Bispectrum Estimation: A Digital Signal Processing Framework

CHRYSOSTOMOS L. NIKIAS, MEMBER, IEEE, AND MYSORE R. RAGHUVEER, MEMBER, IEEE

It is the purpose of this tutorial paper to place bispectrum estimation in a digital signal processing framework in order to aid engineers in grasping the utility of the available bispectrum estimation techniques, to discuss application problems that can directly benefit from the use of the bispectrum, and to motivate research in this area. Three general reasons are behind the use of bispectrum in signal processing and are addressed in the paper: to extract information due to deviations from normality, to estimate the phase of parametric signals, and to detect and characterize the properties of nonlinear mechanisms that generate time series.

I. Introduction

Digital signal processing is a rapidly evolving computer engineering area which plays a key role in the industrial revolution of high technology that we are currently facing. It has been a useful field for more than twenty years where significant contributions have been made by combining ideas and methodologies from systems theory, statistics, numerical analysis, computer science, and very large scale integrated circuits (VLSI) technology. The final objective in a signal processing scenario is to process a finite set of data (either single-sensor or multiple sensor) and extract important information which is "hidden" in the data. This is usually achieved by combining the development of mathematical formulations which reach a certain level of estimation performance with their algorithmic implementation (either in software or hardware) and their application to real data. Various conflicting figures of merit are associated with digital signal processing techniques, namely, quality of the estimates, computational complexity or data throughput rate, cost of implementation, finite word-length effects, and structural properties.

A. Power Spectrum [15], [27], [31], [39], [40], [58], [61], [72], [74], [76]

One of the most fundamental and useful tools in digital signal processing has been the estimation of the power spectral density (PSD) or simply the power spectrum of discrete-time deterministic and stochastic processes. The past fifteen years witnessed an expansion of new power spectrum estimation techniques which have proved essential to the creation of advanced communication, radar, sonar, speech, biomedical, geophysical, and imaging systems. The available power spectrum estimation techniques may be considered in a number of separate classes, namely, conventional (or "Fourier type") methods, maximum-likelihood method of Capon with its modifications, maximum-entropy and minimum-cross-entropy methods, minimum energy, methods based on autoregressive (AR), moving-average (MA) and ARMA models, and harmonic decomposition methods such as Prony, Pisarenko, MUSIC, and Singular Value Decomposition. Research in this area has also led to signal modeling, and to extensions to multi-dimensional, multi-channel, and array processing problems. Each one of the aforementioned techniques has certain advantages and limitations not only in terms of estimation performance but also in terms of computational complexity and, therefore, depending on the signal environment, one has to choose the most appropriate.

In power spectrum estimation, the process under consideration is treated as a superposition of statistically uncorrelated harmonic components and the distribution of power among these frequency components is then estimated. As such, only linear mechanisms governing the process are investigated since phase relations between frequency components are suppressed [37]. The information contained in the power spectrum is essentially that which is present in the autocorrelation sequence; this would suffice for the complete statistical description of a Gaussian process of known mean. However, there are practical situations where we would have to look beyond the power spectrum (autocorrelation) to obtain information regarding deviations from Gaussianness and presence of nonlinearities. Higher order spectra (also known as polyspectra), defined in terms of higher order cumulants of the process, do contain such information. Particular cases of higher order spectra are the third-order spectrum also called the bispectrum which is, by definition, the Fourier transform of the third-order cumulant sequence, and the trispectrum (fourth-order spectrum) which is the Fourier transform of the fourth-order cumulant sequence of a stationary random process. Let us note that the power spectrum is, in fact, a member of the class of higher order spectra, i.e., it is a second-order spectrum.

Manuscript received July 10, 1986; revised January 21, 1987. This work was supported in part by the Office of Naval Research under Contract ONR-N00014-86-K-0219. The submission of this paper was encouraged after review of an advance proposal.

C. L. Nikias is with the Communications and Digital Signal Processing (CDSP) Laboratory, Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA.

M. R. Raghuveer is with Advanced Micro Devices, Inc., Austin, TX 78741, USA.

IEEE Log Number 8714327.

0018-9219/87/0700-0869$01.00 © 1987 IEEE

B. Motivation

The general motivation behind the use of higher order spectra in signal processing is threefold: 1) to extract information due to deviations from Gaussianness (normality), 2) to estimate the phase of non-Gaussian parametric signals, and 3) to detect and characterize the nonlinear properties of mechanisms which generate time series via phase relations of their harmonic components. The first motivation is based on the property that for Gaussian processes only, all polyspectra of order greater than two are identically zero [8], [37], [62]. Thus a nonzero higher order spectrum indicates deviation from normality. For a given zero-mean stationary real random process $\{X_n\}$, nonzero skewness, $E\{X_n^3\} \neq 0$, indicates the existence of its bispectrum. $E\{\cdot\}$ is the expectation operation. Hence, in those signal processing settings where the signal is a non-Gaussian stationary process and the additive noise process is stationary Gaussian, there might be certain advantages estimating signal parameters in higher order spectrum domains [97]. The non-Gaussianess condition is satisfied in many practical applications, since any periodic or quasi-periodic signal can be regarded as a non-Gaussian signal, and self-emitting signals from complicated mechanical systems can also be considered as non-Gaussian signals [99].

The second motivation is based on the fact that higher order spectra preserve the phase information of non-Gaussian parametric signals. For modeling time series data in signal processing problems, least squares estimation is almost exclusively used because it yields maximum-likelihood estimates of the parameters of Gaussian processes and because the equations obtained are usually in a linear form involving autocorrelation samples or their estimates [52]. However, the autocorrelation domain suppresses phase information and therefore least squares techniques (or modeling autocorrelation methods) are incapable of representing non-minimum phase parametric processes [104]. An accurate phase reconstruction in the autocorrelation (or power spectrum) domain can only be achieved if the parametric process is indeed minimum phase. Non-minimum phase estimation is of primary importance in deconvolution problems that arise in geophysics [60], [74], telecommunications [6], [75], etc., in which the wavelet shape must have the correct phase character. One approach to the deconvolution problem that has emerged recently explores the use of higher order spectra to estimate the phase of the wavelet due to the ability of polyspectra to preserve non-minimum phase information [47], [50], [89], [91]. Assuming that the reflectivity series (input sequence) is non-Gaussian white with zero-mean, a mixed-phase wavelet can be reconstructed in the bispectrum domain if the input sequence has nonzero skewness, or in the trispectrum domain if the fourth-order cumulant sequence is different from zero. Let us note, however, that if the reflectivity series is Gaussian, *no* procedure can recover the actual shape of a non-minimum phase wavelet [6].

Finally, introduction of higher order spectra (HOS) is quite natural when we try to analyze the nonlinearity of a system operating under a random input. General relations for arbitrary stationary random data passing through arbitrary linear systems have been studied quite extensively for many years. In principle, most of these relations are based on power spectrum (or autocorrelation) matching criteria. On the other hand, general relations are not available for arbitrary stationary random data passing through arbitrary nonlinear systems. Instead, each type of nonlinearity has to be investigated as a special case. HOS could play a key role in detecting and characterizing the type of nonlinearity in a system from its output data [5], [7], [34], [35], [57], [59], [110], [114]. Consider a linear time-invariant (LTI) system as shown in Fig. 1(a) with input

$$X_k = \sum_m A_m \exp j(\omega_m k + \phi_m) \tag{1.1}$$

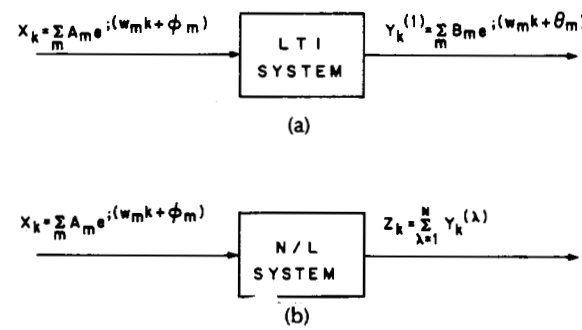


Fig. 1. (a) Output of a Linear Time-Invariant (LTI) system to a sinusoidal input (1.2). (b) Output of a nonlinear (N/L) system (*N*th-order Volterra) to a sinusoidal input (1.3).

where $\{\phi_m\}$ are independent, identically distributed (i.i.d) random variables, uniformly distributed over $[0, 2\pi]$. Then the output of the LTI system, $Y_k^{(1)}$, is given by

$$Y_k^{(1)} = \sum_m B_m \exp j(\omega_m k + \theta_m). \tag{1.2}$$

It can easily be verified that all higher order cumulants of $\{Y_k^{(1)}\}$ of order greater than two are identically zero. We therefore conclude that zero HOS of $\{Y_k^{(1)}\}$ will indicate that only linear mechanisms generate the output time series.

If we now pass the same process $\{X_k\}$ through a general nonlinear system, say an *N*th-order Volterra processor [105]

$$Z_k = \sum_{i_1=0}^{M_1-1} h_1(i_1) X_{k-i_1} + \sum_{i_1=0}^{M_2-1} \sum_{i_2=0}^{M_2-1} h_{12}(i_1, i_2) X_{k-i_1} X_{k-i_2} + \cdots + \sum_{i_1=0}^{M_N-1} \cdots \sum_{i_N=0}^{M_N-1} h_{12\cdots N}(i_1, i_2, \cdots, i_N) X_{k-i_1} \cdots X_{k-i_N}$$

as shown in Fig. 1(b), the output $\{Z_k\}$ will generally take the form

$$Z_k = \sum_{\lambda=1}^{N} Y_k^{(\lambda)} \tag{1.3}$$

where $\{Y_k^{(1)}\}$ is given by (1.2) and

$$Y_k^{(2)} = \sum_m \sum_n C_m C_n \exp j[(\omega_m + \omega_n)k + (\theta_m + \theta_n)]$$

$$Y_k^{(3)} = \sum_m \sum_n \sum_e D_m D_n D_e \cdot \exp j[(\omega_m + \omega_n + \omega_e)k + (\theta_m + \theta_n + \theta_e)]$$

etc.

A nonzero bispectrum (trispectrum) of $\{Z_k\}$ will indicate the existence of the term $\{Y_k^{(2)}\}$ ($\{Y_k^{(3)}\}$) in (1.3) and, therefore, the presence of a quadratic (cubic) nonlinearity in the

system. It therefore follows that nonlinear properties of mechanisms which generate time series can be quantified using HOS due to phase relations of the series harmonic components [8].

C. Application

Flow of energy in underwater (sonar), atmospheric (turbulence), geophysical (seismic), or even biological (brain, myocardium) volumes often takes the form of propagating waves and the goal of digital signal processing is to detect and characterize these waves from measurement sensor data. However, such nonlinearities as quadratic or cubic are usually present in these volumes causing generation of nonlinearly coupled waves. Power spectrum (or autocorrelation) based signal processing techniques applied on the sensor data cannot distinguish nonlinearly coupled from spontaneously excited independent waves with the same resonance conditions. In order to achieve this kind of distinction, one needs higher order spectra [25], [41], [42]. For example, due to quadratic nonlinearity, phase coupling between two frequency components of a process results in a contribution to the power at a frequency equal to their sum (resonance condition). Such coupling affects the third-order cumulant (moment) sequence and hence the bispectrum is used in detecting and characterizing such nonlinear effects.

There are several papers that have been published over the past twenty-five years dealing with the applications of HOS and especially that of the bispectrum. Examples include those in oceanography [20], [25], [66], geophysics and geoscience [14], [26], [33], [36], [88], [125], passive sonar [97], [99], biomedicine [37], [44], [53], telecommunications [6], [75], speech processing [124], economic time series [23], [35], fluid mechanics [29], [30], [32], [43], [48], [51], [55], [56], [62], [63], [120], [121], plasma physics [41], [42], and sunspot data analysis [9], [10]. HOS based on analog or optical estimation procedures have also been applied to the study of laser pulse shapes, radio-image reconstruction, sound quality, and mobility of bacteria [3], [4], [18], [19], [53], [54], [96], [101]–[103], [123]. Digital signal processing techniques that have been developed based on higher order spectra have been used for signal phase estimation and deconvolution [47], [49], [60], [67]–[70], [74], [75], signal detection [17], for identification of nonlinear [11], [13], [34], [35], [79], [95], non-minimum-phase [36], [65], [89], [91], and spike-array type processes [64], for time-delay estimation of arrivals [97], [98] and detection of quadratic phase coupling [41], [42], [82], [87] and discrimination of voices from fricative phonemes in speech [124]. A useful bibliography on HOS from 1953 to 1980 can be found in [115].

While considerable work has been done in the area of higher order spectrum estimation, mainly by a group of statisticians, the field is as yet to be developed and is far from successes achieved in the power spectrum estimation area. There is a very real need to develop better and faster estimation procedures for HOS and to identify application problems that could directly benefit from their use. We must point out, however, that the lack of sufficiently general theory which gives an overall understanding of polyspectra and the computational and statistical difficulties associated with them led people to a somewhat pessimistic attitude towards HOS [1], [2]. Some researchers have gone even as far as to name the subject "controversial" and "not useful."

Probably the most common reason for dissatisfaction has been the difficulty to compute and interpret higher order spectra as a whole [1]. However, the meaningfulness and understandability of HOS can easily be achieved when we have a proper model or good understanding of the environment that generates the observed data. Furthermore, in recent years we have witnessed an expansion of new microelectronic technologies that now permit more efficient algorithmic implementations. Due to these rapid hardware developments, and the growing availability of fast computers, serious consideration of HOS in engineering application problems has become timely.

D. Objective

In this tutorial review the emphasis will be on the third-order spectrum, also called the bispectrum. The goal of the paper is to place bispectrum estimation in a digital signal processing framework in order to

1) aid engineers in grasping the utility of the available conventional and modern bispectrum estimation techniques for specific applications;
2) motivate research in this area for development of more reliable and faster estimation procedures;
3) discuss and encourage identification of practical application problems in sonar, radar, biomedical, geophysical, and telecommunication signal processing which could directly benefit from bispectrum estimation.

We caution the reader to pay attention to the fact that higher order cumulants (or moments) do not resemble, in general, the properties of the autocorrelation function. In their paper, Huber *et al.* [37] sound a warning saying, "the newcomer to the field (meaning bispectrum estimation) still has to learn the hard way from his own mistakes."

The organization of the paper is as follows. Section II introduces the definition of cumulants and higher order spectra. Section III outlines the properties of the bispectrum and several application problems that can directly benefit from it. The family of conventional methods (or "Fourier type") for bispectrum estimation is introduced in Section IV. Parametric modeling of third-order moments for bispectrum estimation using MA, AR, and ARMA models is discussed in Section V. Section VI emphasizes three important signal processing applications of the bispectrum; namely, detection of quadratic phase coupling, signal phase estimation for deconvolution, and time delay estimation. Finally, Section VII is devoted to concluding remarks and possible future trends.

II. Definitions: Cumulants and Higher Order Spectra [1], [8], [90], [91], [106]–[108], [113], [122]

Higher order spectra are defined in terms of cumulants and therefore are also called cumulant spectra. Given a set of n real random variables $\{x_1, x_2, \cdots, x_n\}$, their joint cumulants of order $r = k_1 + k_2 + \cdots + k_n$ are defined as

$$c_{k_1 \cdots k_n} \triangleq (-j)^r \left. \frac{\partial^r \ln \Phi(\omega_1, \omega_2, \cdots, \omega_n)}{\partial \omega^{k_1} \partial \omega^{k_2} \cdots \partial \omega^{k_n}} \right|_{\omega_1 = \omega_2 = \cdots = \omega_n = 0} \tag{2.1}$$

where

$$\Phi(\omega_1, \omega_2, \cdots, \omega_n) \quad E\{\exp j(\omega_1 x_1 + \cdots + \omega_n x_n)\} \tag{2.2}$$

is their joint characteristic function. Let us note that the joint moments of order r of the same set of random variables are given by

$$m_{k_1 \cdots k_n} \triangleq E\{x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}\} = (-j)^r \left. \frac{\partial^r \Phi(\omega_1, \omega_2, \cdots, \omega_n)}{\partial \omega_1^{k_1} \cdots \partial \omega_n^{k_n}} \right|_{\omega_1 = \cdots = \omega_n = 0}. \tag{2.3}$$

Hence, the joint cumulants can be expressed in terms of the joint moments of the random variables. For example, if $m_{1 \cdots 0} = E\{x_1\} = 0$, then

$$\begin{aligned} c_{1 \cdots 0} &= 0 \\ c_{2 \cdots 0} &= m_{2 \cdots 0} = E\{x_1^2\} \\ c_{3 \cdots 0} &= m_{3 \cdots 0} = E\{x_1^3\} \\ c_{4 \cdots 0} &= m_{4 \cdots 0} - 3c_{2 \cdots 0}^2 \\ &= E\{x_1^4\} - 3m_{2 \cdots 0}^2. \end{aligned} \tag{2.4}$$

By taking $\{X(k)\}$, $k = 0, \pm 1, \pm 2, \cdots$ to be a real stationary random process with zero mean, $E\{X(k)\} = 0$, then the moment sequences of the process are related to its cumulants as follows:

$$\begin{aligned} E\{X(k)\, X(k + \tau_1)\} &= m_2(\tau_1) \\ &= c_2(\tau_1) \\ &\quad \text{(autocorrelation sequence)} \\ E\{X(k)\, X(k + \tau_1)\, X(k + \tau_2)\} &= m_3(\tau_1, \tau_2) \\ &= c_3(\tau_1, \tau_2) \\ &\quad \text{(third-order moment or cumulant sequence)} \\ E\{X(k)\, X(k + \tau_1)\, X(k + \tau_2)\, X(k + \tau_3)\} &= m_4(\tau_1, \tau_2, \tau_3) \\ &= c_4(\tau_1, \tau_2, \tau_3) + c_2(\tau_1) \cdot c_2(\tau_3 - \tau_2) \\ &\quad + c_2(\tau_2)\, c_2(\tau_3 - \tau_1) + c_2(\tau_3) \cdot c_2(\tau_2 - \tau_1) \\ &\quad \text{(fourth-order moment sequence)} \end{aligned} \tag{2.5}$$

etc.

