha[i]\; z^{-i}} \quad (8)$$

with $\theta = [\beta[1] \ldots \beta[q]\; \alpha[1] \ldots \alpha[p]]^T$. We suppose that this ARMA(p,q) model, of known orders p and q, is free of zero-pole cancellations (no zeros or poles on the unit circle) and stable with stable inverse.

Let $(P_{\theta,i})_{1\le i\le n_a}$ be the poles inside the unit circle, $(\tilde{P}_{\theta,i})_{1\le i\le n_{\tilde{a}}}$ be the poles outside the unit circle $(n_a + n_{\tilde{a}} = p)$, $(Z_{\theta,i})_{1\le i\le n_b}$ be the zeros inside the unit circle, and $(\tilde{Z}_{\theta,i})_{1\le i\le n_{\tilde{b}}}$ be the zeros outside the unit circle $(n_b + n_{\tilde{b}} = q)$. It can be shown that:

$$\lambda_{3,y} = \gamma_{3,x} \left[\frac{\prod_{i=1}^{n_{\tilde{b}}}(-\tilde{Z}_{\theta,i})}{\prod_{i=1}^{n_{\tilde{a}}}(-\tilde{P}_{\theta,i})} \right]^3 \quad (9)$$

Then, the bicepstrum of $\mathbf{y}$ is defined by [4]:

$$bc_y(k,\ell) = \begin{cases} \log \lambda_{3,y} & k=0,\ \ell=0 \\ -\frac{1}{\ell}\, A_\theta^{(\ell)} & k=0,\ \ell>0 \\ -\frac{1}{k}\, A_\theta^{(k)} & k>0,\ \ell=0 \\ \frac{1}{k}\, B_\theta^{(-k)} & k<0,\ \ell=0 \\ \frac{1}{\ell}\, B_\theta^{(-\ell)} & k=0,\ \ell<0 \\ -\frac{1}{k}\, B_\theta^{(k)} & k=\ell>0 \\ \frac{1}{k}\, A_\theta^{(-k)} & k=\ell<0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where $A_\theta^{(k)}$ and $B_\theta^{(k)}$ are the cepstral parameters, defined in function of the poles and zeros of the ARMA model of parameterized impulse response h_θ as:

$$A_\theta^{(k)} = \sum_{i=1}^{n_b} Z_{\theta,i}^k - \sum_{i=1}^{n_a} P_{\theta,i}^k \quad (11)$$

$$B_\theta^{(k)} = \sum_{i=1}^{n_{\tilde{b}}} \tilde{Z}_{\theta,i}^{-k} - \sum_{i=1}^{n_{\tilde{a}}} \tilde{P}_{\theta,i}^{-k} \quad (12)$$

Let us now suppose that the two processes $\mathbf{y}_j$, for $j = 1, 2$, are represented by ARMA models of parameters θ_j, for $j = 1, 2$ and excited by two stationary non-gaussian white noise $\mathbf{x}_j$, for $j = 1, 2$. So, we can demonstrate from the equation (4) that:

$$d^2(\mathbf{y}_1;\mathbf{y}_2) = |\log\lambda_{3,y_1} - \log\lambda_{3,y_2}|^2 + 3\sum_{k=1}^{+\infty} \frac{1}{k^2} \left[\left(A_{\theta_2}^{(k)} - A_{\theta_1}^{(k)}\right)^2 + \left(B_{\theta_2}^{(k)} - B_{\theta_1}^{(k)}\right)^2 \right] \quad (13)$$

where the cepstral parameters $A_{\theta_j}^{(k)}$ and $B_{\theta_j}^{(k)}$, for $j = 1, 2$, are those of the processes $\mathbf{y}_j$, for $j = 1, 2$.

So, we can show easily by considering equations (5), (6) and (13) that the normalized bicepstral distance, noted subsequently $\overline{d}(\theta_1;\theta_2)$, is a parametric model distance measure obtained by:

$$\overline{d}^2(\theta_1;\theta_2) = 3\sum_{k=1}^{+\infty} \frac{1}{k^2} \left[\left(A_{\theta_2}^{(k)} - A_{\theta_1}^{(k)}\right)^2 + \left(B_{\theta_2}^{(k)} - B_{\theta_1}^{(k)}\right)^2 \right] \quad (14)$$

which depend only, from (11)-(12), on poles and zeros of ARMA models of impulse responses $\mathbf{h}_{\theta_1}$ and $\mathbf{h}_{\theta_2}$.

The interest of this distance is that if $\mathbf{h}_{\theta_1}$ and $\mathbf{h}_{\theta_2}$ are the impulse responses of different but spectrally equivalent ARMA models, then $d(\theta_1;\theta_2)$ will be not null, unlike the normalized cepstral distance.

