Measurement 46 (2013) 2816–2827

Contents lists available at SciVerse ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Review

Higher-order statistics: Discussion and interpretation

Juan José González de la Rosa *,[1], Agustín Agüera-Pérez, José Carlos Palomares-Salas, Antonio Moreno-Muñoz

Research Group PAIDI-TIC-168: Computational Instrumentation and Industrial Electronics (ICEI), Spain
University of Cádiz, Area of Electronics, EPSA, Av. Ramón Puyol S/N, E-11202 Algeciras, Cádiz, Spain

ARTICLE INFO

Article history:
Received 17 January 2012
Received in revised form 22 April 2013
Accepted 29 April 2013
Available online 13 May 2013

Keywords:
Higher-order statistics (HOS)
Statistical signal processing
Time–frequency characterization

ABSTRACT

The focus of the present review paper is the interpretation and estimation of the measurements based in higher-order statistics both in the time and in the frequency domains; and is concerned with statistical signal processing applied in scientific methods. Throughout the work, the mathematical expressions are expanded in order to get a practical interpretation of the estimators and the resulting data structures. A comprehensive selection of the estimators allows the interpretation and the assessment of the numerical and graphical results. The developed expressions complement the examples, giving rise to a practical approach. Conveniently bricked within the sections, an ensemble of tests has been conducted, in which non-Gaussian processes show the need of a higher-order characterization.

Contents

1. The reasons for using HOS 2816
2. Statistical definitions of cumulants and moments 2817
3. Cumulants in multiple data records 2818
   3.1. Third and fourth-order characterization of statistical processes 2820
4. Poly-spectra 2820
5. Multi-dimensional HOS structures 2825
6. Conclusion 2827
   Acknowledgement 2827
   References 2827

1. The reasons for using HOS

In real-life measurement scenarios, non-expected deviations from the forecasted behavior are frequent, and getting an accurate performance of the electronic measurement equipment implies the calculation of as many statistical estimators as possible, specially in control and surveillance applications, dealing with low level signals from sensors, very susceptible to noise. This premise is also present in Nature's modeling, whose complexity implies the implicit use of accurate models, in order to get more understandability of the phenomena under study. As a consequence, computational complexity is increased substantially.

It is assumed that the default characterization for the measured data sequences is sustained in the traditional statistical density functions, like the normal distribution.

* Corresponding author at: Research Group PAIDI-TIC-168: Computational Instrumentation and Industrial Electronics (ICEI), Spain. Tel.: +34 956028020; fax: +34 956028001.

E-mail address: juanjose.delarosa@uca.es (J.J.G. de la Rosa).

URL: http://www.uca.es/grupos-inv/TIC168/ (J.J.G. de la Rosa).

[1] Main Researcher of the Research Unit PAIDI-TIC-168.

0263-2241/$ - see front matter
http://dx.doi.org/10.1016/j.measurement.2013.04.055

Gaussian processes are completely characterized by the autocorrelation sequence and its associated Fourier transform, the power spectrum, in the time and in the frequency domains, respectively. Nevertheless, these complementary estimators, which involve the time-domain product of two sequences of measurements, offer only a primary characterization of the measured data; e.g., in the power spectrum estimation, discrimination among phases of the frequency components is not possible. Consequently, there are numerous situations where we have to look beyond the autocorrelation in order to get extra information, regarding deviations from Gaussian behavior and non-linear characterization. These supplementary features help to distinguish among apparently similar measurement data structures, therefore getting the complete statistical characterization [1,2].

Time-domain estimators which have been obtained after multiplied more than two time-series, are called higher-order statistics. Their Fourier transforms are called polyspectra, and they contain additional information; e.g., the phase of the frequency components. The power spectrum (second-order spectrum) is a particular case of higher-order spectra. The third-ordered is called the bi-spectrum and the fourth-order spectrum is called the tri-spectrum.

In a simplistic manner, cumulants of order higher than 2 are all zero in signals with Gaussian probability density functions, i.e.: all cumulants of the normal distribution beyond the second are zero. Similarly, higher cumulants are blind to any kind of Gaussian process. This is why it is not possible to separate these signals using the statistical approach [3]. These premises are present in almost all HOS applications. In fact, these higher-order cumulants, are used to infer new properties about the data involved in non-Gaussian processes [3,4]. Prior to cumulants, due to the lack of analytical tools, such processes had to be treated as if they were Gaussian. To sum up, cumulants, in the time domain, and their associated Fourier transforms, known as poly-spectra, reveal information about amplitude and phase, whereas second order statistics (power, variance, covariance and spectra) are phase-blind [4].

