# Harmonic Retrieval Using Cumulant-Based Estimator for ARMA Systems

*Shing Tenqchen*[1a*], *Ying-Haw Shu*[1], *Ming-Chang Sun*[1], *Wu-Shiung Feng*[**]

[1]Lab 331, Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

*Chunghwa Telecom Telecommunication Labs., 12, Lane 551, Sec. 5,

Min-Tsu Rd., Yang-Mei Zien, Tao-Yuan County, Taiwan 326, R.O.C.

**Department of Electronics Engineering, Chang Gung University, Taoyuan, Taiwan, R.O.C.

E-mail: d86921028@ntu.edu.tw and stc@cht.com.tw

## ABSTRACT -

In this paper, we propose a cumulant-based estimator for ARMA systems. The polyspectra, cumulants, and various other related statistics, such as bicepstra and bicoherence are all used to develop cumulant-based algorithms for estimating the parameters of linear (e.g. ARMA) or nonlinear processes. The use of cumulant-based estimator is useful: (1) if the additive noise is Gaussian and the signal is non-Gaussian, (2) the linear system is non-minimum phase (that is, mixed-phase), or (3) the process is nonlinear. The singular value plot shows several significant singular values corresponding to one quadratically coupled triplet; as in the case of the power spectrum (harmonics in noise), overestimating the number of harmonics usually leads to better results.

## 1. Introduction

Blind Source Separation (BSS) is a fundamental problem in signal processing with a large number of applications in speech processing, array signal processing, multiuser communications, etc. [1]-[3]. The BSS problem consists in recovering the original sources from the observations only, without knowing the transmitted channel and sources. Most of the existing approaches to BSS have been developed for instantaneous mixtures. In practical situations, however, this hypothesis is not true and it is more common to find convolutive mixtures. Different algorithms [2]-[3] have been proposed to separate convolutive mixtures of sources in a blind way. In this paper we present a new approach based on the higher-order cumulants. Why do we need higher-order statistics? The motivation to use cumulants and polyspectra of order k>2 is given by the following ($(\mathbf{m}_k = (m_1, \cdots, m_{k-1}))$:

1) the power spectrum does not carry any information about the phase of H(f). In contrast, if $u(n)$ is non-Gaussian,

this phase information can be recovered from the higher-order polyspectra. Thus, the standard minimum-phase assumption may be dropped, which is necessary when the process is Gaussian or only second-order statistics are used.

2) The autocorrelation sequence cannot give any evidence of nonlinearity. In contrast, the higher-order cumulants can give evidence of nonlinearity.

This paper is organized as follows. The signal model of the preliminary results is presented in Section 2. Section 3 will discuss the cumulant-based estimator based on the autocorrelation carry over to the fourth-order cumulant. A simulation with ARMA model will be presented in Section 4 to show our result.

## 2. Signal Model: The Preliminary Results

The power spectrum is formally defined as the Fourier Transform (FT) of the autocorrelation sequence (the Wiener-Khintchine theorem [4])

$$P_{xx}(f) = \sum_{m=-\infty}^{\infty} R_{xx}(m) \exp(-j2\pi fm), \qquad (1)$$

where

$$R_{xx}(m) := E\{x^*(n)x(n+m)\}, \qquad (2)$$

and $f$ denotes the frequency. An equivalent definition is given by

$$P_{xx}(f) := E\{X(f)X^*(f)\} \qquad (3)$$

where X(f) is the Fourier Transform of $x(n)$

$$X(f) = \sum_{n=-\infty}^{\infty} x(n) \exp(-j2\pi fn). \qquad (4)$$

A sufficient, but not necessary, condition for the existence of the power spectrum is that the autocorrelation be absolutely summable. Higher-order moments are natural generations of the autocorrelation, and cumulants are specific nonlinear combinations of these moments. The fisrt order cumulant of a stationary process is the mean, $C_{1x} := E\{x(t)\}$. The higher-order cumulants are invariant to a shift of mean; hence, it is convenient to define them

under the assumption of zero mean; if the process has non-zero mean, we subtract the mean, and then apply the following definitions to the resulting process. The second-, third- and fourth-order cumulants of a zero-mean stationary process are defined by [4]. We omit here to save space. The $k$th-order polyspectrum is defined as the FTs of the corresponding cumulant sequence:

$$S_{2x}(f) = \sum_{k=-\infty}^{\infty} C_{2x}(k)e^{-j2\pi fk}, \qquad (5)$$

$$S_{3x}(f_1,f_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} C_{3x}(k,l)e^{-j2\pi f_1 k}e^{-j2\pi f_2 l}, \quad (6)$$