While the third-order moments and third-order cumulants are identical, this is not true for the fourth-order statistics. In order to generate the fourth-order cumulant sequence, we need knowledge of the fourth-order moment and autocorrelation sequences.

The Nth-order spectrum $C(\omega_1, \omega_2, \cdots, \omega_{N-1})$ of the process $\{X(k)\}$ is defined as the Fourier transform of its Nth-order cumulant sequence $c_N(\tau_1, \tau_2, \cdots, \tau_{N-1})$, i.e.,

$$C(\omega_1, \omega_2, \cdots, \omega_{N-1}) = \sum_{\tau_1 = -\infty}^{+\infty} \cdots \sum_{\tau_{N-1} = -\infty}^{+\infty} \cdot c_N(\tau_1, \tau_2, \cdots, \tau_{N-1}) \cdot \exp\{-j(\omega_1 \tau_1 + \cdots + \omega_{N-1} \tau_{N-1})\}. \tag{2.6}$$

In general, $C(\omega_1, \cdots, \omega_{N-1})$ is complex and a sufficient condition for its existence in that $c_N(\tau_1, \tau_2, \cdots, \tau_{N-1})$ is absolutely summable. The notion of considering a spectral representation for a cumulant sequence as shown in (2.6) (cumulant spectrum) is acknowledged to be due to Kolmogorov [8]. It should be noted that the term polyspectrum is due to Tukey [115] whereas the term higher order spectrum is due to Brillinger and Rosenblatt [9], [10].

The power spectrum, bispectrum, and trispectrum are special cases of the Nth-order spectrum defined by (2.6), i.e.,

Power Spectrum: $N = 2$

$$C(\omega_1) = \sum_{\tau_1 = -\infty}^{+\infty} c_2(\tau_1) \exp\{-j(\omega_1 \tau_1)\} \quad \text{(Wiener-Kinchine).} \tag{2.7}$$

Bispectrum: $N = 3$

$$C(\omega_1, \omega_2) = \sum_{\tau_1 = -\infty}^{+\infty} \sum_{\tau_2 = -\infty}^{+\infty} c_3(\tau_1, \tau_2) \exp\{-j(\omega_1 \tau_1 + \omega_2 \tau_2)\}. \tag{2.8}$$

Trispectrum: $N = 4$

$$C(\omega_1, \omega_2, \omega_3) = \sum_{\tau_1 = -\infty}^{+\infty} \sum_{\tau_2 = -\infty}^{+\infty} \sum_{\tau_3 = -\infty}^{+\infty} c_4(\tau_1, \tau_2, \tau_3) \cdot \exp\{-j(\omega_1 \tau_1 + \omega_2 \tau_2 + \omega_3 \tau_3)\}. \tag{2.9}$$

At this point, a natural question that arises is why the Nth-order spectrum (or polyspectrum) is defined as the Fourier transform of the cumulant rather than of the moment sequence of $\{X(k)\}$. The reason is twofold: a) if $\{X(k)\}$ is a stationary Gaussian random process, than all its Nth-order moments for $N \geq 3$ do not provide any additional information pertaining to the process. It is, therefore, better to have a function that shows this fact explicitly. The cumulant spectrum function does so since higher order ($N \geq 3$) cumulants are zero for Gaussian processes; b) if the random variables $\{x_1, \cdots, x_n\}$ can be divided into any two or more groups which are statistically independent, their Nth-order cumulants are identically zero [8]. Hence, cumulant spectra provide a suitable measure of statistical dependence. Finally, Brillinger [8] points out that ergodicity requirements are met more easily with cumulants than moments.

III. Properties of the Bispectrum

Throughout this paper we concentrate on real discrete processes. Let $\{X(k)\}$ be a real, discrete, zero-mean stationary process with power spectrum $P(\omega)$, defined as

$$P(\omega) = \sum_{\tau = -\infty}^{+\infty} r(\tau) \exp -j(\omega\tau), \qquad |\omega| < \pi \tag{3.1}$$

where

$$r(\tau) = E\{X(k)\, X(k + \tau)\} \tag{3.2}$$

is its autocorrelation sequence. If $R(m, n)$ denotes the third moment sequence of $\{X(k)\}$, i.e.,

$$R(m, n) = E\{X(k)\, X(k + m)\, X(k + n)\} \tag{3.3}$$

then its bispectrum is defined as

$$B(\omega_1, \omega_2) \quad \sum_{m = -\infty}^{+\infty} \sum_{n = -\infty}^{+\infty} R(m, n) \exp -j(\omega_1 m + \omega_2 n). \tag{3.4}$$

Since the third-order moments and cumulants are identical, the bispectrum is a third-order cumulant spectrum.

The physical significance of the power spectrum and bispectrum becomes apparent when expressed in terms of the components $dZ(\omega)$ of the Fourier–Stieltjes representation of $X(k)$ (Cramér spectral representation) [16], [25]

$$X(k) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{j\omega k}\, dZ(\omega) \tag{3.5a}$$

for all k, where

$$E\{dZ(\omega)\} = 0$$

$$E\{dZ(\omega_1)\, dZ^*(\omega_2)\} = \begin{cases} 0, & \omega_1 \neq \omega_2 \\ 2\pi P(\omega)\, d\omega, & \omega_1 = \omega_2 = \omega \end{cases}$$

and

$$E\{dZ(\omega_1)\, dZ(\omega_2)\, dZ^*(\omega_3)\} = \begin{cases} B(\omega_1, \omega_2)\, d\omega_1\, d\omega_2, & \omega_1 + \omega_2 = \omega_3 \\ 0, & \omega_1 + \omega_2 \neq \omega_3. \end{cases} \tag{3.5b}$$

It is therefore apparent that the power spectrum $P(\omega)$ represents the contribution to the mean product of two Fourier components whose frequencies are the same whereas the bispectrum $B(\omega_1, \omega_2)$ represents the contribution to the mean product of three Fourier components where one frequency equals the sum of the other two.

Important symmetry conditions follow from the above definitions. From (3.1) and (3.2) we have

$$r(\tau) = r(-\tau)$$

$$P(\omega) = P(-\omega)$$

$$P(\omega) \geq 0 \quad \text{(real, nonnegative function).} \tag{3.6}$$

From (3.3) it follows that the third moments obey the symmetry properties [84], [85], [88], [89]

$$\begin{aligned} R(m, n) &= R(n, m) \\ &= R(-n, m - n) \\ &= R(n - m, -m) \\ &= R(m - n, -n) \\ &= R(-m, n - m). \end{aligned} \tag{3.7}$$

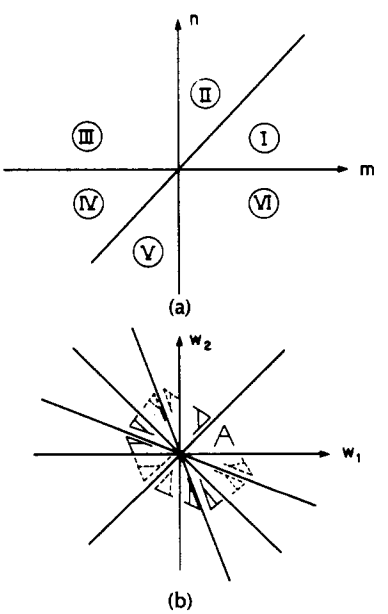


Fig. 2. (a) Symmetry regions of third-order moments (3.7). (b) Symmetry regions of the bispectrum (3.10) (after [64]).

As a consequence, knowing the third moments in any *one* of the six sectors, (I) through (VI), shown in Fig. 2(a), would enable us to find the entire third moment sequence. These sectors include their boundaries so that, for example, sector (I) is an infinite wedge bounded by the lines $m = 0$, and $m = n$; $m, n \geq 0$.

From the definition of the bispectrum in (3.4) and the properties of third moments in (3.7), it follows that

i) $B(\omega_1, \omega_2)$ is generally complex, i.e., it has magnitude and phase

$$B(\omega_1, \omega_2) = |B(\omega_1, \omega_2)| \exp j\psi_B(\omega_1, \omega_2). \tag{3.8}$$

ii) $B(\omega_1, \omega_2)$ is doubly periodic with period 2π

$$B(\omega_1, \omega_2) = B(\omega_1 + 2\pi, \omega_2 + 2\pi). \tag{3.9}$$

iii) $$\begin{aligned} B(\omega_1, \omega_2) &= B(\omega_2, \omega_1) = B^*(-\omega_2, -\omega_1) \\ &= B^*(-\omega_1, -\omega_2) = B(-\omega_1 - \omega_2, \omega_2) \\ &= B(\omega_1, -\omega_1 - \omega_2) = B(-\omega_1 - \omega_2, \omega_1) \\ &= B(\omega_2, -\omega_1 - \omega_2). \end{aligned} \tag{3.10}$$

Thus knowledge of the bispectrum in the triangular region $\omega_2 \geq 0$, $\omega_1 \geq \omega_2$, $\omega_1 + \omega_2 \leq \pi$ shown in Fig. 2(b) is enough for a complete description of the bispectrum. It is worth noting that the computation of $B(\omega_1, \omega_2)$ in (3.4) is done over one of the twelve sectors shown in Fig. 2(b) and the symmetries (3.10) are then utilized.

Additional properties of the bispectrum that make it very attractive in practical application problems are outlined below.

1) Gaussian Processes: If $\{X(k)\}$ is a stationary zero-mean Gaussian process, its third-moment sequence $R(m, n) = 0$ for all (m, n) and therefore its bispectrum $B(\omega_1, \omega_2)$ is identically zero [37].

2) Linear Phase Shifts: Given $\{X(k)\}$ with power spectrum $P_x(\omega)$ and bispectrum $B_x(\omega_1, \omega_2)$, the process $Y(k) = X(k - N)$, where N is a constant integer, has power spectrum $P_y(\omega) = P_x(\omega)$ and bispectrum $B_y(\omega_1, \omega_2) = B_x(\omega_1, \omega_2)$, i.e., the second- and third-order moments suppress *linear* phase information. Let us note, however, that while the power spectrum (autocorrelation) suppresses *all* phase information, the bispectrum (third-moment sequence) does not [47], [49], [50].

3) Non-Gaussian White Noise: If $\{W(k)\}$ is a stationary non-Gaussian process with $E\{W(k)\} = 0$, $E\{W(k)\, W(k + \tau)\} = Q \cdot \delta(\tau)$, and $E\{W(k)\, W(k + \tau)\, W(k + \rho)\} = \beta \cdot \delta(\tau, \rho)$, its power spectrum and bispectrum are both flat, i.e., $P(\omega) = Q$ and $B(\omega_1, \omega_2) = \beta$.

4) Quadratic Phase Coupling: There are situations in practice where because of interaction between two harmonic components of a process there is contribution to the power at their sum and/or difference frequencies. Such a phenomenon which could be due to quadratic nonlinearities gives rise to certain phase relations called quadratic phase coupling. In certain applications it is necessary to find out if peaks at harmonically related positions in the power spectrum are, in fact, coupled. Since the power spectrum suppresses all phase relations it cannot provide the

answer. The bispectrum, however, is capable of detecting and quantifying phase coupling [41], [42], [82], [86], [87]. This is best illustrated by the following simple example. Consider the processes

$$X_I(k) = \cos(\lambda_1 k + \varphi_1) + \cos(\lambda_2 k + \varphi_2) + \cos(\lambda_3 k + \varphi_3) \quad (3.11a)$$

and

$$X_{II}(k) = \cos(\lambda_1 k + \varphi_1) + \cos(\lambda_2 k + \varphi_2) + \cos(\lambda_3 k + (\varphi_1 + \varphi_2)) \quad (3.11b)$$

where $\lambda_3 = \lambda_1 + \lambda_2$, i.e., $(\lambda_1, \lambda_2, \lambda_3)$ are harmonically related and $\varphi_1, \varphi_2, \varphi_3$ are independent random variables uniformly distributed between $[0, 2\pi]$. From (3.11a), it is apparent that λ_3 is an independent harmonic component because φ_3 is an independent random-phase variable. On the other hand, λ_3 of $X_{II}(k)$ in (3.11b) is a result of phase coupling between λ_1 and λ_2. One can easily verify that $X_I(k)$ and $X_{II}(k)$ have *identical* power spectra ($P_I(\omega) = P_{II}(\omega)$) consisting of impulses at λ_1, λ_2, and λ_3. However, the bispectrum of $X_I(k)$ is identically zero whereas the bispectrum of $X_{II}(k)$ shows an impulse in the triangular region $\omega_2 \geq 0$, $\omega_1 \geq \omega_2$, $\omega_1 + \omega_2 \leq \pi$. The impulse is located at $\omega_1 = \lambda_1, \omega_2 = \lambda_2$ if $\lambda_1 \geq \lambda_2$. The quadratic phase coupling property of the bispectrum for this example is illustrated in Fig. 3. We elaborate more on quadratic phase coupling later in the paper.

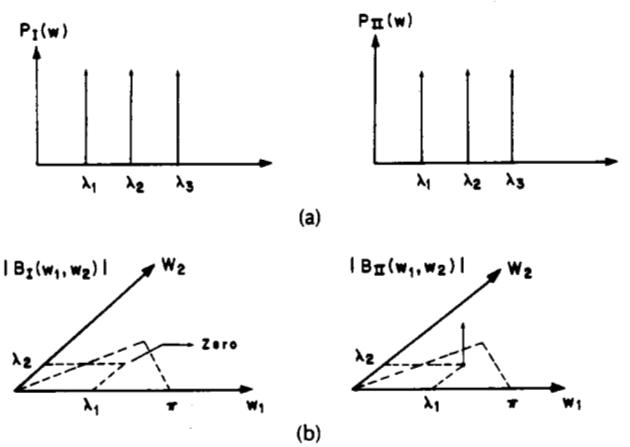


Fig. 3. Quadratic phase coupling. (a) Power spectrum and magnitude bispectrum of $X_I(k)$ described by (3.11a) and (b) of $X_{II}(k)$ described by (3.11b).

5) Non-Gaussian Process into a Linear Filter [10], [12], [47], [49], [50], [89], [94]: Let $\{X(k)\}$ be a zero-mean stationary non-Gaussian process with power spectrum $Px(\omega)$ and bispectrum $Bx(\omega_1, \omega_2)$. Put

$$Y(k) = \sum_{s=-\infty}^{+\infty} h(k - s)\, X(s) \quad (3.12)$$

where

$$h(k) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} H(\omega) \exp(j\omega k)\, d\omega. \quad (3.13)$$

Then the power spectrum of $\{Y(k)\}$ is

$$Py(\omega) = |H(\omega)|^2\, Px(\omega) \quad (3.14)$$

and the bispectrum is

$$By(\omega_1, \omega_2) = H(\omega_1)\, H(\omega_2)\, H^*(\omega_1 + \omega_2)\, Bx(\omega_1, \omega_2). \quad (3.15)$$

If

$$H(\omega) = |H(\omega)| \exp j\varphi(\omega) \quad (3.16)$$

then (3.15) can be written as

$$|By(\omega_1, \omega_2)| = |H(\omega_1)| \cdot |H(\omega_2)| \cdot |H(\omega_1 + \omega_2)| \cdot |Bx(\omega_1, \omega_2)| \quad (3.17a)$$

and

$$\psi_y(\omega_1, \omega_2) = \varphi(\omega_1) + \varphi(\omega_2) - \varphi(\omega_1 + \omega_2) + \psi_x(\omega_1, \omega_2). \quad (3.17b)$$

For the special case where $\{X(k)\}$ is non-Gaussian white, then, $|Bx(\omega_1, \omega_2)| = \beta$ and $\psi_x(\omega_1, \omega_2) = 0$. If $\{X(k)\}$ is stationary Gaussian, $Bx(\omega_1, \omega_2) = 0$ and hence $By(\omega_1, \omega_2) = 0$.

6) Minimum-, Maximum-, and Mixed-Phase Linear Filters [49], [67]–[70], [74], [75]: The fact that the third-moment sequence retains phase information unlike the autocorrelation (except, of course, for the linear phase component) also makes the bispectrum useful in identifying non-minimum (mixed) phase systems or sequences. We demonstrate this property of the bispectrum by a simple example. Consider the following second-order Finite Impulse Response (FIR) filters driven by a non-Gaussian stationary white noise process $\{W(k)\}$ with $E\{W(k)\} = 0$, $E\{W(k)\, W(k + \tau)\} = Q \cdot \delta(\tau)$, and $E\{W(k)\, W(k + \tau)\, W(k + \rho)\} = \beta \cdot \delta(\tau, \rho)$:

- *Minimum-Phase Filter:*

$$H_1(z) = (1 - az^{-1})(1 - bz^{-1})$$

$$Y_1(n) = W(n) - (a + b)\, W(n - 1) + abW(n - 2) \quad (3.18a)$$

where $0 < a < 1$, $0 < b < 1$.