In practice, the formula (13) and (14) are not usable due to the infinite sums. But since the cepstral parameters $A_{\theta_j}^{(k)}$ and $B_{\theta_j}^{(k)}$ decay exponentially [4], the series in (13) and (14) converge rapidly. So, a truncated series provides a good approximation of the exact normalized bicepstral distance:

$$\overline{d}_L^2(\theta_1;\theta_2) = 3\sum_{k=1}^{L} \frac{1}{k^2} \left[\left(A_{\theta_2}^{(k)} - A_{\theta_1}^{(k)}\right)^2 + \left(B_{\theta_2}^{(k)} - B_{\theta_1}^{(k)}\right)^2 \right] \quad (15)$$

where the integer L is chosen using [5], namely $L = \ln c / \ln a$ where c is a very small constant (say 10^{-4}) and $1 > a > \max\{|P_{\theta_j,i}|, |\tilde{P}_{\theta_j,i}^{-1}|, |Z_{\theta_j,i}|, |\tilde{Z}_{\theta_j,i}^{-1}|\}$.

Note that the normalized bicepstral distance $\overline{d}_L$ (15) is a bispectral distance between ARMA models which is not null if the two models have the same amplitude but a different phase.

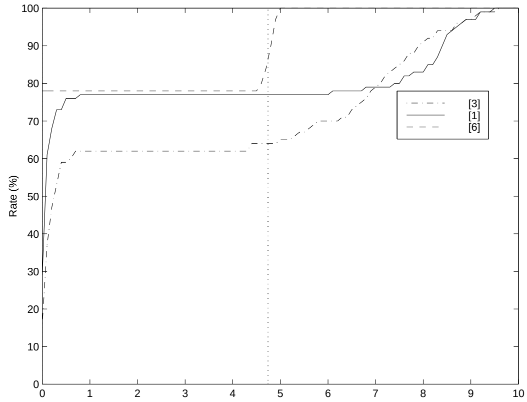

Figure 1: bicepstral distance distributions.

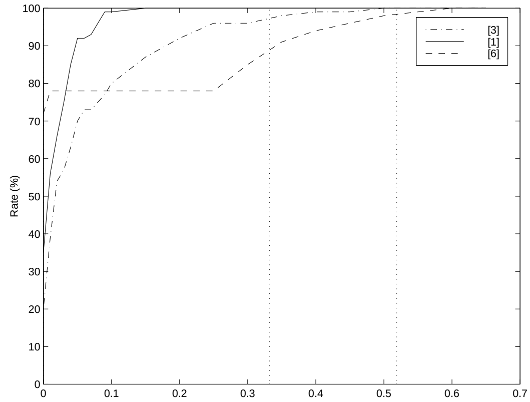

Figure 2: euclidean distance distributions.

4 COMPARISON OF IDENTIFICATION METHODS

The purpose of this section is to use the normalized bicepstral distance, and its properties, to measure the quality of ARMA parameters estimators and to compare them in simulations. Indeed, the classical way to compare identification methods consists in doing Monte-Carlo simulations (see [1, 3, 6]) with different models, length of signals and type of non-gaussian input noises. Then one compare the means and standard deviations of the parameter estimates calculated, for each method, on T independant Monte-Carlo runs.

So, we present a new way of comparison, using the normalized bicepstral distance (15), with the following principle:

1. Calculation, for each method, of the distance $\overline{d}$ (eq. 15) between the actual parameter θ_0 and the T estimated parameters $\widehat{\theta}_n$ for $n \in [1; T]$.

2. Graphic presentation of these results under the form of a distribution, for each method, of $\overline{d}^2(\theta_0; \widehat{\theta})$ where we find in ordinate the rate of estimated models having a distance to the actual model lower than the corresponding value obtained in abscissa.

3. Comparison of the obtained distributions: the best method is the one having a corresponding curve the nearest from the Y-axis.

Table 1: estimated parameters: mean $\pm$ std.

True	[3]	[1]	[6]
-2.05	-1.9718 $\pm$ 0.1718	-2.0642 $\pm$ 0.1136	-2.0475 $\pm$ 0.2031
1	0.9540 $\pm$ 0.1702	1.0154 $\pm$ 0.0936	1.0194 $\pm$ 0.2115

To illustrate this new way and the problems of the classical way, we will considering the following example. The actual MA(2) model is:

$$H_{\theta_0}(z) = 1 - 2.05\, z^{-1} + z^{-2}$$

with zeros located at 0.8 and $1/0.8 = 1.25$.