and

$$S_{4x}(f_1,f_2,f_3) = \sum_{k,l,m=-\infty}^{\infty} C_{4x}(k,l,m)e^{-j2\pi(f_1 k+f_2 l+f_3 m)}, \quad (7)$$

which are respectively the power spectrum, the bispectrum, and the trispectrum. The symmetry of bispectrum and trispectrum are given by [5]. Thus, it is also omitted here. Similar to cross-correlation, we can define cross-cumulants; for example,

$$C_{xyz}(k,l) = E\{x^*(n)y(n+k)z(n+l)\}. \qquad (8)$$

The cross-bispectrum is defined by

$$S_{xyz}(f_1,f_2) = \sum_{k=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{xyz}e^{-j2\pi f_1 k}e^{-j2\pi f_2 l}. \quad (9)$$

Note that the bispectrum $S_{3x}(f_1,f_2)$ is a special case of the cross-bispectrum obtained when $x = y = z$. The cross-bicoherence is another useful statistic which is defined as,

$$\text{bic}_{xyz}(f_1,f_2) = \frac{S_{xyz}(f_1,f_2)}{\sqrt{S_{2x}(f_1+f_2)S_{2y}(f_1)S_{2z}(f_2)}}. \quad (10)$$

The corss-bicepstrum of three processes is defined by

$$b_{xyz}(m,n) = \int\int \ln(S_{xyz}(f_1,f_2))e^{j2\pi f_1 m}e^{j2\pi f_2 n}df_1 df_2, \quad (11)$$

and is well-defined only if $S_{xyz}(f_1,f_2)$ is nonzero everywhere. The bias, by itself, does not completely characterize the estimate. If the estimate is good, we expect that $\hat{s}_N$ will take on values around the true quantity $s$. The nature measure of the spread is the squared deviation around the true quantity, $s$, $E\{|\hat{s}_N - s|^2\}$. The estimate is said to be asymptotically consistent if the squared deviation goes to zero, as $N \to \infty$. This condition is sometimes called mean square consistency. A consistent estimate is necessarily unbiased.

In practice, we have a finite amount of data, $\{x(n)\}_{n=0}^{N-1}$, and we must obtain consistent estimates of cumulants. The sample estimates are given by,

$$\hat{C}_{xy}(k) = \frac{1}{N_3} \sum_{n=N_1}^{N_2} x^*(n)y(n+k), \qquad (12)$$

$$\hat{M}_{xy}(k) = \frac{1}{N_3} \sum_{n=N_1}^{N_2} x(n)y(n+k) \qquad (13)$$

$$\hat{C}_{xyz}(k,l) = \frac{1}{N_3} \sum_{n=N_1}^{N_2} x^*(n)y(n+k)z(n+l), \quad (14)$$

$$\hat{C}_{wxyz}(k,l,m) = \frac{1}{N_3} \sum_{n=N_1}^{N_2} w^*(n)x(n+k)y(n+l)z^*(n+m)$$
$$\qquad (15)$$

$$-\hat{C}_{wx}(k)C_{yz}(l-m) - \hat{C}_{wy}(l)C_{xz}(k-m) - \hat{M}_{wz}(m)\hat{M}_{xy}(l-k).$$

Consider the harmonic retrieval problem. The observed data are of the form

$$y(n) = x(n) + w(n) = \sum_{k=1}^{P} \alpha_k \exp(j2\pi nf_k + j\phi_k) + w(n), (16)$$

where $w(n)$ is additive noise; $\alpha_k$'s are the amplitudes, $f_k$'s are the frequencies, and $\phi_k$'s are the phases. Additional assumption is that are often made regarding the phases are independent random variables uniformly distributed over $[0, 2\pi]$.

## 3. General Case: Cumulant-Based Estimators

In [6], it is estimated that the fourth-order cumulants of the process in (16) are given by

$$C_{4y}(l,m,n) = -\sum_{k=1}^{P} e^{j2\pi f_k(l+m-n)}|\alpha_k|^4 + C_{4w}(l,m,n).(17)$$

In particular, the diagonal slice is given by,

$$C_{4y}(m,m,m) = -\sum_{k=1}^{P} e^{j2\pi f_k m}|\alpha_k|^4 + C_{4w}(m,m,m).(18)$$

If, $w(n)$ is Gaussian, then $C_{4w}(m,m,m) \equiv 0$. If $w(n)$ is independent identical distribution (i.i.d.) non-Gaussian, then $C_{4w}(l,m,n)$ is a delta function at the origin. Consequently, all of the analyses based on the autocorrelation carry over to the fourth-order cumulant. In particular, the development of spectral estimates based on second-order statistics, can also be based on fourth-order statistics [7]. Similarly we can base ESPRIT on the fourth-order cross-cumulant matrix as well. The fourth-order statistics are most useful when the additive noise is narrow-band Gaussian. The number of

harmonics can usually be determined by examining the singular value plot generated by harmest routine [8].