- *Maximum-Phase Filter:*

$$H_2(z) = (1 - az)(1 - bz)$$

$$Y_2(n) = W(n) - (a + b)\, W(n + 1) + abW(n + 2). \quad (3.18b)$$

- *Mixed-Phase Filter:*

$$H_3(z) = (1 - az)(1 - bz^{-1})$$

$$Y_3(n) = -aW(n + 1) + (1 + ab)\, W(n) - bW(n - 1). \quad (3.18c)$$

The output sequences $\{Y_1(n)\}$, $\{Y_2(n)\}$, and $\{Y_3(n)\}$ have identical autocorrelation sequences given by

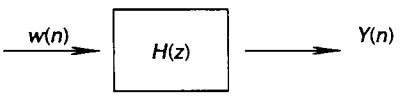
$$\begin{aligned} r(\tau) &= E\{Y_1(n)\, Y_1(n + \tau)\} \\ &= E\{Y_2(n)\, Y_2(n + \tau)\} \\ &= E\{Y_3(n)\, Y_3(n + \tau)\} \\ r(0) &= 1 + a^2b^2 + (a + b)^2 \\ r(1) &= -(a + b)(1 + ab) \\ r(2) &= ab \\ r(\tau) &= 0, \quad \text{for } \tau \geq 3 \end{aligned} \quad (3.19)$$

which implies that they have identical power spectral densities, i.e.,

$$P(\omega) = |H_1(z)|^2 = |H_2(z)|^2 = |H_3(z)|^2, \quad z = \exp j\omega \quad (3.20)$$

Table 1 Minimum, Maximum and Mixed-Phase Process with Identical Power Spectra (or Autocorrelations): $0 < a < 1, 0 < b < 1$

		Minimum Phase	Maximum Phase	Mixed Phase
	$H(z)$	$(1 - az^{-1})(1 - bz^{-1})$	$(1 - az)(1 - bz)$	$(1 - az)(1 - bz^{-1})$
	$Y(n)$	$w(n) - (a + b)w(n - 1) + abw(n - 2)$	$w(n) - (a + b)w(n + 1) + abw(n + 2)$	$-aw(n + 1) + (1 + ab)w(n) - bw(n - 1)$
Third Moments	$R(0, 0)$	$1 - (a + b)^3 + a^3b^3$	$1 - (a + b)^3 + a^3b^3$	$(1 + ab)^3 - a^3 - b^3$
	$R(1, 1)$	$(a + b)^2 - (a + b)a^2b^2$	$-(a + b) + ab(a + b)^2$	$-a(1 + ab)^2 + (1 + ab)b^2$
	$R(2, 2)$	a^2b^2	ab	$-ab^2$
	$R(1, 0)$	$-(a + b) + ab(a + b)^2$	$(a + b)^2 - (a + b)a^2b^2$	$a^2(1 + ab) - (1 + ab)^2b$
	$R(2, 0)$	ab	a^2b^2	$-a^2b$
	$R(2, 1)$	$-(a + b)ab$	$-(a + b)ab$	$ab(1 + ab)$
Auto-correlations	$r(0)$		$1 + a^2b^2 + (a + b)^2$	
	$r(1)$		$-(a + b)(1 + ab)$	
	$r(2)$		ab	



which is a somewhat expected result because the magnitude squared of an FIR transfer function does not recognize a zero (z_0) from its reciprocal ($1/z_0^*$). On the other hand, bispectral measurements provide a way of making the correct identification because $\{Y_1(n)\}$, $\{Y_2(n)\}$, and $\{Y_3(n)\}$ have different third-moment sequences and thus different bispectra as illustrated in Table 1. This property of the bispectrum has important implications for deconvolution in geophysics [60], communications [6], [75], etc.

7) Processes whose Spectrum and Bispectrum are Modeled by Different Linear Filters [86], [87]: Suppose a process has spectrum $P(\omega)$ and bispectrum $B(\omega_1, \omega_2)$. Then finding a linear filter with transfer function $H(\omega)$ to match the power spectrum amounts to solving

$$P(\omega) = |H(\omega)|^2 \tag{3.21a}$$

whereas finding a linear filter with transfer function $T(\omega)$ to match the bispectrum is equivalent to solving

$$B(\omega_1, \omega_2) = T(\omega_1)\, T(\omega_2)\, T^*(\omega_1 + \omega_2). \tag{3.21b}$$

The solution to (3.21b) when it exists is generally different from that of (3.21a). Consider the process $Z(n)$ in Fig. 4. As can be seen, it is the sum of two processes, one which is the output of an AR filter driven by WGN $\epsilon(n)$ and the other the output of an AR filter driven by NGWN $W(n)$ of the type described in subsection 3. Further, $\epsilon(n)$ and $W(n)$ are independent which implies that $X(n)$ and $Y(n)$ are also statistically independent from which it follows that the bispectrum of $Z(n)$ is the sum of the individual bispectra. Since $X(n)$ is Gaussian its bispectrum is zero. So the bispectrum of $Z(n)$ is the bispectrum of $Y(n)$, i.e., it is given by (3.21b) with

$$T(\omega) = 1/A(\omega). \tag{3.22a}$$

where

$$A(\omega) = \left[1 + \sum_{i=1}^{p} a_i \exp(-j\omega i)\right].$$

The power spectrum of $Z(n)$ is the sum of the spectrum $X(n)$ and the spectrum of $Y(n)$, i.e., it is given by (3.21a) with

$$|H(\omega)|^2 = \frac{|A(\omega)|^2 + |B(\omega)|^2}{|A(\omega)|^2 \cdot |B(\omega)|^2} \tag{3.22b}$$

where

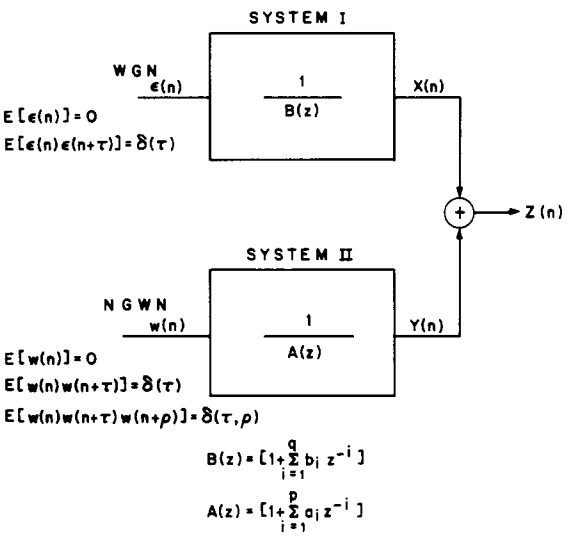


Fig. 4. The output process $Z(n)$ has ARMA power spectrum (3.22b) and AR bispectrum (3.22a). Input proccesses $\{\epsilon(n)\}$ and $\{W(n)\}$ are independent.

$$B(\omega) = \sum_{i=1}^{q} b_i \exp(-j\omega i).$$

This example clearly shows that using a model that describes the bispectrum of a process is generally different from one that describes its power spectrum. This example also shows that bispectral analysis has a possible role in identification of processes involving Gaussian and non-Gaussian components.

8) Gaussian Process into a Nonlinear Device [34], [37], [57], [59]: Let $\{X(k)\}$ be a Gaussian stationary process with zero-mean, autocorrelation $\{r_x(\tau)\}$, and power spectrum $P_x(\omega)$. Let

$$Y(k) = G[X(k)] \tag{3.23}$$

where $G[\cdot]$ is a nonlinear function. If $G[\cdot]$ is odd, $G[-k] = -G[k]$, the bispectrum of $\{Y(k)\}$ vanishes identically. Let us assume that $G[\cdot]$ contains an even component, for example, quadratic

$$Y(k) = X(k) + aX^2(k) \tag{3.24}$$

where a is small such that a^3 is negligible. Then the process $\{Y(k)\}$ has an autocorrelation sequence and a power spectrum [37], [59]

$$r_y(\tau) = r_x(\tau) + a^2 r^2(0) + 2a^2 r_x^2(\tau)$$

$$P_y(\omega) = P_x(\omega) + a^2 r^2(0)\,\delta(\omega) + 2a^2 \cdot \int P_x(\lambda)\, P_x(\omega - \lambda)\, d\lambda$$

respectively. Its bispectrum is given by

$$\begin{aligned} B_y(\omega_1, \omega_2) = {} & 2a\{P_x(\omega_1)\, P_x(\omega_2) \\ & + P_x(\omega_2)\, P_x(\omega_1 + \omega_2)\} \\ & + ar(0)\,\{P_x(\omega_1)\,\delta(\omega_2) \\ & + P_x(\omega_2)\,\delta(\omega_1) \\ & + P_x(\omega_2)\,\delta(\omega_1 + \omega_2)\} \\ & + 0(a^3). \end{aligned}$$

Let us note that when $a = 0$, $P_y(\omega) = P_x(\omega)$, and $B_y(\omega_1, \omega_2) = 0$. For the particular case of (3.24) of a quadratic system, there exist results concerning system identification by means of Gaussian input [11], [114]. For a general system identification case $G[\cdot]$ of (3.23) detailed results can be found in [5], [8], [13].

9) Poisson Triggered Process: Consider a process of the form [37]

$$X(k) = \sum_m h(k - T_m) \tag{3.25a}$$

where $\cdots, T_{-1}, T_0, T_1, T_2, \cdots$ are the times of events of a Poisson process, with $E\{T_{m+1} - T_m\} = \mu$. Assuming that $h(k)$ is related to $H(\omega)$ as shown in (3.13), the power spectrum and bispectrum of the process are

$$P_x(\omega) = \frac{1}{\mu}\,|H(\omega)|^2 \tag{3.25b}$$

$$B_x(\omega_1, \omega_2) = \frac{1}{\mu}\,H(\omega_1)\,H(\omega_2)\,H^*(\omega_1 + \omega_2). \tag{3.25c}$$

Let us note that $P_x(\omega) = B_x(\omega, 0)/H(0) = B_x(0, \omega)/H(0)$.

When the process $\{X(k)\}$ of (3.25a) acts as an additive noise (impulsive or shapes of spikes) [64] to a Gaussian random process $\{Y(k)\}$, i.e.,

$$Z(k) = Y(k) + X(k) \tag{3.26a}$$

where $\{Y(k)\}$, $\{X(k)\}$ are independent, then the power spectra and bispectra are

$$P_z(\omega) = P_y(\omega) + P_x(\omega) \tag{3.26b}$$

$$B_z(\omega_1, \omega_2) = B_x(\omega_1, \omega_2). \tag{3.26c}$$

$P_x(\omega)$ can be reconstructed from $B_x(\omega_1, \omega_2)$ and used in (3.26b) to reconstruct $P_y(\omega)$ from $P_z(\omega)$.

10) Bicoherence Index: This is a function that several authors have found useful in practical applications. It combines two completely different entities, namely, the bispectrum $B(\omega_1, \omega_2)$ and the power spectrum $P(\omega)$, of a process and is defined as [20], [37], [41], [42], [86]

$$b(\omega_1, \omega_2) \triangleq \frac{B(\omega_1, \omega_2)}{P(\omega_1)\,P(\omega_2)\,P(\omega_1 + \omega_2)}. \tag{3.27}$$

Let us note that for the processes whose power spectrum and bispectrum are modeled by the same linear filter, the magnitude bicoherence index is identical to a constant, i.e., $|b(\omega_1, \omega_2)| = C$ for all (ω_1, ω_2). The bicoherence index may also provide the degree of phase coupling in cases where harmonically related peaks are identified in the power spectrum domain [41], [42], [86], [87].

IV. Conventional Bispectrum Estimators

The problem met within practice is one of estimating the bispectrum of a process when a finite set of observation measurements is given. There are two chief approaches that have been used to estimate the bispectrum, namely, the conventional ("Fourier type") and the parametric approach which is based on autoregressive (AR), moving average (MA), and ARMA models. The conventional methods for bispectrum estimation and their properties is the subject of discussion in this section. They may be classified into the following two classes: 1) *Indirect class* of techniques which are approximations of the definition of the bispectrum given by (3.3) and (3.4), and 2) *Direct class* of techniques which approximate an equivalent definition of the bispectrum described by (3.5a) and (3.5b). While these approximations are straightforward, limitations on statistical variance of the estimates, computer time, and memory impose severe problems on their implementation. In fact, the computations may be surprisingly expensive despite the use of fast Fourier transform (FFT) algorithms [37].

A. Indirect Class of Conventional Bispectrum Estimators [8], [90], [98], [122]

Let $\{X(1), X(2), \cdots, X(N)\}$ be the given data set. Then we have the following.

1) Segment the data into K records of M samples each, i.e., $N = KM$.

2) Subtract the average value of each record.

3) Assuming that $\{x^{(i)}(k), k = 0, 1, \cdots, M-1\}$ is the data set per segment $i = 1, 2, \cdots, K$, obtain an estimate of the third-moment sequence

$$r^{(i)}(m, n) = \frac{1}{M} \sum_{\ell=s_1}^{s_2} x^{(i)}(\ell)\, x^{(i)}(\ell + m)\, x^{(i)}(\ell + n) \tag{4.1}$$

where

$$i = 1, 2, \cdots, K$$

$$s_1 = \max(0, -m, -n)$$

$$s_2 = \min(M - 1, M - 1 - m, M - 1 - n).$$

4) Average $r^{(i)}(m, n)$ over all segments

$$\hat{R}(m, n) = \frac{1}{K} \sum_{i=1}^{K} r^{(i)}(m, n). \tag{4.2}$$

5) Generate the bispectrum estimate

$$\hat{B}_{IN}(\omega_1, \omega_2) = \sum_{m=-L}^{L} \sum_{n=-L}^{L} \hat{R}(m, n)\, W(m, n) \cdot \exp\{-j(\omega_1 m + \omega_2 n)\} \tag{4.3}$$

where $L < M - 1$ and $W(m, n)$ is a two-dimensional window function. Let us note that the computations of the bispectrum estimate in (4.3) may be substantially reduced if the

symmetry properties of third moments (3.7) are taken into account for the calculations of $r^{(i)}(m, n)$ in (4.1) and if the symmetry properties of the bispectrum shown in (3.10) are incorporated in the computations of (4.3).

As in the case of conventional power spectrum estimation, to get better estimates, suitable windows should be used. Two-dimensional windows for bispectrum estimation have been derived and discussed by Sasaki, Sato, and Yamashita [98]. The window functions should satisfy the following constraints:

a) $W(m, n) = W(n, m) = W(-m, n-m) = W(m-n, -n)$ (symmetry properties of third moments);
b) $W(m, n) = 0$ outside the region of support of $\hat{R}(m, n)$;
c) $W(0, 0) = 1$ (normalizing condition);
d) $W(\omega_1, \omega_2) \geq 0$, for all (ω_1, ω_2). (4.4)

A class of functions which satisfies constraints (4.4), for $W(m, n)$, is the following [2], [98]:

$$W(m, n) = d(m)\, d(n)\, d(n-m) \tag{4.5}$$

where

$$d(m) = d(-m) \tag{4.6a}$$

$$d(m) = 0, \quad m > L \tag{4.6b}$$

$$d(0) = 1 \tag{4.6c}$$

$$D(\omega) \geq 0, \quad \text{for all } \omega. \tag{4.6d}$$

Equations (4.5) and (4.6) allow a reconstruction of two-dimensional window functions for bispectrum estimation using standard one-dimensional lag windows. However, not all conventional power spectrum windows satisfy constraint (4.6d). For example, the Hanning window has negative sidelobes in the frequency domain. Windows that satisfy (4.6a)–(4.6d) have been reported in [98]. These are as follows:

a) *Optimum window (minimum bispectrum bias supremum):*

$$d_0(m) = \begin{cases} \frac{1}{\pi}\left|\sin\frac{\pi m}{L}\right| + \left(1 - \frac{|m|}{L}\right)\left(\cos\frac{\pi m}{L}\right), & |m| \leq L \\ 0, & |m| > L. \end{cases} \tag{4.7}$$

b) *Parzen window:*

$$d_p(m) = \begin{cases} 1 - 6\left(\frac{|m|}{L}\right)^2 + 6\left(\frac{|m|}{L}\right)^3, & |m| \leq L/2 \\ 2\left(1 - \frac{|m|}{L}\right)^3, & L/2 \leq |m| \leq L \\ 0, & |m| > L. \end{cases} \tag{4.8}$$

c) *Uniform window in frequency domain:*

$$W_u(\omega_1, \omega_2) = \begin{cases} 4/3\left(\frac{\pi}{\Omega_0}\right), & |\omega| < \Omega_0 = a_0/L \\ 0, & |\omega| > \Omega_0 = a_0/L \end{cases} \tag{4.9}$$

where $|\omega| = \max[|\omega_1|, |\omega_2|, |\omega_1 + \omega_2|]$ and a_0 a constant parameter. This window is distributed uniformly in the hexagonal frequency region shown by Fig. 2(b), and it does not belong to the separable class described by (4.5). It rather belongs to the general class defined by (4.4).

The above three windows have been evaluated in terms of bispectrum bias spectrum (J) and approximate normalized (by triple product of power spectra) bispectrum variance (V) defined as

$$J \triangleq \frac{1}{(2\pi)^2}\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi}(\omega_1 - \omega_2)^2\, W(\omega_1, \omega_2)\, d\omega_1\, d\omega_2 \tag{4.10a}$$

and

$$V \triangleq \sum_{m=-L}^{L}\sum_{n=-L}^{L}|W(m, n)|^2 \tag{4.10b}$$

respectively. The results are illustrated in Table 2. From this table, it is apparent that under the condition $V_u = V_0$,

Table 2 Bispectrum Bias Supremum and Variance of Window Functions for Conventional Bispectrum Estimation

Windows	Bias Supremum (J)	Variance
Optimum	$J_0 = \frac{6\pi^2}{L^2}$	$V_0 \sim 0.05L^2$
Parzen	$J_p = \frac{72}{L^2}$	$V_p \sim 0.037L^2$
Uniform in Frequency Domain	$J_u = \frac{5}{6}\left(\frac{a_0}{L}\right)^2$	$V_u = \frac{4\pi^2}{3}\left(\frac{L}{a_0}\right)^2$

$J_u \cong 3.7J_0$. Comparing the optimum window with the Parzen window, we find that $J_p = 1.215J_0$ and $V_p = 0.74V_0$. In other words, the optimum window gives a variance which is about 26 percent larger than that of the Parzen window while the former achieves a bias which is about 18 percent smaller than that of the latter [98].