The identification methods compared here are those presented in [1], [3] and [6]. The methods of [1] and [3] are "linear algebra solutions" and use different slices of second and third-order cumulants. The one of [6] is a maximum likelihood approach and to avoid convergence to false local minima, the method of [1] is used to initialized the maximization procedure in [6].

So, the signal length used is $N = 2048$ samples and the input white noise $\mathbf{x}$ used is exponentially distributed ($\gamma_{2,x} = 1$ and $\gamma_{3,x} = 2$). Table 1 displays the estimated parameters (mean and standard deviation, averaged on $T = 100$ Monte-Carlo runs) for each method. So classically, from the Table 1, we could say that the method of [1] is the best one with lower bias and standard deviation.

Look at now the Figure 1, where we present the distributions of $\overline{d}^2(\theta_0; \widehat{\theta})$. We see that the methods of [1] and [6] provide about 80% of estimated models near of the actual model in the bispectral sense, against 60% only for [3]. And from point 2) of our comparison method, the method of [6] is the best one since its corresponding distribution is the nearest to the Y-axis. Note that we have indicated by vertical dotted lines the distance between the actual model and its exact spectrally equivalent models. So, from the Figure 1, the amplitude of the 20% of bad estimated models provided by [6] is correctly estimated and only their phase is false. On the other hand, for the method of [1], 20% of estimated models have erroneous amplitude and phase. Then, how to explain the results of Table 1?

We present in Figure 2 the distribution of the euclidean distance d_{eu} between the actual model and the estimated models:

$$d_{\mathrm{eu}}(\theta_0;\widehat{\theta}) = \sqrt{(\theta_0 - \widehat{\theta})^T\,(\theta_0 - \widehat{\theta})} \qquad (16)$$

Globally, the comparison obtained from Table 1 is the same as the one obtain with the distributions of d^2_{eu}. This is due to the fact that for each estimation of a parameter, the standard deviation is the euclidean distance to the mean value. Nevertheless, it was underlined that the euclidean distance between ARMA models is a bad dissimilarity measure (see [2]), because two close models in the sense of this distance may have very distinct spectral behaviours, the opposite being true too. This is confirmed by the simulations above: in Figure 2, all the estimated parameters provided by [1] seem close to the actual one, whereas Figure 1 shows that 20% of them correspond to highly different spectral behaviour.

So, it is clear from this example that the normalized bicepstral distance is a good tool to measure in simulations, the quality of ARMA parameters estimators in the non-gaussian case, unlike the classical method of bias and standard deviation. Furthermore, it permits to distinguish different types of estimation errors (on the phase and/or the amplitude) and to estimate the rate of estimated models having spectral and bispectral contents close to those of the actual one.

5 CONCLUSION

We proposed in this paper a new way to measure the quality of different ARMA model identification methods. It is based on a new distance measure: the bicepstral distance. This distance is a third-order extension of the well-known cepstral distance [2] and its interest comes from the fact that it can differentiate spectrally equivalent non-gaussian processes with different phases, unlike the cepstral distance.

This distance becomes particularly interesting in the practical case where an ARMA representation of observed signals is suitable and its normalized form is nothing but a distance between non-minimum phase ARMA models. So, we use this property to develop our new comparison method. Its interests, compared to the classical way of comparison used in the literature, is that it is a very discriminating graphical method which permits to distinguish different types of estimation errors and to detect methods which provides estimated models parametrically close to the actual one but spectrally far from it.

References

[1] S.A. Alshebeili, A.N. Venetsanopoulos, and A.E. Çetin. Cumulant based identification approaches for nonminimum phase FIR systems. *IEEE Trans. on S.P.*, 41(4):1576–1588, April 1993.

[2] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18:349–369, 1989.

[3] J.A.R. Fonollosa and J. Vidal. System identification using a linear combination of cumulant slices. *IEEE Trans. on S.P.*, 41(7):2405–2412, July 1993.

[4] C.L. Nikias and A.P. Petropulu. *Higher-order spectra analysis: a nonlinear signal processing framework*. Signal Processing Series. Prentice Hall, Englewood Cliffs, New Jersey, alan v. oppenheim, series editor edition, 1993.

[5] R. Pan and C.L. Nikias. The complex cepstrum of higher-order cumulants and nonminimum phase system identification. *IEEE Trans. on A.S.S.P.*, ASSP-36(2):186–205, February 1988.

[6] J.L. Vuattoux and E. Le Carpentier. Efficient ARMA parameter estimation of non-gaussian processes by minimization of the fisher information under cumulant constraints. In *Proceedings of IEEE Signal Processing Workshop SSAP'96*, pages 218–221, Corfou, Greece, June 1996.