The simplest nonlinear system is the second-order Voltera system whose input-output relationship is defined

$$y(n) = \sum_{k=0}^{\infty} h(k)x(n-k) + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} q(k,l)x(n-k)x(n-l). \quad (19)$$

The corresponding frequency domain representation is

$$Y(f) = H(f)X(f) + \sum_{f_1+f_2=f} Q(f_1+f_2)X(f_1)X(f_2), (20)$$

and is obtained by Fourier Transforming both side of (19). The cross-spectrum is given by

$$S_{yx}(f) = E\{Y(f)X^*(f)\} = H(f)E\{X(f)X^*(f)\} = H(f)S_{xx}(f). \quad (21)$$

Consider the cross-cumulant

$$C_{xxy}(\tau,\rho) = E\{x(n)x(n+\tau)y(n+\rho)\}$$

$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} q(k,l)E\{x(n)x(n+\tau)x(n+\rho-k)x(n+\rho-l)\}. \quad (22)$$

Assuming Gaussianity, we can rewrite the last equation in frequency domain, as

$$S_{yxx}(f_1,f_2) = 2Q(f_1,f_2)S_{xx}(f_1)S_{xx}(f_2) + S_{xx}(f_2)\delta(f_1+f_2)E\{y(n)\}. \quad (23)$$

## 4. Simulation

We simulate a non-Gaussian ARMA process, and then estimate its cumulants:
rand('seed',0); randn('seed',0);

u=rpiid(1024,'exp'); n=25;

y=filter([1,-2], [1,-1.5,0.8], u);

for k=-n:n,

cmat(:, k+n+1)=cumest(y, 3, n, 128, 0, 'biased', k);

end.

Time-series $y$ is segmented into records of 128 samples each, with no overlap; biased estimates of the third-order cumulants are obtained from each segment and then averaged; the (i j) element of cmat will contain the estimate

of $C_{3y}(i-n-1, j-n-1)$, for i, j = 1, ..., 2*n+1. You can use the function cumtrue [8] to compute and display the true cumulants.
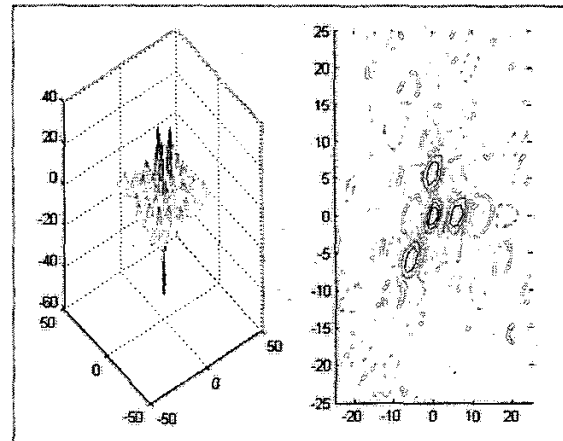


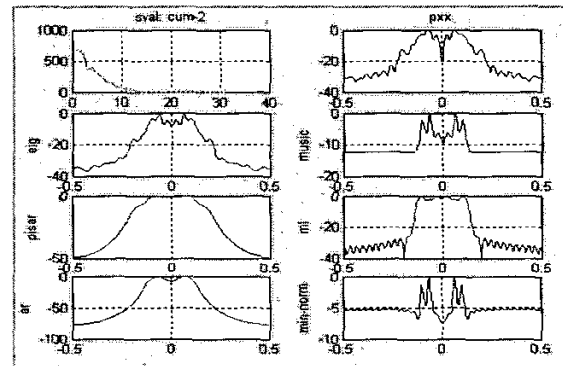Fig 1: Estimated Third-Order Cumulants of an ARMA(2,1) Process.



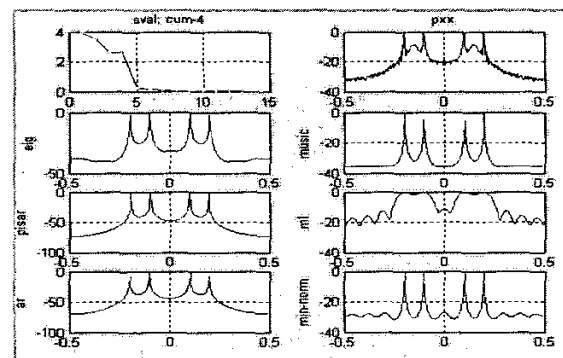Fig2: The output of Estimated 2nd-Order Cumulants of an ARMA(2,1) Process.