B. Direct Class of Conventional Bispectrum Estimators [9], [26], [28], [33], [37], [44], [66]

The definition of the bispectrum, as given by (3.5a) and (3.5b), provides a way to estimate it. If a finite segment of the process $X(k)$ has been observed, its Fourier transform coefficients will be approximations of the quantities $dZ(\omega)$ and the averaged triple products of the Fourier coefficients are then estimates of its bispectrum.

Let $\{X(1), X(2), \cdots, X(N)\}$ be the available set of observations for bispectrum estimation. Let us assume that f_s is the sampling frequency and $\Delta_0 = f_s/N_0$ is the required spacing between frequency samples in the bispectrum domain along horizontal or vertical directions. Thus N_0 is the total number of frequency samples.

a) Segment the data into K segments of M samples each, i.e., $N = KM$, and subtract the average value of each segment. If necessary, add zeros at each segment to obtain a convenient length M for the FFT.

b) Assuming that $\{X^{(i)}(k), k = 0, 1, 2, \cdots, M-1\}$ are the data of segment $\{i\}$, generate the DFT coefficients

$$Y^{(i)}(\lambda) = \frac{1}{M}\sum_{k=0}^{M-1}X^{(i)}(k)\exp(-j2\pi k\lambda/M), \quad \lambda = 0, 1, \cdots, M/2$$
$$i = 1, 2, \cdots, K. \tag{4.11}$$

c) In general, $M = M_1 \times N_0$, where M_1 is a positive integer (assumed odd number), i.e., $M_1 = 2L_1 + 1$. Since M is even and M_1 is odd, we compromise on the value of N_0 (closest

$$\hat{b}_i(\lambda_1, \lambda_2) = \frac{1}{\Delta_0^2} \sum_{k_1=-L_1}^{L_1} \sum_{k_2=-L_1}^{L_1} Y^{(i)}(\lambda_1 + k_1) \cdot Y^{(i)}(\lambda_2 + k_2) Y^{(i)*}(\lambda_1 + \lambda_2 + k_1 + k_2) \quad (4.12a)$$

over the triangular region $0 \le \lambda_2 \le \lambda_1, \lambda_1 + \lambda_2 \le f_s/2$ (Fig. 2(b)). For the special case where no averaging is performed in the bispectrum domain $M_1 = 1$, $L_1 = 0$ and therefore

$$\hat{b}_i(\lambda_1, \lambda_2) = \frac{1}{\Delta_0^2} Y^{(i)}(\lambda_1) \cdot Y^{(i)}(\lambda_2)\, Y^{(i)*}(\lambda_1 + \lambda_2). \quad (4.12b)$$

d) The bispectrum estimate of the given data is the average over the K pieces

$$\hat{B}_D(\omega_1, \omega_2) = \frac{1}{K} \sum_{i=1}^{K} \hat{b}_i(\omega_1, \omega_2) \quad (4.13)$$

where

$$\omega_1 = \left(\frac{2\pi f_s}{N_0}\right)\lambda_1 \quad \text{and} \quad \omega_2 = \left(\frac{2\pi f_s}{N_0}\right)\lambda_2.$$

An equivalent procedure for the direct estimation of the bispectrum was suggested by Godfrey [23] in terms of complex demodulates. The procedure is general in the sense that, once the complex demodulates have been computed, cumulant spectra of any order may easily be computed. Assuming that the DFT coefficients $\{Y^{(i)}(\lambda)\}$ have been computed in (4.11) the complex demodulation approach proceeds as follows [23], [37]:

a) Form the sequence

$$Y_k^{(i)}(\lambda) = \begin{cases} Y^{(i)}(\lambda + k), & \text{for } |k| < L_1 \\ 0, & \text{otherwise} \end{cases} \quad (4.14a)$$

and

$$\tilde{Y}_k^{(i)}(\lambda) = \begin{cases} Y^{(i)}(\lambda + k), & \text{for } |k| < 2L_1 \\ 0, & \text{otherwise.} \end{cases} \quad (4.14b)$$

b) Generate the complex demodulates by transforming back to the time domain

$$V_s^{(i)}(\lambda) = \sum_{k=-N'/2}^{N'/2} Y_k^{(i)}(\lambda) \exp(j2\pi sk/N') \quad (4.15a)$$

$$\tilde{V}_s^{(i)}(\lambda) = \sum_{k=-N'/2}^{N'/2} \tilde{Y}_k^{(i)}(\lambda) \exp(j2\pi sk/N') \quad (4.15b)$$

where $N' \ge 2L_1$.

c) Estimate the bispectrum per segment

$$\hat{b}_i(\omega_1, \omega_2) = \frac{1}{N'\Delta_0^2} \sum_{s=0}^{N'-1} V_s^{(i)}(\lambda_1) \cdot V_s^{(i)}(\lambda_2)\, \tilde{V}_s^{(i)*}(\lambda_1 + \lambda_2), \quad i = 1, 2, \cdots, K. \quad (4.16)$$

d) Average as in (4.13) to obtain the bispectrum estimate of the data.

It is easy to verify that (4.12a) and (4.16) are identical estimates.

C. Properties of Conventional Estimators

In general, the bispectrum estimates (4.3) and (4.13) are different. They become identical, however, if on the one hand (4.3) is computed for $L = M - 1$ without using a window and on the other hand (4.12b) is adopted for the estimation of (4.13). It has been shown that the conventional estimates are asymptotically unbiased and consistent [8], [90], [122]. They have distributions which tend to complex normal distributions [122]. For sufficiently large M and N, both direct and indirect estimators provide approximately unbiased estimates

$$E\{\hat{B}_{IN}(\omega_1, \omega_2)\} \cong E\{\hat{B}_D(\omega_1, \omega_2)\} \cong B(\omega_1, \omega_2) \quad (4.17)$$

with asymptotic variances [33], [37], [46], [98]

$$\text{var}\{R_e\hat{B}_{IN}(\omega_1, \omega_2)\} = \text{var}\{I_m\hat{B}_{IN}(\omega_1, \omega_2)\} \cong \frac{V}{(2L+1)^2K} P(\omega_1)\, P(\omega_2)\, P(\omega_1 + \omega_2) \quad (4.18a)$$

and

$$\text{var}\{R_e\hat{B}_D(\omega_1, \omega_2)\} = \text{var}\{I_m\hat{B}_D(\omega_1, \omega_2)\} \cong \frac{1}{KM_1} P(\omega_1)\, P(\omega_2)\, P(\omega_1 + \omega_2) \quad (4.18b)$$

where V is the energy of the window given by (4.10b) and $P(\omega)$ is the true power spectrum of the process. Let us note that when no window is used in (4.3), i.e., $V/(2L+1)^2 = 1$, and no averaging is performed by the direct approach in the frequency domain ($M_1 = 1$), (4.18a) becomes identical to (4.18b).

Conventional estimators are generally of high variance and therefore a large number of records (K) is required to obtain smooth bispectral estimates. From (4.18a) and (4.18b), it is apparent that the variance is reduced by increasing K or M_1. However, increasing the number of segments is demanding on computer time and may introduce potential nonstationarities [46]. Frequency-domain averaging over small rectangles, in addition to increasing computations, may increase bias and does not help near-bispectral peaks. In the case of "short" data records, the K could be increased by using overlapping records [46].

The conventional methods have the advantages of ease of implementation—FFT algorithms can be used—and good estimates with very long data records [86]. However, because of the "uncertainty principle" of the Fourier transform, the ability of the conventional methods to resolve harmonic components in the bispectrum domain is limited. This could pose a problem in detecting quadratic phase coupling at closely spaced frequency pairs [86], [87]. Finally, poor bispectral fidelity is achieved by conventional methods in the case of parametric processes [86], [87].

Additional references for the estimation of bispectra via conventional methods may be found in Lii and Helland [46], Hinich and Clay [33], Van Ness [122], Hasselman *et al.* [25], Raghuveer and Nikias [86], [87], Kim and Powers [41], [42], Akaike [2], etc.

V. Parametric Models for Bispectrum Estimation

One of the most popular and useful approaches for the interpretation of time series data is the construction of white noise driven linear parametric models from the underlying physical process. However, the time series literature is dominated by a Gaussian assumption and/or power spectrum matching (or modeling autocorrelation) techniques. Two limitations are associated with these techniques: 1) If the process is non-Gaussian, the power spectrum matching models will not capture all the information "hidden" in the data, and 2) non-Gaussian processes that are non-minimum-phase will be identified as being minimum-phase by autocorrelation-based models. There are three main motivations to look for a non-Gaussian white noise driven parametric model for bispectrum estimation: 1) to recover phase information accurately, 2) to increase the resolution capability of an estimator in resolving closely spaced peaks in bispectrum domain, and 3) to increase bispectrum fidelity in the situations where non-Gaussian processes are indeed parametric or may well be approximated by parametric models.

Consider a real autoregressive moving-average ARMA (p, q) process $\{X(n)\}$ described by

$$\sum_{i=0}^{p} a_i X(n-i) = \sum_{k=0}^{q} W(n-k), \qquad a_0 = 1 \tag{5.1}$$

where $\{W(n)\}$ are i.i.d. with $E\{W(n)\} = 0$, $E\{W(n)\,W(n+k)\} = Q \cdot \delta(k)$, $E\{W(n)\,W(n+k)\,W(n+m)\} = \beta \cdot \delta(k, m)$, and $X(m)$ is independent of $W(n)$ for $m < n$. Let us note that $\{W(n)\}$ and $\{X(n)\}$ are non-Gaussian. Since $\{W(n)\}$ is third-order-stationary it follows that $\{X(n)\}$ is also third-order-stationary assuming it is a stable ARMA model. The power spectrum of $\{X(n)\}$ is

$$P(\omega) = Q|H(z)|^2, \qquad \text{for } z = \exp(j\omega) \tag{5.2}$$

where

$$H(z) = N(z)/D(z)$$

$$N(z) = \sum_{k=0}^{q} b_k z^{-k}$$

$$D(z) = \sum_{i=0}^{p} a_i z^{-i}, \qquad a_0 = 1 \tag{5.3}$$

and its bispectrum is

$$B(\omega_1, \omega_2) = \beta H(\omega_1)\,H(\omega_2)\,H^*(\omega_1 + \omega_2) \tag{5.4}$$

where

$$H(\omega) = H(z)|_{z=\exp j\omega} = |H(\omega)|\exp j[h(\omega)]. \tag{5.5}$$

If $\{W(n)\}$ is Gaussian (and hence $\{X(n)\}$) then any real root (z_r) of $N(z)$ or $D(z)$ can be replaced by its inverse $(1/z_r)$ and pairs of nonzero conjugate roots (z_0) by their paired conjugated inverses $(1/z_0^*)$ without changing the power spectrum $P(\omega)$ in (5.2) and therefore the probability structure of $\{X(n)\}$ [47], [49], [50], [89], [91]. For example, with real distinct roots there are 2^{p+q} ways of choosing the roots without changing $P(\omega)$, i.e., resulting in the same autocorrelation sequence. This implies that there are 2^{p+q} ways of choosing the phase $h(\omega)$ in (5.5) without changing the magnitude $|H(\omega)|$.

A great body of results on the estimation of the coefficients $\{a_i\}$, $\{b_k\}$ of (5.1) exists based on least squares criteria which, in the Gaussian case, are essentially equivalent asymptotically to maximum-likelihood approaches. In the non-Gaussian case, these procedures provide least square solutions but not a maximum-likelihood solution. In both cases, however, the estimated coefficients correspond to roots inside the unit circle (minimum-phase ARMA model) [47], [49]. So, how do we correctly identify the ARMA (p, q) frequency transfer function? The following comments are in order:

1) If the process is Gaussian and $H(\omega)$ minimum-phase, autocorrelation-based methods (least squares) will correctly identify both magnitude $|H(\omega)|$ and phase $h(\omega)$ [27], [40], [58].

2) If $\{W(n)\}$ is Gaussian and $H(\omega)$ non-minimum-phase, *no procedure* can recover the true phase $h(\omega)$ [6].

3) If the process is non-Gaussian and $H(\omega)$ non-minimum-phase, autocorrelation-based methods will correctly identify $|H(\omega)|$ but not $h(\omega)$ [6], [47], [49].

4) If the process is non-Gaussian and $H(\omega)$ non-minimum-phase, true magnitude and phase can be recovered by *knowing* the actual non-Gaussian distribution of $\{W(n)\}$. This is achieved via maximum-likelihood estimate of $\{a_i\}$, $\{b_k\}$, but at the expense of solving nonlinear sets of equations [6].

5) If the process is non-Gaussian and $H(\omega)$ non-minimum-phase, $|H(\omega)|$ and $h(\omega)$ can correctly be recovered *without knowing* the actual non-Gaussian distribution of $\{W(n)\}$. This is achieved by estimating the parameters $\{a_i\}$, $\{b_k\}$ or directly magnitude $|H(\omega)|$ and phase $h(\omega)$ in a higher order spectrum domain such as the bispectrum or trispectrum.

Since higher order spectrum methods do not require the knowledge of the non-Gaussian distribution, they become more attractive in practical application problems of signal processing because the non-Gaussian distribution of the data is usually unknown or difficult to estimate (geophysics, multipath telecommunications, etc.). In this section we address the bispectrum estimation problem via MA, AR, and ARMA models and in the applications section we elaborate further on the phase estimation $h(\omega)$ for deconvolution. The problem herein can be stated as follows: Given only a finite length of data $\{X(1), X(2), \cdots, X(N)\}$, it is required to estimate the bispectrum of the underlying discrete random process via parametric modeling of its third moments.

A. MA Methods [47], [49], [89]

Assuming that $\{X(n)\}$ is truly an MA(q) process which can be obtained from (5.1) by setting $p = 0$, i.e.,

$$X(n) = \sum_{k=0}^{q} b_k W(n-k) \tag{5.6}$$

our problem is to estimate $\{b_k\}$ from a finite set of observations $\{X(1), \cdots, X(N)\}$ in such a way that the third-moment structure of the data is taken into account. Considering

$$R(\tau, \tau) = E\{X(n)\,X^2(n+\tau)\} \tag{5.7}$$

and substituting (5.6) into (5.7), we obtain

$$R(\tau, \tau) = \beta \sum_{k=0}^{q} b_k b_{k+\tau}^2, \qquad \tau = -q, \cdots, 0, \cdots, q. \tag{5.8}$$

Two methods for the estimation of MA coefficients have been suggested based on (5.8): namely, nonlinear least squares and the searching linear programming approach. Both methods employ estimates of third moments, $\hat{R}(\tau, \tau)$, which are generated following the procedure described by (4.1) and (4.2).

The nonlinear least squares approach solves the extremal problem

$$\text{minimize} \sum_{\tau=-q}^{+q} (\hat{R}(\tau, \tau) - \beta \sum_{k=0}^{q} b_k b_{k+\tau}^2)^2 \qquad (5.9)$$

with respect to $q + 2$ unknowns $\{b_k\}$ and β.

In the searching linear programming approach, one first uses a typical second-order method to estimate the coefficients of the MA process that will accurately reflect the autocorrelation structure of the data [47], [89]. Assuming that the resulting coefficients are $\{b_k^{(2)}\}$ then the roots, r_j, $j = 1, 2, \cdots, q$ of the polynomial

$$B(z) = \sum_{k=0}^{q} b_k^{(2)} z^{-k} \qquad (5.10)$$

have magnitude less than one (minimum phase). An accurate estimate of the distribution of roots could be obtained by taking the conjugated inverse of an appropriate number of r_j's. As pointed out earlier, there are 2^q possible sets of roots that yield the same autocorrelation sequence and 2^q distinct sets of third-moment sequences. Therefore, the method is seeking through linear programming the set of roots and thus coefficients $\{b_k\}$ that minimize (5.9) among all possible sets of roots. Let us note that if the roots are complex, the inverse complex conjugates are taken in pairs [47], [89].

Both methods described above generate the MA coefficients by utilizing a subset of third-moment estimates; namely, $\hat{R}(\tau, \tau), \tau = -q, \cdots, q$. However, a larger set of third moments could be included in the estimation procedures by modifying (5.8) to

$$R(\tau, \rho) = \beta \sum_{k=0}^{q} b_k b_{k+\tau} b_{k+\rho}, \qquad \tau, \rho = -q, \cdots, q \qquad (5.11a)$$

and rewriting (5.9) as

$$\sum_{\tau=-q}^{+q} \sum_{\rho=-q}^{+q} (\hat{R}(\tau, \rho) - R(\tau, \rho))^2. \qquad (5.11b)$$

It is important to note that both methods are based on a least squares third-moment matching criterion, (5.9) or (5.11b). While the first method is a general one and could be used on processes with both power spectrum and bispectrum described by two different linear filters, the second approach (searching linear programming) can only be used on processes of which the power spectrum and bispectrum can be described by the same linear filter, say MA(q).

B. AR Methods [74], [75], [84]–[87]

Assuming now that $\{X(n)\}$ is truly a pth-order AR(p) process obtained from (5.1) by setting $q = 0$, $a_0 = b_0 = 1$, i.e.,

$$X(n) + \sum_{i=1}^{p} a_i X(n - i) = W(n) \qquad (5.12)$$

the third-moment sequence $R(\tau, \rho) = E\{X(n)\, X(n + \tau)\, X(n + \rho)\}$ satisfies the following *third-order* recursion [86]:

$$R(-k, -\ell) + \sum_{i=1}^{p} a_i R(i - k, i - \ell) = \beta \cdot \delta(k, \ell), \qquad (k, \ell) \geq 0 \qquad (5.13)$$

where $\delta(k, \ell)$ is the two-dimensional unit impulse function. From (5.13), it follows that $2p + 1$ third-moment values on the $\tau = p$ line satisfy the matrix equation [86], [87].

$$\boldsymbol{R}_c\, \boldsymbol{a} = \boldsymbol{\beta} \qquad (5.14)$$

where

$$\boldsymbol{R}_c = \begin{bmatrix} R(0, 0) & R(1, 1) & \cdots & R(p, p) \\ R(-1, -1) & R(0, 0) & \cdots & R(p - 1, p - 1) \\ \vdots & & & \\ R(-p, -p) & R(-p + 1, -p + 1) & \cdots & R(0, 0) \end{bmatrix}$$

$$\boldsymbol{a} = [1, a_1, a_2, \cdots, a_p]^T$$

$$\boldsymbol{\beta} = [\beta, 0, 0, \cdots, 0]^T.$$

The matrix $\boldsymbol{R}_c$ is Toeplitz but in general is not symmetric. Needless to say that $\boldsymbol{R}_c$ will be Hankel if the parameter vector $\boldsymbol{a}$ is written as $\boldsymbol{a} = [a_p, \cdots, a_i, 1]$.