The goal of this work is to explore and review HOS with a twofold purpose. By one side, mathematical expressions for the cumulants sequences have been expanded in order to get a more comprehensive view of the graphical representations of the data structures. Secondly, several examples are integrated in the work to show interpretation techniques regarding the results with cumulants and the practical limits of the measurement algorithms. The paper is structured as follows: in Section 2 a thorough revision is performed, gathering details and tricks when dealing with expressions. The following Section 3, provides with techniques to deal with multiple measurement channels. In this section, the statistical processes are characterized by using 3rd and 4th-order cumulants'. Section 4 comprises higher-order spectra and it shows a bi-spectrum estimator for quadratic phase coupling detection, and the performance of an estimator of the Spectral Kurtosis (SK) over a set of synthetic signals, which clearly shows the capability of noise rejection by HOS, along with performance issues regarding the recommended number of averaged data-registers (measurement realizations). Section 5 shows performance of HOS dealing with multidimensional data, which complements the example outlined in Section 4, and introduces the reader to the need of higher-order characterization and to the multi-dimensional data structures associated to HOS. Finally, conclusions are drawn in Section 6.

2. Statistical definitions of cumulants and moments

A continuous random variable X-mean μ; variance σ, and probability density $p(x)$-is completely characterized by its moments and cumulants. Moments are defined using the moment-generating function $M(t) \equiv E[e^{tX}]$,[2] and the *McLaurin* series. Similarly, cumulants are defined using the cumulant-generating function, $K(t)$, which is the Neperian logarithm of $M(t)$ [5]:

$$K(t) \equiv \log[M(t)] = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!} = \mu t + \sigma^2 \frac{t^2}{2!} + \cdots; \tag{1}$$

where it is assumed the existence of an $h > 0$, $h \in \Re$, such that for $|t| < h$; the existence of $K(t)$ is guaranteed. The first derivative, $K^{(1)}(t)$, and its polynomial expansion is detailed in the following equation:

$$\begin{aligned}
&K^{(1)}(t) = \sum_{r=0}^{\infty} \kappa_{r+1} \frac{t^r}{r!} = \underbrace{\mu}_{K^{(1)}(0)=\kappa_1} + \underbrace{\sigma^2}_{K^{(2)}(0)=\kappa_2} t + \cdots,\\
&\kappa_1 = K^{(1)}(0) = \mu,\\
&\kappa_2 = K^{(2)}(0) = \sigma^2,\\
&\cdots\\
&\kappa_r = K^{(r)}(0).
\end{aligned} \tag{2}$$

The coefficients of Eq. (2), are the cumulants; i.e. generically, $\kappa_r = K^{(r)}(0)$, is the rth-order cumulant [5].

While it is immediate to interpret cumulants up to a degree of 2, the higher-order cumulants are neither moments nor central moments, but rather more complicated polynomial expressions of the moments. They are related to each other via a recursive formula, described in the following equation:

$$\kappa_r = \mu'_r - \sum_{k=1}^{r-1} \binom{r-1}{k-1} \kappa_k \mu'_{r-k}. \tag{3}$$

As an example, and resulting from Eq. (3), the rth-order moment, μ'_r, is an rth-degree polynomial in the first r cumulants[3]:

$$\begin{aligned}
&\mu'_1 = \kappa_1,\\
&\mu'_2 = \kappa_2 + \kappa_1^2,\\
&\mu'_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3,\\
&\mu'_4 = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4,\\
&\mu'_5 = \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 15\kappa_2^2\kappa_1\\
&\quad + 10\kappa_2\kappa_1^3 + \kappa_1^5,\\
&\mu'_6 = \kappa_6 + 6\kappa_5\kappa_1 + 15\kappa_4\kappa_2 + 15\kappa_4\kappa_1^2 + 10\kappa_3^2\\
&\quad + 60\kappa_3\kappa_2\kappa_1 + 20\kappa_3\kappa_1^3 + 15\kappa_2^3 + 45\kappa_2^2\kappa_1^2 + 15\kappa_2\kappa_1^4 + \kappa_1^6,\\
&\ldots,
\end{aligned} \tag{4}$$

[2] $E[\cdot]$ is the expectation value operator.

[3] The apostrophe corresponds to non-central moments.