Fig3: The output of Estimated 4th-Order Cumulants of an ARMA(2,1) Process.
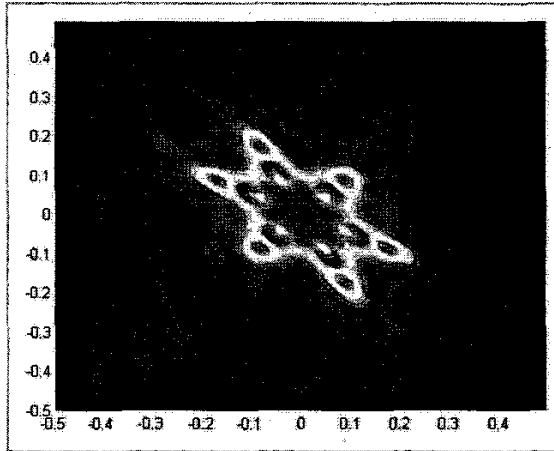


Fig4: The gaussianity and linearity tests of an ARMA(2,1) Process
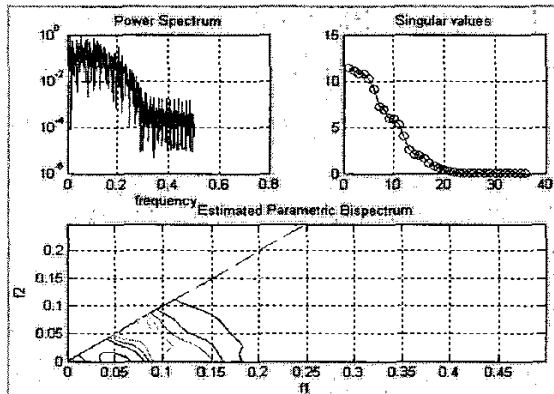


Fig5: The output of the estimated parametric Bispectrum for an ARMA(2,1) Process.

We see the display on Figure 5. Note that $f_1 + f_2 = 0.25$, which corresponds to the third peak in the amplitude spectrum, so that we may conclude that three of the four harmonics are quadratically phase coupled. The singular value plot shows several significant singular values corresponding to one quadratically coupled triplet; as in the case of the power spectrum (harmonics in noise), overestimating the number of harmonics usually leads to better results.

## Conclusion

In this paper, we presented a blind source separation method for nonlinear system. Note that both algorithms require access to inputs and outputs. Quadratic-phase coupling (QPC) can be detected. Cumulants are estimated for each segment and are then averaged across the set of segments; such segmentation usually speed up calculations, at a slightly loss in statistical efficiency. Note that segmented estimates can never better than unsegmented estimates. In the case of bispectra, we can apply an appropriate smoothing window. Be aware that the higher-order moments and cumulants of complex processed can be defined in different ways, and their polyspectra do not possess all the symmetry properties of their real counterparts.

## References

[1] K. I. Diamantaras and S. Y. Kung. *Principal Component Neural Networks*. Theory and Applications. Adaptive and Learning Systems for Signal Processing, Communications and Control. John Wiley & Sons Inc., New York, 1996.

[2] A. Cichocki, and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley & Sons, LTD, 2002.

[3] A. Cichocki, and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. John Wiley & Sons, New York, 1994. new revised and improved edition.

[4] Brillinger, D.R., "An Introduction to Polyspectra," *Ann. Math. Stat.*, Vol. 36, pp. 1351-1374, 1965.

[5] Rioul, O. and M. Vetterli, "Wavelets and signal Processing," *IEEE Signal Processing Magazine*, pp. 14-38, Oct. 1991.

[6] Swami, A. and J. M. Mendel, "Adaptive Cumulant-Based estimation of ARMA Parameters," *Proc. Amer. Control Conf.*, *ACC-88*, Atlanta, GA, pp. 2114-19, June 1988.

[7] Pan, R. and C. L. Nikias, "Harmonic decomposition methods in cumulant domains," *Proc. ICASSP-88*, pp. 2356-59, New York, 1988.

[8] Swami, A., J.M. Mendel, and C. L. Nikias, "Higher-Order Spectral Analysis Toolbox," *MatlabUser's Guide*, 2003.