Another representation of (5.13) is possible by letting (k, ℓ) run on portions of the triangle in Fig. 2(a). The $p + 1$ equations corresponding to this representation are

$$R(-k, -\ell) + \sum_{i=1}^{p} a_i R(i - k, i - \ell) = \beta \cdot \delta(k, \ell)$$

$$k = 0, 1, \cdots, L_1$$

$$\ell = \begin{cases} 0, \cdots, k, & \text{for } k < L_1 \\ 0, \cdots, L_2, & \text{for } k = L_1 \end{cases} \qquad (5.15)$$

where L_1 and L_2 are chosen such that $L_2 \leq L_1$, and

$$p = 1 + L_2 + \frac{(L_1 - 1)(L_1 + 2)}{2}.$$

The matrix corresponding to the equations in (5.15) does not possess the Toeplitz structure. However, this approach draws information from samples over a finite area rather than just a portion of a straight line and therefore is more general than the case in (5.14).

Given $2p + 1$ samples of the true third-moment sequence of a process, one at the origin and p points on either side of it on the $m = n$ line in Fig. 2(a), (5.14) can be used to fit a pth AR model. This model is fitted in the sense of perfect matching between the third-moment sequence of the output of the AR model and the given samples at correspond-

ing lags, i.e., if we were to insist on finding an NGWN-driven pth-order AR model whose output third-moment sequence equaled the given samples at lags $m = n = 0, \pm 1, \cdots, \pm p$, then the parameters of such a model would satisfy (5.14) as a necessary condition. More than $2p + 1$ samples are required to fit such a model using (5.15). If the samples are from the third moments of a true pth-order process satisfying all the model assumptions then the parameters $a_1, \cdots, a_p$ in both cases are the same. Otherwise, the two solutions may be different.

The problem of matching a given bispectrum by that of an AR process can also be viewed from a different angle. Let $B_x(\omega_1, \omega_2)$ be the bispectrum of a certain stationary process $X(n)$. Let us form a linear predictor of $X(n)$ that makes use of the preceding p values as below

$$\hat{X}(n) = -\sum_{i=1}^{p} a_i X(n - i). \tag{5.16}$$

The prediction error $e(n)$ is given by

$$e(n) = X(n) - \hat{X}(n) = X(n) + \sum_{i=1}^{p} a_i X(n - i). \tag{5.17}$$

The bispectrum $E(\omega_1, \omega_2)$ of $e(n)$ can be easily shown to be given by

$$E(\omega_1, \omega_2) = A(\omega_1)\, A(\omega_2)\, A^*(\omega_1 + \omega_2)\, B(\omega_1, \omega_2) \tag{5.18}$$

where

$$A(\omega) = \left[1 + \sum_{i=1}^{p} a_i \exp - j\omega i\right].$$

If the prediction parameters in (5.16) are such that

$$E(\omega_1, \omega_2) = \beta \tag{5.19}$$

where β is a real constant, then we would have

$$B(\omega_1, \omega_2) = \frac{\beta}{A(\omega_1)\, A(\omega_2)\, A^*(\omega_1 + \omega_2)} \tag{5.20}$$

i.e., the bispectrum of $X(n)$ is given exactly by that of an AR process with parameters $\{a_i, i = 1, \cdots, p\}$ with a driving NGWN of skewness. Thus the closeness of the error bispectrum $E(\omega_1, \omega_2)$ to "flatness" gives a measure of the closeness of the corresponding AR bispectrum to the bispectrum of the process being considered [87]. A detailed discussion on the concept of higher order "flatness" or "whiteness" may be found in [9], [10].

We now present two methods for estimating the bispectrum of a process when we are given a single finite realization of the process. The methods differ in the way they form estimates of the third-order equations in (5.13).

The Third-Order Recursion (TOR) Method [86]: Given the data samples $\{X(1), \cdots, X(n)\}$, third moments $\hat{R}(\tau, \rho)$ are estimated using (4.1) and (4.2) and used in place of the true ones in (5.14) to obtain

$$\hat{\boldsymbol{R}}_c \hat{\boldsymbol{a}} = \hat{\boldsymbol{\beta}} \tag{5.21}$$

where

$$\hat{\boldsymbol{R}}_c = \begin{bmatrix} \hat{R}(0, 0) & \hat{R}(1, 1) & \cdots & \hat{R}(p, p) \\ \hat{R}(-1, -1) & \hat{R}(0, 0) & \cdots & \hat{R}(p - 1, p - 1) \\ \vdots & & & \\ \hat{R}(-p, -p) & \hat{R}(-p + 1, -p + 1) & \cdots & \hat{R}(0, 0) \end{bmatrix}$$

$\hat{\boldsymbol{a}} \triangleq [1, \hat{a}_1, \cdots, \hat{a}_p]^T$ are the estimates of the AR parameters and $\hat{\boldsymbol{\beta}} = [\hat{\beta}, 0, \cdots, 0]^T$ is the estimate of the third moment of the driving white noise. Instead of using (5.14), we can use (5.15) to form the AR parameter estimates. An assumption made in using these procedures is that the process whose samples are given is third-order ergodic. It is shown in [86] that if the process is in fact of the type in (5.12), then the TOR method provides consistent estimates of the AR parameters.

The Constrained Third-Order Mean (CTOM) Method [87]: Let $\{X(1), \cdots, X(N)\}$ be samples of a stationary process whose bispectrum is to be estimated. Consider

$$\hat{q}_m(k, i) = X(m - i)\, X^2(m - k), \qquad i, k = 1, \cdots, p. \tag{5.22}$$

We see that

$$E\{\hat{q}_m(k, i)\} = R(i - k, i - k) \tag{5.23}$$

where $R(\tau, p)$ denotes the third-moment sequence of the process $X(n)$ whose samples are given above.

If we were given samples of $R(\tau, p)$ itself then to fit a pth-order AR model, we would be solving equations corresponding to (5.13), i.e., we would have

$$E\{\hat{q}_m(k, 0) + \sum_{i=1}^{p} \hat{a}_i \hat{q}_m(k, i)\} = 0, \qquad k = 1, \cdots, p. \tag{5.24a}$$

We denote the quantity within the braces by $\hat{C}(m, k)$. Thus with true third-moment samples we would solve

$$E\{\hat{C}(m, k)\} = 0, \qquad k = 1, \cdots, p. \tag{5.24b}$$

We refer to $\hat{C}(m, k)$ as the third-order (TOR) error process. With the given data samples we can obtain $N - p$ samples of $\hat{C}(m, k)$ for every value of k, i.e., we can have $\hat{C}(m, k)$ for $m = p + 1, p + 2, \cdots, N$; $k = 1, 2, \cdots, p$.

Instead of taking expectations as in (5.24b) we now equate the sample mean of the $\hat{C}(m, k)$ to zero to get the p linear equations to be solved for the parameters

$$\frac{1}{N - p} \sum_{m=p+1}^{N} \hat{C}(m, k) = 0, \qquad k = 1, \cdots, p. \tag{5.25}$$

In matrix form we have

$$\hat{\boldsymbol{Q}}\hat{\boldsymbol{a}} = \hat{\boldsymbol{q}} \tag{5.26}$$

where

$$\hat{\boldsymbol{Q}} = \begin{bmatrix} \hat{q}_{11} & \cdots & \hat{q}_{1p} \\ \vdots & & \\ \hat{q}_{p1} & \cdots & \hat{q}_{pp} \end{bmatrix}$$

$$\hat{\boldsymbol{a}} = [\hat{a}_1 \cdots \hat{a}_p]^T$$

$$\hat{\boldsymbol{q}} = [\hat{q}_{10} \cdots \hat{q}_{p0}]^T, \qquad \hat{q}_{ij} = \sum_{m=p+1}^{N} \hat{q}_m(i, j).$$

For simplicity of notation we considered an unbroken segment of N samples of data in outlining the CTOM approach. However, we may choose to divide the given data into records, form $\hat{C}(m, k)$ of (5.24a) for each segment, average $\hat{C}(m, k)$'s over all segments, and finally equate the sample mean to zero as in (5.25) to get the CTOM equations. It is shown in [87] that asymptotically, the CTOM method and the TOR method are equivalent for a given model order. The proposed methods are only AR bispectrum estimation methods. Since no entropy considerations are involved, they are not "maximum-entropy" bispectrum estimation methods [71].

It should be emphasized that the appropriate AR model for obtaining the bispectrum estimate of the data $\{X(1), \cdots, X(N)\}$ will generally be different from that for AR power spectrum estimation of the same data. Some of the well-known AR model order selection criteria such as the FPE, AIC, and CAT [39] depend on autocorrelations and hence cannot be used. However, AR model order selection criteria based on matrix rank determination may be adapted to the third-order case [86].

The TOR and CTOM methods were based on a causal AR model described by (5.12). However, we can also consider anti-causal and noncausal representations. The third-order recursion equations for these will be different from what we have in (5.13). For example, the anti-causal AR(p) model

$$X(n) + \sum_{i=1}^{p} b_i X(n + i) = W(n) \tag{5.27a}$$

has third-order recursion equation

$$R(k, \ell) + \sum_{i=1}^{p} b_i R(k - i, \ell - i) = \beta \cdot \delta(k, \ell), \quad \text{for } (k, \ell) \geq 0. \tag{5.27b}$$

It therefore follows that $2p + 1$ third moments on the $\tau = \rho$ line would satisfy the matrix equation

$$\boldsymbol{R}_a \boldsymbol{b} = \boldsymbol{\beta} \tag{5.27c}$$

where

$$\boldsymbol{R}_a = \begin{bmatrix} R(0,0) & R(-1,-1) & R(-2,-2) & \cdots & R(-p,-p) \\ R(1,1) & R(0,0) & R(-1,-1) & \cdots & R(-p+1,-p+1) \\ \vdots & & & & \\ R(p,p) & R(p-1,p-1) & R(p-2,p-2) & \cdots & R(0,0) \end{bmatrix}$$

$$\boldsymbol{b} = [1, b_1, \cdots, b_p]^T$$

$$\boldsymbol{\beta} = [\beta_2, 0, \cdots, 0]^T.$$

Comparing (5.14) with (5.27c) we see that $\boldsymbol{R}_a = \boldsymbol{R}_c^T$ wherer "T" denotes the matrix transpose operation. Thus $\boldsymbol{b}$ is different from $\boldsymbol{a}$.

Let us note that causal AR models can only provide minimum-phase information accurately whereas anti-causal AR models maximum-phase information [74], [75]. If the two models are properly utilized on the same data $\{X(1), \cdots, X(N)\}$, then the true mixed phase of the signal can be obtained [74], [75]. This is discussed later in the paper. Non-minimum-phase information could also be captured via noncausal AR model formulations based on TOR or CTOM parameter estimates [68], [69]. Extension of the TOR formulations to the multichannel case was provided in [83].

Work has been done on non-Gaussian AR processes to discriminate between a process with all its poles inside the unit circle and a process having the same power spectrum (or autocorrelation) but with at least one pole outside the unit circle. Huzii developed a method based on a higher order moment to achieve such discrimination [38]. However, no bispectral computations were involved. For example, in the case of an AR(1) process $\{1, a_1\}$, the proposed test statistic is [38]

$$\pi_k = R(0, k)/R(k, k) \tag{5.28a}$$

where k is a positive integer. The detection scheme is

$$\pi_\rho \underset{H_0}{\overset{H_1}{\gtrless}} 1 \tag{5.28b}$$

with hypothesis H_1: $|a_1| < 1$ and hypothesis H_0: $|a_1| > 1$.

C. ARMA Methods [45], [47], [49], [67], [89], [91]

One of the currently available methods for the estimation of the ARMA(p, q) parameters $\{a_i\}, \{b_i\}$ in (5.1) from a finite set of observations $\{X(1), \cdots, X(N)\}$ is a four-step procedure. First, the power spectrum of the data, $P(\omega)$, is estimated. Second, the phase $h(\omega)$ of the transfer function (5.5) is reconstructed from a conventional bispectrum estimate of the data and third, the ARMA(p, q) impulse response is generated using inverse FFT algorithm. Finally, the model parameters are obtained from the impulse response via Pade rational approximants [45]. The method, therefore, is based on conventional bispectrum estimation techniques and also utilizes information that is present in the autocorrelation domain.

From (5.2) and (5.5), it follows that the frequency transfer function is given by

$$H(\omega) = [P(\omega)/Q]^{1/2} \exp jh(\omega). \tag{5.29}$$

There is a big body of literature concerning the estimation of $P(\omega)$ and Q from the available observations $\{X(1), \cdots, X(N)\}$ [15], [27], [39], [40], [47]. On the other hand, Brillinger [12], Lii and Rosenblatt [47], and Matsuoka and Ulrych [60] have proposed methods to reconstruct the phase information $h(\omega)$ from the bispectrum $B(\omega_1, \omega_2)$. These methods are explained later in the Applications section of the paper. So, having computed $P(\omega)$, Q, and $h(\omega)$ and assuming that the ARMA (p, q) model is stable and causal, its impulse response, $\{c_\lambda\}$, is obtained from [45]

$$c_\lambda = \frac{1}{2\pi} \int_{-\pi}^{+\pi} [P(\omega)/Q]^{1/2} \exp jh(\omega) \exp j\omega\lambda \, d\omega. \tag{5.30}$$

The ARMA parameters are computed utilizing the identity

$$\frac{N(Z)}{D(Z)} = C(Z) = \sum_{\lambda=0}^{\infty} c_\lambda Z^{-\lambda} \tag{5.31}$$

where $N(Z)$, $D(Z)$ are given by (5.3). It is assumed that $N(Z)$ and $D(Z)$ have no common factors. Specifically, let

$$\hat{C}(Z) = \sum_{\lambda=0}^{L} \hat{c}_\lambda Z^{-\lambda}$$

be the estimated coefficients, where L is sufficiently large and $L > p + q$. Assuming $p > q$, then our problem is

$$\hat{N}(Z)/\hat{D}(Z) = \hat{C}(Z) \tag{5.31a}$$

or

$$\begin{aligned}\{\cdots, 0, \hat{b}_0, \hat{b}_1, \cdots, \hat{b}_q, 0, \cdots\} \\ = \{\cdots, 0, \hat{a}_0, \hat{a}_1, \cdots, \hat{a}_p, 0, \cdots\} \\ * \{\cdots, 0, \hat{c}_0, \hat{c}_1, \cdots, \hat{c}_L, 0, \cdots\}\end{aligned} \tag{5.31b}$$

where "*" denotes linear convolution. Since $\hat{b}_n = 0$ for $n \geq q + 1$ and $\hat{a}_0 = 1$, it follows that

$$\begin{bmatrix} \hat{c}_{q+r} & \hat{c}_{q+r-1} & \cdots & \hat{c}_{q+r-p+1} \\ \hat{c}_{q+r+1} & \hat{c}_{q+r} & \cdots & \hat{c}_{q+r-p+2} \\ \vdots & \vdots & & \vdots \\ \hat{c}_{q+r+p-1} & \hat{c}_{q+r+p-2} & & \hat{c}_{q+r} \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = - \begin{bmatrix} \hat{c}_{q+r+1} \\ \hat{c}_{q+r+2} \\ \vdots \\ \hat{c}_{q+r+p} \end{bmatrix} \tag{5.31c}$$

for any $r > (p - q)$. Similarly, for $n = 0, 1, 2, \cdots, q$, we have

$$\sum_{i=0}^{p} \hat{a}_i \hat{c}_{n-i} = \hat{b}_n$$

or in a matrix form

$$\begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_q \end{bmatrix} = \begin{bmatrix} \hat{c}_0 & 0 & 0 & \cdots & 0 \\ \hat{c}_1 & \hat{c}_0 & 0 & \cdots & 0 \\ \vdots & & & & \\ \hat{c}_q & \hat{c}_{q-1} & \hat{c}_{q-2} & \cdots & \hat{c}_0 \end{bmatrix} \begin{bmatrix} 1 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_q \end{bmatrix}. \tag{5.31d}$$

Obviously, the method requires *a priori* knowledge of the polynomial orders of p, q. These can be estimated by constructing the c-table associated with $C(z)$ and identifying its breaking point [45]. In other words, we use the $\{\hat{c}_\lambda\}$ to construct estimates

$$\hat{C}_{t,s} = (-1)\frac{s(s-1)}{2} \det [\hat{c}_{t+i-j}, \quad i, j = 1, 2, \cdots, s] \tag{5.32}$$

which, ideally, satisfy the following conditions: i) $c_{q,p} \neq 0$ and ii) $c_{q+1,p+1} = 0$ ("breaking point"). In practice, of course, $\hat{c}_{q+1,p+1} \cong 0$.

Extension of this method to trispectrum estimation can found in [49], [50], [91], [95]. Estimation of the ARMA (p, q) parameters based on matching criteria that employ autocorrelations and fourth-order cumulants were proposed in [116]–[118]. An ARMA method which estimates a minimum-phase AR part from autocorrelations and non-minimum-phase all-pass MA part from higher order cumulants was proposed in [21], [22]. ARMA trispectrum methods based on causal and anti-causal models were introduced in [68].

VI. Applications of the Bispectrum

It is the intent of this section to discuss in detail three specific applications of the bispectrum in signal processing problems, namely, detection and quantification of quadratic phase coupling, phase estimation of non-Gaussian parametric signals for deconvolution, and time-delay estimation from two-sensor measurements.

A. Quadratic Phase Coupling

Quadratic phase coupling can arise only among harmonically related components. Three frequencies are harmonically related when one of them is the sum of the other two. A special case is when we have two components with one being at twice the frequency of the other. In certain applications, such as EEG data analysis [37], oceanography [25], and plasma physics [41], [42], it is necessary to find out if peaks at harmonically related positions in the power spectrum are, in fact, phase-coupled. Since the power spectrum suppresses all phase relations, it cannot provide the answer.

One might ask how can techniques such as the conventional and parametric bispectrum estimation methods that employ linear filters (FIR or IIR) be used to detect something that might have resulted from quadratic nonlinearities. To answer this we must explain how the bispectrum itself is capable of detecting phase coupling. This is best illustrated by an example.

Consider the process [86], [87]

$$X(n) = \sum_{i=1}^{6} \cos(\lambda_i n + \varphi_i) \tag{6.1}$$

where $\lambda_1 > \lambda_2 > 0$, $\lambda_4 > \lambda_5 > 0$, $\lambda_3 = \lambda_1 + \lambda_2$, $\lambda_6 = \lambda_4 + \lambda_5$, $\varphi_1, \varphi_2, \cdots, \varphi_5$ are all independent, uniformly distributed r.v.s. over $(0, 2\pi)$, and $\varphi_6 = \varphi_4 + \varphi_5$. In (6.1) while $(\lambda_1, \lambda_2, \lambda_3)$ and $(\lambda_4, \lambda_5, \lambda_6)$ are at harmonically related positions, only the component at λ_6 is a result of phase coupling between those at λ_4 and λ_5 while the one at λ_3 is an independent harmonic component. The power spectrum of the process consists of impulses at λ_i; $i = 1, 2, \cdots, 6$ as illustrated in Fig. 5(a). Looking at the spectrum one cannot say if the harmonically related components are, in fact, involved in quadratic phase-coupling relationships. The third-moment sequence $R(k, \ell)$ of $X(n)$ can be easily obtained as [86]

$$\begin{aligned} R(k, \ell) = \tfrac{1}{4}\{&\cos(\lambda_5 k + \lambda_4 \ell) + \cos(\lambda_6 k + \lambda_4 \ell) \\ &+ \cos(\lambda_4 k + \lambda_5 \ell) + \cos(\lambda_6 k - \lambda_5 \ell) \\ &+ \cos(\lambda_4 k - \lambda_6 \ell) + \cos(\lambda_5 k - \lambda_6 \ell)\}. \end{aligned} \tag{6.2}$$

It is important to observe that in (6.2) *only the phase coupled components appear.* Consequently, the bispectrum evaluated in the triangular region of Fig. 2(a) shows an impulse only at (λ_4, λ_5) indicating that only this pair is phase-coupled (see also Fig. 5(b)). In the total absence of phase coupling, the third-moment sequence and hence the bispectrum are both zero. Thus the fact that *only phase coupled components contribute to the third-moment sequence* of a process is what makes the bispectrum a useful tool for

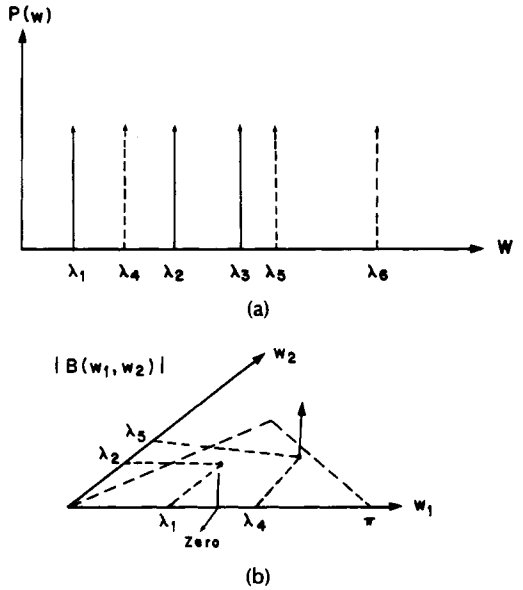


Fig. 5. Quadratic phase coupling. (a) Power spectrum of the process described by (6.1). (b) Its magnitude bispectrum.

detecting quadratic phase coupling and discriminating phase-coupling components from those that are not.

Any of the bispectrum estimation techniques described in Section V can be used for the analysis of quadratic phase-coupling phenomena. However, each of those techniques will exhibit certain advantages and a number of limitations. For example, the conventional techniques can serve as better quantifiers of phase coupling whereas the parametric methods (AR and ARMA) are better as detectors rather than quantifiers [86], [87].

The use of conventional methods for bispectrum estimation in conjunction with the bicoherence index $b(\omega_1, \omega_2)$ defined by (3.27) has been used extensively for the detection and quantification of quadratic phase coupling in [41], [42], [86]. When the bicoherence index takes on a value close to unity at a frequency pair where phase coupling has occurred this indicates an almost 100-percent degree of phase coherence. On the other hand, a near-zero value of $b(\omega_1, \omega_2)$ at harmonically related frequency pairs will suggest an absence of phase coupling (an almost 0.0-percent degree of phase coupling) [41], [42]. Certainly, one of the advantages of using the conventional approach to bicoherence index estimation is its ability to serve as a good quantifier by providing good estimates of the degree of phase coupling at harmonically related frequency pairs. For example, the conventional bicoherence estimate of the process

$$X(n) = \sum_{i=1}^{4} \cos(\lambda_i n + \varphi_i) \tag{6.3}$$

where $\lambda_3 = \lambda_1 + \lambda_2$, $\varphi_3 = \varphi_1 + \varphi_2$, $\lambda_4 = \lambda_1 + \lambda_2$, and φ_1, φ_2, φ_4 are i.i.d, uniformly distributed r.v.s over $[0, 2\pi]$, will show a peak at (λ_1, λ_2) of magnitude approximately 0.5 indicating a 50-percent degree of phase coherence only. Possible limitations for using conventional methods for the analysis of quadratic phase coupling are the high variance of bispectrum estimates and the resolution when harmonically related frequency pairs are close to each other [86], [87]. Let us note that the effect of these limitations is reduced as the number and/or length of the data segments increase.

Quadratic phase coupling has also been studied in [84]–[87], using AR models for bispectrum estimation. The motivation to use AR techniques was to take advantage of the high-resolution capability and low variance estimates associated with AR modeling. The justification for the use of AR methods for detection of quadratic phase coupling was provided in [86] by proving that the third-moment sequence of a process consisting of a sum of cosinusoids of which N pairs are coupled can be modeled exactly by an AR($6N$) model. For example, the $R(k, \ell)$ described by (6.2) would satisfy exactly the third-order recursion (5.13) if the AR parameters take the values

$$\begin{aligned} a_1 &= -2(\cos\lambda_4 + \cos\lambda_5 + \cos\lambda_6) \\ a_2 &= 3 + 4(\cos\lambda_4 \cos\lambda_5 + \cos\lambda_5 \cos\lambda_6 \\ &\quad + \cos\lambda_6 \cos\lambda_4) \\ a_3 &= -4(\cos\lambda_4 + \cos\lambda_5 + \cos\lambda_6 \\ &\quad + 2\cos\lambda_4 \cos\lambda_5 \cos\lambda_6) \\ a_4 &= a_2 \\ a_5 &= a_1, \end{aligned}$$

and

$$a_6 = 1. \tag{6.4}$$

It was also shown in [86] that we cannot have just a single segment of data (i.e., just one set of fixed values for φ_1, φ_2, φ_3, φ_4 in (6.3)) for detecting quadratic phase coupling between pairs of sinusoids. Segmentation of data into records is necessary to obtain consistent estimates of the third moments. The advantages of AR techniques arise in situations where the conventional estimators completely fail to resolve closely spaced frequency pairs [86], [87]. On the other hand, the limitation of the AR techniques lies on their inability to provide an accurate estimate of the degree of phase coherence when phase coupling does occur at harmonically related frequency pairs. Simulation examples that demonstrate the performance of AR and conventional methods can be found in [84]–[87].

B. Deconvolution

Assuming that $\{X(n)\}$ is the observed output of a linear, time-invariant system with input $\{W(n)\}$ being a non-Gaussian zero-mean white noise process, i.i.d, the following discrete-time convolution relationship may be adopted

$$X(n) = \sum_{k=0}^{L} h_k W(n-k) \tag{6.5}$$

where $\{h_k\}$ represents the system (or wavelet) parameters. The deconvolution problem is to recover $\{h_k\}$ first (or, equivalently, the magnitude and phase of the system transfer function) and then to process $\{X(n)\}$ through the inverse system in order to generate the reflectivity series $\{W(n)\}$. This problem arises in data communication when a receiver is obliged to achieve a blind starting phase, i.e., without the transmission by the transmitter of a known sequence of data [6], [81]. It also arises in geophysical and other deconvolution scenarios [24], [36], [60], [104], [109].

For classical deconvolution problems, an accurate phase reconstruction in the autocorrelation (or power-spectrum) domain can only be achieved if the system is indeed minimum-phase, i.e., the zeros of the polynomial

$$A(z) = \sum_{k=0}^{L} h_k z^{-k} \tag{6.6}$$

be inside the unit circle $|z| = 1$. As we mentioned earlier, if the driving white noise is Gaussian, no procedure can recover the actual shape of a non-minimum-phase system [6].

Various approaches have been suggested for non-minimum-phase wavelet estimation based on different principles and assumptions. The first-order system case was solved in [104] using the absolute value norm (L_1). The general non-minimum-phase problem was addressed in [6] assuming exactly known non-Gaussian distribution and solving a highly nonlinear system of equations. Other approaches include predictive deconvolution, homomorphic, minimum entropy, and maximum-likelihood deconvolution (see [109] and references therein). Each one of these techniques has a number of advantages and limitations depending on the data environment. Perhaps, the most popular deconvolution scheme is the predictive one which makes the assumption that the wavelet is minimum-phase. However, this assumption is often found to be incorrect.

Another approach to the wavelet estimation problem that has recently been suggested explores the use of higher order spectra [47]–[51], [78], [84]–[95]. Assuming that $\{W(n)\}$ in (6.5) has nonzero skewness, i.e., $E\{W^3(n)\} = \beta \neq 0$, then the bispectrum of $\{X(n)\}$ contains the non-minimum-phase information of $\{a_i\}$. It is given by

$$B_x(\omega_1, \omega_2) = |B_x(\omega_1, \omega_2)| \exp j\, \Psi_x(\omega_1, \omega_2) \tag{6.7}$$

where

$$B_x(\omega_1, \omega_2) = |A(\omega_1)|\, |A(\omega_2)|\, |A(\omega_1 + \omega_2)| \tag{6.8}$$

$$\Psi_x(\omega_1, \omega_2) = \varphi_A(\omega_1) + \varphi_A(\omega_2) - \varphi_A(\omega_1 + \omega_2) \tag{6.9}$$

and

$$A(\omega) = A(z)\,|_{z=\exp j\omega} = |A(\omega)| \exp j\, \varphi_A(\omega). \tag{6.10}$$

Methods have been suggested for phase estimation in the bispectrum domain by Brillinger [11], Lii–Rosenblatt [45], [47], and Matsuoka–Ulrych [60] by generating bispectrum estimates $B_x(\omega_1, \omega_2)$ via conventional methods. This means that once a conventional estimate $\hat{\Psi}_x(\omega_1, \omega_2)$ is obtained, the $\hat{\varphi}_A(\omega)$ is reconstructed using (6.9) in some manner. The Brillinger (BR) and Lii–Rosenblatt (LR) algorithms reconstruct $\hat{\varphi}_A(\omega)$ from $\hat{\Psi}_x(\omega_1, \omega_2)$ recursively whereas the Matsuoka–Ulrych (MU) method does so in a nonrecursive least squares manner. The magnitude $|A(\omega)|$ is estimated in the autocorrelation (or power spectrum) domain. Another approach to the deconvolution problem in the bispectrum domain was suggested in [67], [74], [75] by employing causal and anticausal ARMA models and reconstructing both $|A(\omega)|$ and $\varphi_A(\omega)$ in the bispectrum domain. Finally, non-minimum-phase signal reconstruction via cepstrum modeling of third-order moments was established in [70]. The details of the algorithms are summarized below.

1) The BR Algorithm: Brillinger originally suggested the recursive equation [11]

$$\varphi_A(\omega) = \left\{2 \int_0^{\omega} \varphi_A(\lambda)\, d\lambda - \int_0^{\omega} \Psi_x(\lambda, \omega - \lambda)\, d\lambda\right\} \Big/ \omega \tag{6.11}$$

which was modified in [60] for digital computations resulting in the expression

$$\varphi_A(n) = \left\{2 \sum_{i=0}^{n-1} \varphi_A(i) - S(n)\right\} \Big/ (n-1), \qquad n = 2, 3, \cdots, N \tag{6.12a}$$

where

$$S(n) = \sum_{i=0}^{n} \Psi_x(i, n-i) \tag{6.12b}$$

$n = 0$ corresponds to $\omega = 0$ and $n = N$ to $\omega = \pi$. Two initial conditions need to be specified, namely, $\varphi_A(0)$ and $\varphi_A(1)$. The value of $\varphi_A(0)$ may be zero or $\pm\pi$ and is determined from $\Psi_x(0, 0)$. Estimation of $\varphi_A(1)$ is obtained from [60]

$$\varphi_A(1) = \left\{\sum_{n=2}^{N} (S(n) - S(n-1))/(n(n-1))\right\} + \varphi_A(N)/N \tag{6.12c}$$

where $\varphi_A(N)$ can be set to zero or $k\pi$, i.e., being equivalent to pure time delay. Let us note that the algorithm utilizes all the bispectrum values of the first triangular region in Fig. 2(b) to reconstruct the wavelet phase.

2) The LR Algorithm [47], [49]: The recursive formula is given by [60]

$$\varphi_A(n) = -\sum_{i=0}^{n-1} \Psi_x(i, 1) + \varphi_A(0) + n\varphi_A(1) \tag{6.13}$$

where the initial conditions are determined as before. However, this algorithm is limited to the use of the bispectrum values along a straight line parallel to either ω_1 or ω_2 and, therefore, does not utilize all the available information.

3) The MU Algorithm [60]: Defining

$$\boldsymbol{\varphi} = [\varphi_A(1), \varphi_A(2), \cdots, \varphi_A(N-1)]^T$$

$$\boldsymbol{\Psi} = [\Psi_x(1, 1), \Psi_x(1, 2), \cdots, \Psi_x(2, 2), \Psi_x(2, 3), \cdots, \Psi_x(N/2 \cdot N/2)]^T$$

the following set of equations could be formed using (6.10):

$$\boldsymbol{A}\,\boldsymbol{\varphi} = \boldsymbol{\Psi} \tag{6.14a}$$

where $\boldsymbol{A}$ is a sparse coefficient matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & -1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 2 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 0 & -1 & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \end{bmatrix}$$

of size $(N/2)^2 \times (N-1)$ for N even and

$$\left[\frac{(N-1)(N+1)}{4}\right] \times [(N-1)]$$

for N odd. The unknown phase vector is obtained using a least squares solution

$$\boldsymbol{\varphi} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\Psi}. \tag{6.14b}$$

The MU algorithm is a nonrecursive method that utilizes all the bispectrum values available in the first triangular region of Fig. 2(b).

4) The ARMA Bispectrum Approach [67], [74], [75]: Consider the data record $\{X(n)\}$ generated by (6.5). Assuming a generally mixed-phase wavelet, (6.6) can be written as

$$A(z) = I(z^{-1})\,O(z) \tag{6.15}$$

where $I(z^{-1})$ and $O(z)$ are polynomial with all their zeros, respectively, inside and outside the unit circle. Combining (6.15) and (6.5) we obtain

$$X(z) = W(z)\,I(z^{-1})\,O(z). \tag{6.16}$$

It was suggested in [67], [74], [75] that (6.16) can be rewritten as

$$X(z)/I(z^{-1}) = W(z)\,O(z)$$

and be approximated arbitrarily closely by a causal (or forward) stable ARMA (p, q) model of the form

$$\sum_{i=0}^{p} a_i X(n-i) \simeq \sum_{k=0}^{q} b_k W(n+k), \qquad a_0 = 1. \tag{6.17a}$$

The AR coefficients $\{a_i\}$ contain the minimum-phase component of the mixed-phase wavelet. The model (6.17a) satisfies the third-order recursion equation

$$\sum_{i=0}^{p} a_i R(-\tau+i, -\rho+i) = 0, \qquad \text{for } \tau > q, \rho > q \tag{6.17b}$$

where $R(m, n)$ is the third-moment sequence of $\{X(n)\}$. From (6.17b) an overdetermined system of equations is formed and solved via least squares. The minimum-phase component is then obtained from

$$|A_1(\omega)| = \text{Magn}\left\{1\Big/\sum_{i=0}^{p} a_i \exp j(-\omega i)\right\} \tag{6.17c}$$

$$\varphi_1(\omega) = \text{Arg}\left\{1\Big/\sum_{i=0}^{p} a_i \exp j(-\omega i)\right\}. \tag{6.17d}$$

Note that $\{R(m, n)\}$ in (6.17b) is replaced by estimates $\{\hat{R}(m, n)\}$ obtained following (4.1) and (4.2).

Equation (6.16) can also be written as

$$X(z)/O(z) = W(z)\,I(z^{-1})$$

and similarly be approximated arbitrarily closely by an anticausal (backward) stable ARMA (r, s) model of the form

$$\sum_{i=0}^{r} c_i X(n+i) \simeq \sum_{k=0}^{s} d_k W(n-k), \qquad c_0 = 1. \tag{6.18a}$$

The AR coefficients $\{c_i\}$ contain the maximum-phase component of the mixed-phase wavelet. The third-order recursion of (6.18b) is given by

$$\sum_{i=0}^{r} c_i R(\tau-i, \rho-i) = 0, \qquad \text{for } \tau > s, \rho > s \tag{6.18b}$$

which is solved for $\{c_i\}$ using the pseudo-inverse least squares approach. The maximum-phase component of (6.6) is obtained from

$$|A_0(\omega)| = \text{Magn}\left\{1\Big/\sum_{i=0}^{r} c_i \exp j(\omega i)\right\} \tag{6.18c}$$

$$\varphi_0(\omega) = \text{Arg}\left\{1\Big/\sum_{i=0}^{r} c_i \exp j(\omega i)\right\}. \tag{6.18d}$$

The magnitude and phase of the wavelet (system) are easily obtained by combining the results of (6.17c), (6.17d), and (6.18c) and (6.18d) as follows:

$$|A(\omega)| = |A_1(\omega)| \cdot |A_0(\omega)| \tag{6.19a}$$

$$\varphi_A(\omega) = \varphi_1(\omega) + \varphi_0(\omega). \tag{6.19b}$$

The ARMA bispectrum approach generates the minimum ($\varphi_1(\omega)$) and the maximum ($\varphi_0(\omega)$) phase components of the wavelet directly from the data $\{X(n)\}$ bypassing an overall estimation of the bispectrum. Hence, the reconstruction algorithms that generate $\varphi_A(\omega)$ from $\Psi_x(\omega_1, \omega_2)$ become unnecessary for the parametric approach. Extension of the ARMA method to the trispectrum case as well as model order-selection criteria can be found in [68]. Comparisons with conventional methods are provided in [67], [74], [75].

Phase reconstruction methods can also be found in [3], [4], [53] [54]. Phase estimation in the trispectrum domain has been discussed in [60], [77] and investigated further in [45], [50], [91], [116]–[118].

5) Cepstrum Modeling of Third-Order Moments and Non-Minimum-Phase Signal Reconstruction [70]: One of the early approaches for non-minimum-phase signal reconstruction and deconvolution has been homomorphic filtering based on complex cepstrum [109] or differential cepstrum [80] operations in the available data $\{X(n)\}$. However, these cepstrum approaches work well only if $\{X(n)\}$ is the result of a linear convolution between an impulse train and a non-minimum-phase impulse response signal. In the case where non-Gaussian white noise is convolved with the impulse response signal, cepstrum approaches on $\{X(n)\}$ will fail to provide satisfactory results. The basic idea explored in [70] was to perform non-minimum-phase impulse response reconstruction by applying cepstrum operations on the higher order cumulants of $\{X(n)\}$.

Assuming that the observed sequence is generated by an ARMA system (rather than MA as in (6.5)) described by

$$X(k) = -\sum_{i=1}^{L_3} d_i X(k-i) + \sum_{i=-L_1}^{L_2} u_i W(k+i) \tag{6.20a}$$

where $\{W(k)\}$ is non-Gaussian, white, i.i.d with $E\{W(k)\} = 0$, $E\{W(k)\,W(k+\tau)\} = Q \cdot \delta(\tau)$, and $E\{W(k)\,W(k+\tau)\,W(k+\rho)\} = \beta \cdot \delta(\tau, \rho)$, the problem is to reconstruct the magnitude and phase response of the system from the given output data $\{X(n)\}$ or equivalently its impulse response. Since the system is generally non-minimum-phase and stable, its transfer function may be written in terms of poles and zeros as follows:

$$H(z) = I(Z^{-1})\,O(Z) \tag{6.20b}$$

where

$$I(Z^{-1}) = \frac{\prod_{i=1}^{L_1}(1 - a_i Z^{-1})}{\prod_{i=1}^{L_3}(1 - c_i Z^{-1})}, \qquad |a_i| < 1, |c_i| < 1 \tag{6.20c}$$

is the minimum-phase component with poles $\{c_i\}$ and zeros $\{a_i\}$ inside the unit circle and

$$O(Z) = \prod_{i=1}^{L_2} (1 - b_i Z), \qquad |b_i| < 1 \tag{6.20d}$$

is the maximum-phase component with zeros outside the unit circle. Similarly, the minimum-phase impulse response of the ARMA system is given by

$$i(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} I(-\omega) \exp j(\omega n)\, d\omega, \qquad n \geq 0 \text{ (causal)} \tag{6.20e}$$

and the maximum-phase component

$$o(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} O(\omega) \exp j(\omega n)\, d\omega, \qquad n \leq 0 \text{ (anticausal).} \tag{6.20f}$$

Combining (3.15) with (6.20b), it follows that the bispectrum of the observed data $\{X(n)\}$ is given by

$$B_x(Z_1, Z_2) = \beta \cdot I(Z_1^{-1}) \cdot I(Z_2^{-1}) \cdot I(Z_1 Z_2) \cdot O(Z_1) \cdot O(Z_2) \cdot O(Z_1^{-1} Z_2^{-1}). \tag{6.20g}$$

Therefore, the Z-transform of the two-dimensional complex cepstrum of third moments

$$\{R_x(\tau, \rho)\},\ R_x(\tau, \rho) \triangleq E\{X(n)\, X(n + \tau)\, X(n + \rho)\}$$

can be defined as

$$\begin{aligned} C_x(Z_1, Z_2) \triangleq \ln [B_x(Z_1, Z_2)] &= \ln |\beta| + \ln I(Z_1^{-1}) \\ &+ \ln I(Z_2^{-1}) + \ln I(Z_1 Z_2) + \ln O(Z_1) \\ &+ \ln O(Z_2) + \ln O(Z_1^{-1} Z_2^{-1}). \end{aligned} \tag{6.20h}$$

Combining (6.20g) and (6.20h) and using power series expansion, it can be easily shown that the complex cepstrum $c_x(m, n)$ of third moments in [70]

$$c_x(m, n) \triangleq Z^{-1}[C_x(Z_1, Z_2)] = \begin{cases} \ln |\beta|, & m = 0, n = 0 \\ -\frac{1}{n} A^{(n)}, & m = 0, n > 0 \\ -\frac{1}{m} A^{(m)}, & n = 0, m > 0 \\ -\frac{1}{m} B^{(-m)}, & n = 0, m < 0 \\ \frac{1}{n} B^{(-n)}, & m = 0, n < 0 \\ -\frac{1}{n} B^{(n)}, & m = n > 0 \\ \frac{1}{n} A^{(-n)}, & m = n < 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.20i}$$

where

$$A^{(k)} \triangleq \sum_{i=1}^{L_1} a_i^k - \sum_{i=1}^{L_3} c_i^k$$

$$B^{(k)} \triangleq \sum_{i=1}^{L_2} b_i^k$$

are parameters which contain the minimum- and maximum-phase information, respectively.

Since $C_x(Z_1, Z_2)$ is analytic in its region of convergence which contains the unit surface, the following linear convolution (*) operation has been shown to be valid [70]:

$$R_x(m, n) * [-mc_x(m, n)] = -mR_x(m, n). \tag{6.20j}$$

Substitution of (6.20d) into (6.20j) results in the following key *cepstral equation*:

$$\begin{aligned} \sum_{k=1}^{\infty} \{A^{(k)} [R_x(m - k, n) - R_x(m + k, n + k)] \\ + B^{(k)} [R_x(m - k, n - k) - R_x(m + k, n)]\} = -mR_x(m, n) \end{aligned}$$

which provides a direct relationship between parameters $\{A^{(k)}\}$ $\{B^{(k)}\}$ and third moments $\{R_x(m, n)\}$. Since $|a_i|, |b_i| < 1$ and $|c_i| < 1$ for all $\{i\}$, we can always tuncate it arbitrarily closely and obtain the approximate cepstral equation

$$\begin{aligned} \sum_{i=1}^{p} \{A^{(i)} [R_x(m - i, n) - R_x(m + i, n + i)] \\ + \sum_{j=1}^{q} B^{(j)} [R_x(m - j, n - j) - R_x(m + j, n)]\} \\ \cong -mR_x(m, n). \end{aligned} \tag{6.20k}$$

Equation (6.20k) may easily be written as an overdetermined system of equations with unknowns the cepstral parameters $\{A^{(i)}\}$, $\{B^{(i)}\}$. A least squares solution is then obtained for the cepstral parameters. The elegancy of the result shown by (6.20i) and (6.20k) lies on its connection with the differential cepstrum [80] of the system's impulse response.

From (6.20b), (6.20e), and (6.20f), the system impulse response $\{h(n)\}$ is given by

$$h(n) = i(n) * o(n).$$

The differential cepstrum of $h(n)$ was defined in [80] as

$$h_d(n) \triangleq Z^{-1}\left\{\frac{1}{H(Z)} \frac{\partial H(Z)}{\partial Z}\right\}.$$

From the properties of differential cepstrum, we have [80]

$$h_d(n) = i_d(n) + o_d(n)$$

where $\{i_d(n)\}$, $\{o_d(n)\}$ are the differential cepstra of $\{i(n)\}$ and $\{o(n)\}$, respectively. It was shown in [70] that

$$i_d(n) = \begin{cases} A^{(n-1)}, & n \geq 2 \\ 0, & n \leq 1 \end{cases} \tag{6.20ℓ}$$

$$o_d(n) = \begin{cases} 0, & n \geq 1 \\ -B^{(1-n)}, & n \leq 0 \end{cases}. \tag{6.20m}$$

In practice, the third moments of $\{X(n)\}$ are estimated as in (4.1), (4.2) and the cepstral parameters $\{A^{(i)}\}$, $\{B^{(i)}\}$ are computed from (6.20k) via least squares. Then the differential cepstra are generated from (6.20ℓ) and (6.20m). The minimum- and maximum-phase impulse response components described by (6.20e) and (6.20f), respectively, are

generated using the recursive relations [80]

$$i(n) = -\frac{1}{n}\sum_{k=2}^{n+1} i_d(k)\, i(n-k+1), \qquad \text{for } n \geq 1$$

$$o(n) = -\frac{1}{n}\sum_{k=n+1}^{0} o_d(k)\, o(n-k+1), \qquad \text{for } n \leq -1$$

with $i(0) = o(0) = 1$.

Several properties of this approach have been outlined in [70]: 1) The minimum- and maximum-phase components of the impulse response are estimated separately. 2) The method is computationally attractive as it involves a solution of a linear system of equations. 3) It works for MA, AR, or ARMA signals (or systems). 4) It does not require model order selection criteria. 5) The length of minimum and maximum impulse responses is determined by the algorithm itself based on prechosen threshold value. 6) The method does not require *a priori* knowledge of the type of the model (AR, MA, or ARMA).

C. Time-Delay Estimation [97], [99]

Time-delay estimation techniques have been found useful in many fields such as sonar, radar, biomedicine, geophysics [39], [40], and optics [54]. Let us assume that $\{X(n)\}$ and $\{Y(n)\}$ are two available sensor measurements satisfying

$$\begin{aligned} X(n) &= S(n) + E(n) \\ Y(n) &= S(n-D) + G(n) \end{aligned} \tag{6.21}$$

where $S(n)$ is an unknown signal and $S(n-D)$ is the same signal delayed (or advanced) in time, and $E(n)$, $G(n)$ are unknown noise sources. The problem is to find from finite records of $X(n)$ and $Y(n)$ an estimate of the time delay D.

The basic approach to the solution of the time-delay estimation problem is to shift $X(n)$ with respect to $Y(n)$ and compare similarities between the two records at each shift. Best match will occur at a shift equal to D. Assuming that the noise sources are zero-mean independent random processes, the fundamental operation adopted to "compare similarities" between $X(n)$ and $Y(n)$ is the cross correlation which ideally is

$$r_{xy}(\tau) = E\{X(n)\, Y(n+\tau)\} = R_{ss}(\tau - D), \qquad -\infty < \tau < \infty. \tag{6.22}$$

The $r_{xy}(\tau)$ peaks at $\tau = D$. However, in practical application problems, due to finite-length data records and not exactly independent noise sources, the $r_{xy}(\tau)$ does not necessarily show a peak at the time-delay position. Various window functions (prefilters) have been suggested to better shape the cross-correlation function via convolution operation, namely ROTH, SCOT, PHAT, Eckart, Hannan–Thompson, to name a few [111]. Parameter estimation approaches have also been suggested for time-delay estimation based on autocorrelations and cross correlation such as the least squares and Wiener filtering method [111]. Each of those techniques has certain limitations depending upon the nature of the signal and noise sources.

In practical application problems where the signal $\{S(n)\}$ can be regarded as non-Gaussian stationary random process, and the noise sources $\{E(n)\}$, $\{G(n)\}$ independent stationary Gaussian, the similarities between $\{X(n)\}$ and $\{Y(n)\}$ could also be "compared" in higher order spectrum domains such as the bispectrum. Let us note that self-emitting signals from complicated mechanical systems contain strong quasi-periodic components and therefore can be regarded as non-Gaussian signals [97], [99], [100]. The main motivation behind the use of higher order spectra for time-delay estimation under the aforementioned assumptions is the fact that they are free of Gaussian noise (in theory). Assuming that $\{S(n)\}$ has also a nonzero measure of skewness, the following identities hold:

$$\begin{aligned} R_{xxx}(\tau, \rho) &= E\{X(n)\, X(n+\tau)\, X(n+\rho)\} = R_s(\tau, \rho) \\ R_{xyx}(\tau, \rho) &= E\{(X(n)\}\, Y(n+\tau)\, X(n+\rho)\} = R_s(\tau - D, \rho) \end{aligned} \tag{6.23}$$

because the third-moment sequence of a zero-mean Gaussian process is identically zero. Obtaining the bispectra of the third-moment sequences in (6.23) we have

$$\begin{aligned} B_{xxx}(\omega_1, \omega_2) &= \text{FT}\,\{R_{xxx}(\tau, \rho)\} = B_s(\omega_1, \omega_2) \\ B_{xyx}(\omega_1, \omega_2) &= \text{FT}\,\{R_{xyx}(\tau, \rho)\} = B_s(\omega_1, \omega_2)\, e^{j\omega_1 D} \end{aligned} \tag{6.24}$$

where $\text{FT}\{\cdot\}$ denotes the two-dimensional Fourier transform operation. The hologram signal can be reconstructed using (6.24) by the ratio

$$\begin{aligned} I(\omega_1, \omega_2) &= B_{xyx}(\omega_1, \omega_2)/B_{xxx}(\omega_1, \omega_2) \\ I(\omega_1, \omega_2) &= e^{j\omega_1 D} \end{aligned} \tag{6.25}$$

where $B_{xxx}(\omega_1, \omega_2)$ is assumed to be nonzero. This equation also indicates that the hologram is free of Gaussian noises. One way of computing the time delay D is to form the function

$$T(\tau) = \int_{-\pi}^{+\pi}\int e^{-j(\tau - D)\omega_1}\, d\omega_1\, d\omega_2 \tag{6.26}$$

which peaks at $\tau = D$. There are, of course, several other ways that can be adopted to compute D either by directly using (6.23) and cross moments or (6.25) and least squares straight line (fitted on the phase of $I(\omega_1, \omega_2)$). Simulation examples that demonstrate better performance of (6.26) over (6.22) have been reported in [97], [99]. However, time-delay estimation based on (6.23)–(6.26) and finite-length data records is still an open problem for research.

VII. Conclusions and Future Trends

This paper presented, within a digital signal processing framework, a tutorial review of bispectrum estimation and modeling of third-order moments. The general motivations behind the use of the bispectrum has been discussed, namely, deviations from normality, phase estimation, and detection and characterization of nonlinear mechanisms that generate time series. In addition, various areas of application that can directly benefit from the use of the bispectrum have been identified and discussed in the paper. Although the paper has emphasized bispectrum estimation, let us note that the extension of the discussion to the trispectrum case is somewhat straightforwad.

There is plenty of room for research in this area with several lines of investigation to be pursued. For example, one could study the use of trispectrum estimation and the mod-

eling of fourth-order cumulants via AR and ARMA models for phase estimation. This approach will turn out to be very useful in the case of non-Gaussian processes with zero measure of skewness. Model order selection criteria for general ARMA bispectrum and trispectrum estimation could be developed. Adaptive parametric schemes that capture the information present in higher order cumulants of a given process is another topic worth looking into. These adpative schemes can be very useful for adaptive deconvolution, communication channels (equalizers), and array processing. One could also investigate the use of nonlinear Volterra filters for higher order spectrum estimation. Filtering of impulsive noise via polyspectra and time-delay estimation of arrivals in the polyspectrum domain are yet two more problems worth investigating. Finally, image reconstruction could be carried out in polyspectrum domain, and detection schemes could be devised based on higher order cumulants.

References

[1] H. Akaike, "Note on higher order spectra," *Ann. Inst. Statist. Math.*, vol. 18, pp. 123–126, 1966.

[2] ——, "Notes on lectures at the Institute of Statistical Mathematics," *The Statistical Theory of Spectral Estimation*, pp. 1–30, 1964.

[3] H. O. Bartelt, A. W. Lohmann, and B. Wirnitzer, "Speckle masking in astromony—Triple correlation theory and applications," *Appl. Opt.*, vol. 22, no. 24, Dec. 1983.

[4] ——, "Phase and amplitude recovery from bispectra," *Appl. Opt.*, vol. 23, pp. 3121–3129, Sept. 1984.

[5] J. S. Bendat and A. G. Piersol, "Spectral analysis of nonlinear systems involving square-law operations," *J. Sound Vibr.*, vol. 81, pp. 199–212, 1982.

[6] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of nonminimum phase system: Blind adjustment of linear equalizer in data communications," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 385–398, June 1980.

[7] A. Blanc-Lapierre and R. Forter, *Theorie des Fonctions Aleatoires*. Paris, France: Masson, 1953.

[8] D. R. Brillinger, "An introduction to polyspectra," *Ann. Math. Statist.*, vol. 36, pp. 1351–1374, 1965.

[9] D. R. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of k-th order spectra," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York, NY: Wiley, 1967, pp. 153–188.

[10] ——, "Computation and interpretation of k-th order spectra," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York, NY: Wiley, 1967, pp. 189–232.

[11] D. R. Brillinger, "The identification of a particular nonlinear time series system," *Biometrika*, vol. 65, pp. 509–515, 1977.

[12] ——, *Time Series, Data Analysis and Theory* (expanded edition). San Francisco, CA: Holden-Day, 1981.

[13] ——, "The identification of polynomial systems by means of higher order spectra," *J. Sound Vibr.*, vol. 12, pp. 301–313, 1970.

[14] D. E. Cartwright, "A unified analysis of tides and surges round North and East Britain, *Philos. Trans. Roy. Soc. London*, ser. A 263, pp. 1–55, 1968.

[15] D. G. Childers, Ed. *Modern Spectrum Analysis*. New York, NY: IEEE PRESS, 1978.

[16] H. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[17] P. F. Dwyer, "Detection of non-Gaussian signals by frequency domain kurtosis estimation," in *Proc. ICASSP'83*, vol. 2, pp. 607–610 (Boston, MA, Apr. 1983).

[18] J. M. Florence and J. H. Song, "Real-time acousto-optic bispectral processor, *Proc. SPIE*, vol. 431, pp. 431–436, 1986.

[19] H. Gamo, "Triple correlator of photo electric fluctuations as a spectroscopic tool," *J. Appl. Phys.*, vol. 34, no. 4, pp. 875–876, Apr. 1963.

[20] V. Kh. German, S. P. Levikov, and A. S. Tsvetsinskii, "Bispectral analysis of sea-level variations," *Sov. Meteorol. Hydrol.*, no. 11, pp. 50–57, 1980.

[21] G. B. Giannakis and J. M. Mendel, "Stochastic realization of non-minimum phase systems," in *Proc. American Control Conf. (ACC)*, pp. 1254–1259 (Seattle, WA, June 1986).

[22] ——, "Approximate realization and model reduction of non-minimum phase systems," in *Proc. 25th IEEE Conf. on Decision and Control (CDC)* (Athens, Greece, Dec. 1986).

[23] M. D. Godfrey, "An exploratory study of the bispectrum of economic time series," *Appl. Statist.*, vol. 14, pp. 48–69, 1965.

[24] R. Godfrey and F. Rocca, "Zero-memory nonlinear deconvolution," *Geophys. Prospect.*, vol. 29, pp. 189–228, 1981.

[25] K. Hasselman, W. Munk, and G. MacDonald, "Bispectra of ocean waves," in *Time Series Analysis*, M. Rossenblatt, Ed. New York, NY: Wiley, 1963, pp. 125–130.

[26] R. A. Haubrich, "Earth noise, 5 to 500 millicycles per second," *J. Geophys. Res.*, vol. 70, pp. 1415–1427, 1965.

[27] S. Haykin, ed., *Nonlinear Methods of Spectral Analysis*. New York, NY: Springer-Verlag, 1979.

[28] K. N. Helland, E. C. Itsweire, and K. S. Lii, "A program for computation of bispectra with application to spectral energy transfer in fluid turbulence," *Adv. Eng. Software*, vol. 7, no. 1, pp. 22–27, 1985.

[29] K. N. Helland, K. S. Lii, and M. Rosenblatt, "Bispectra of atmospheric and wind tunnel turbulences," in *Applications of Statistics*, P. R. Krishnaiah, Ed. New York, NY: North Holland, 1977, pp. 223–248.

[30] ——, "Bispectra and energy transfer in grid-generated turbulence," in *Developments in Statistics*, vol. 2, P. R. Krishnaiah, Ed. New York, NY: Academic Press, 1978, pp. 123–155.

[31] B. I. Helme and C. L. Nikias, "Improved spectrum performance via a data-adaptive weighted Burg technique," *IEEE Trans. Acoustics, Speech Signal Processing*, vol. ASSP-33, no. 4, pp. 903–910, Aug. 1985.

[32] J. R. Herring, "Theoretical calculations on turbulent bispectra," *J. Fluid Mech.*, vol. 97, pp. 193–204, 1980.

[33] M. J. Hinich and C. S. Clay, "The application of the discrete Fourier transform in the estimation of power spectra coherence and bispectra of geophysical data," *Rev. Geophys.*, vol. 6,no. 3, pp. 347–363, 1968.

[34] M. J. Hinich and D. M. Patterson, "Identification of the coefficients in a non-linear time series of the quadratic type," *J. Econometrics*, vol. 30, pp. 269–288, 1985 (North-Holland).

[35] ——, "Evidence of nonlinearlity in daily stock returns," *J. Bus. Econ. Stat.*, vol. 3, no. 1, pp. 69–77, 1985.

[36] J. W. J. Hosken, "A stochastic model of seismic reflections," presented at the *50th Ann. Int. Meet. of SEG*, 1980.

[37] P. J. Huber, B. Kleiner, T. Gasser, and G. Dumermuth, "Statistical methods for investigating phase relations in stationary stochastic processes," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 78–86, 1971.

[38] M. Huzii, "Estimation of coefficient of an autoregressive process by using a higher order moment," *J. Time Ser. Anal.*, vol. 2, pp. 87–93, 1981.

[39] S. M. Kay and L. Marple, "Spectrum analysis: A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1418, Nov. 1981.

[40] S. B. Kesler, Ed., *Modern Spectrum Analysis II*. New York, NY: IEEE PRESS, 1986.

[41] Y. C. Kim, J. M. Beall, E. J. Powers, and R. W. Miksad, "Bispectrum and nonlinear wave coupling," *Phys. Fluids*, vol. 23, no. 2, pp. 258–263, Feb. 1980.

[42] Y. C. Kim and E. J. Powers, "Digital bispectral analysis of self-excited fluctuations spectra," *Phys. Fluids*, vol. 21, no. 8, pp. 1452–1453, 1978.

[43] K. Knast and W. Chmielowski, "Study of the high-order correlation functions of numer fluctuations in simple fluids with radial hand-body interactions," *J. Phys.* (France), vol. 42, no. 10, pp. 1373–1385, Oct. 1981.

[44] J. Korein, L. J. Tick, R. A. Zeitlin, and C. T. Randt, "Linear and nonlinear spectral analytic techniques applied to the human electroencephalogram," *Corr. Sci. Health Res. Council*, vol. 4, pp. 1126–1128, 1968.

[45] K. S. Lii, "Non-Gaussian ARMA model identification and estimation," *Proc. Bus. and Econ. Statistics* (ASA), pp. 135–141, 1982.

[46] K. S. Lii and K. N. Helland, "Cross-bispectrum computation

and variance estimation," *ACM Trans. Math. Software*, vol. 7, pp. 284–294, 1981.
[47] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Ann. Statist.*, vol. 10, pp. 1195–1208, 1982.
[48] K. S. Lii, M. Rosenblatt, and C. Van Atta, "Bispectral measurements in turbulence," *J. Fluid Mech.*, vol. 77, pp. 45–62, 1976.
[49] K. S. Lii and M. Rosenblatt, "Non-Gaussian linear processes, phase, and deconvolution," in *Statistical Signal Processing*, E. J. Wegman and J. G. Smith, Eds., 1984, pp. 51–58.
[50] ——, "A fourth order deconvolution technique for non-Gaussian linear processes" in *Multivariate Analysis VI*, P. R. Krishnaiah, Ed. Amsterdam, The Netherlands: Elsevier, 1985, pp. 395–410.
[51] K. S. Lii, K. N. Helland, and M. Rosenblatt, "Estimating three-dimensional energy transfer in isotropic turbulence," *J. Time. Ser. Anal.*, vol. 3, no. 1, pp. 1–28, 1982.
[52] L. Ljung, "Consistency of the least-squares identification method," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 779–781, Oct. 1976.
[53] A. W. Lohmann, G. P. Weigelt, and B. Wirnitzer, "Speckle masking in astronomy-triple correlation theory and applications," *App. Opt.*, vol. 22, no. 24, pp. 4028–4037, Dec. 1983.
[54] A. W. Lohmann and B. Wirnitzer, "Triple correlations," *Proc. IEEE.*, vol. 72, no. 7, pp. 889–901, 1984.
[55] J. L. Lumley and K. Takeuchi, "Application of central-limit theorems to turbulence and higher-order spectra," *J. Fluid Mech.*, vol. 74, pp. 433–468, 1976.
[56] G. MacDonald, "The bispectra of atmospheric pressure records," in *Proc. IBM Sci. Comput. Symp. on Statistics* (White Plains, NY: IBM) pp. 247–264, 1963.
[57] T. A. Magness, "Spectral response of a quadratic device to non-Gaussian noise," *J. Appl. Phys.*, vol. 25, pp. 1357–1365, 1954.
[58] J. H. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.
[59] V. Z. Marmarelis and D. Sheby, "Bispectral analysis of weakly nonlinear quadratic systems," in *Proc. IEEE 3rd ASSP Workshop on Spectrum Estimation and Modeling*, pp. 14–16, (Boston, MA, Nov. 17–18, 1986).
[60] T. Matsuoka and T. J. Ulrych, "Phase estimation using the bispectrum," *Proc. IEEE*, vol. 72, pp. 1403–1411, 1984.
[61] J. H. McClellan, "Multidimensional spectral estimation," *Proc. IEEE*, vol. 70, pp. 1029–1039, Sept. 1982.
[62] C. H. McComas, "Bispectra of internal waves," Woods Hole Tech. Note 02543, Woods Hole, MA, 1978.
[63] C. H. McComas and M. G. Briscoe, "Bispectra of internal waves," *J. Fluid Mech.*, vol. 97, pp. 205–213, 1980.
[64] Y. Nagata, "Bispectra of spike-array type time series and their application to the analysis of oceanic microstructures," *J. Oceanographical Soc. Japan*, vol. 34, pp. 204–216, Oct. 1978.
[65] S. Neshyba and H. Crew, "Rotary cross-bispectra and energy transfer functions between non-Gaussian vector processes," *J. Phys. Oceanogr.*, vol. 7, no. 6, pp. 892–903, Nov. 1977.
[66] S. Neshyba and E. J. C. Sobey, "Vertical cross coherence and cross bispectra between internal waves measured in a multiple-layered ocean," *J. Geophys. Res.*, vol. 80, no. 9, pp. 1152–1162, Mar. 1975.
[67] C. L. Nikias, "ARMA bispectrum approach to nonminimum phase system identification," *Tech. Tep. CDSP-86-102*, Dept. Electrical and Computer Eng., Northeastern Univ., Boston, MA, Aug. 1986.
[68] ——, "Parametric trispectrum estimation," in *Proc. IEEE 3rd ASSP Workshop on Spectrum Estimation and Modeling*, pp. 17–20 (Boston, MA, Nov. 17–18, 1987).
[69] C. L. Nikias and H. H. Chiang, "Non-causal autoregressive bispectrum estimation and deconvolution," in *Proc. IEEE Int. Conf. ASSP (ICASSP'87)* (Dallas, TX, Apr. 6–9, 1987).
[70] C.L. Nikias and R. Pan, "Nonminimum phase system identification via cepstrum modeling of higher-order cumulants," in *Proc. IEEE Int. Conf. ASSP (ICAASSP'87)*, (Dallas, TX, Apr. 6–9, 1987).
[71] C. L. Nikias and M. R. Raghuveer, "Discussion: Higher-order autospectra by maximum entropy method," *Geophysics*, vol. 50, no. 1, pp. 165–166, Jan. 1985.
[72] C. L. Nikias and M. R. Raghuveer, "Multidimensional parametric spectral estimation," *Signal Processing*, vol. 9, no. 3, Oct. 1985.
[73] C. L. Nikias and P. D. Scott, "The covariance least-squares algorithm for spectral estimation of processes of short data length," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-21, no. 2, pp. 180–190, Apr. 1983.
[74] C. L. Nikias, A. N. Venetsanopoulos, and B. Paramasivaiah, "Mixed-phase wavelet estimation via autoregressive bispectrum," in *Proc. 1985 Int. Geosc. Remote Sensing Symp.*, vol. 1. Amherst, MA: U. Mass Press, Oct. 1985, pp. 66–71.
[75] ——, "Identification of nonminimum phase communication channels via parametric modeling of third moments," in *Proc. Int. Conf. Communications* (Toronto, Ont., Canada, June 1986).
[76] T. Ning and C. L. Nikias, "Multichannel AR spectrum estimation: The optimum approach in the reflection coefficient domain," *IEEE Trans. Acoustics, Speech Signal Processing*, vol. ASSP-34, no. 5, pp. 1139–1152, Oct. 1986.
[77] R. Pan and C. L. Nikias, "Phase reconstruction in the trispectrum domain," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 895–897, June 1987.
[78] E. Parzen, "Time series analysis for models of signal plus white noise, in *Spectral Analysis of Time Series*, B. Harris, ED. New York, NY: Wiley, 1967, pp. 233–257.
[79] B. Picinbono, "Quadratic filters," in *Proc. ICASSP'82*, pp. 298–301 (Paris, France, 1982).
[80] A. Polydoros and A. Fam, "The differential cepstrum: definition and properties," in *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 77–80, Apr. 1981.
[81] J. G. Proakis, "Advances in equalization for intersymbol interference," in *Advances in Communication Systems*, vol. 4. New York, NY: Academic Press, 1975, pp. 124–194.
[82] M. R. Raghuveer, "Bispectrum and multidimensional power spectrum estimation algorithms based on parametric models with applications to the analysis of ECG data," Ph.D. dissertation, Univ. of Connecticut, Storrs, Dec. 1984.
[83] ——, "Multichannel bispectrum estimation," in *Proc. 3rd IEEE ASSP Workshop on Spectrum Estimation and Modeling*, pp. 21–24 (Boston, MA, Nov. 27–28, 1986).
[84] M. R. Raghuveer and C. L. Nikias, "Bispectrum estimation for short length data," in *Proc. ICASSP'85*, pp. 1352–1355, 1985.
[85] ——, "A parametric approach to bispectrum estimation," in *Proc. ICASSP'84*, pp. 38.1.1–38.1.4, 1984.
[86] ——, "Bispectrum estimation: A parametric approach," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, no. 5, pp. 1213–1230, Oct. 1985.
[87] ——, "Bispectrum estimation via AR modeling," *Signal Processing (Special Issue on Modern Trends in Spectral Analysis)*, vol. 9, no. 1, pp. 35–48, Jan. 1986.
[88] G. I. Roden and D. J. Bendiner, "Bispectra and cross-bispectra of temperature, salinity, sound velocity and density fluctuations with depth off Northeastern Japan," *J. Phys. Ocean*, vol. 3, no. 3, pp. 308–317, 1973.
[89] M. Rosenblatt, "Linear processes and bispectra," *J. Appl. Probab.*, vol. 17, pp. 265–270, 1980.
[90] M. Rosenblatt and J. W. Van Ness, "Estimation of the bispectrum," *Ann. Math. Statist.*, vol. 36, pp. 1120–1136, 1965.
[91] M. Rosenblatt, "Cumulants and cumulant spectra," in *Time Series in the Frequency Domain*, D. Brillinger and P. Krishnaiah, Eds. Amsterdam, The Netherlands: North-Holland, 1983, pp. 369–382.
[92] M. Rosenblatt, *Random Processes*. New York, NY: Oxford Press, 1962.
[93] ——, "Statistical analysis of stochastic processes with stationary residuals," in *Probability and Statistics—The Harald Cramer Volume*, U. Grenander, Ed. New York, NY: Wiley, 1959, pp. 125–139.
[94] ——, "Linear random fields," in *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York, NY: Academic Press, 1983, pp. 299–309.

[95] ——, "Linearity and nonlinearity in time series: Prediction," in *Proc. 42nd Session Int. Statis. Inst. (Manila), Book 1*, pp. 423–434, 1979.
[96] O. Sasaki, T. Sato, and T. Oda, "Laser Doppler vibration measuring system using bispectral analysis," *Appl. Opt.*, vol. 19, no. 1, pp. 151–153, Jan. 1980.
[97] O. Sasaki, T. Sato, and Y. Nakamura, "Holographic passive sonar," *IEEE Trans. Sonics Ultrason.*, vol. SU-24, no. 3, pp. 193–200, May 1977.
[98] K. Sasaki, T. Sato, and Y. Yamashita, "Minimum bias windows for bispectral estimation," *J. Sound Vib.*, vol. 40, no. 1, pp. 139–148, 1975.
[99] T. Sato and K. Sasaki, "Bispectral holography," *J. Acoust. Soc. Amer.*, vol. 62, pp. 404–408, 1977.
[100] T. Sato, K. Sasaki, and Y. Nakamura, "Real-time bispectral analysis of gear noise and its application to contactless diagnosis," *J. Acoust. Soc. Amer.*, vol. 62, pp. 382–387, 1977.
[101] T. Sato and O. Sasaki, "New 3-D laser Doppler velocimeter using cross-bispectral analysis," *Appl. Opt.*, vol. 17, no. 24, pp. 3890–3894, Dec. 1978.
[102] T. Sato, S. Wadaka, J. Ishii, and T. Sunada, "A new ultrasonic imaging system using a rotating random phase disc and power spectral and third order correlation analysis," in *7th Int. Symp. on Acoustic Imaging and Holography Proc.*, vol. 16, pp. 167–178, 1976.
[103] T. Sato, S. Wadaka, J. Yamamoto, and J. Ishii, "Imaging system using an intensity triple correlator," *Appl. Opt.*, vol. 17, no. 3, pp. 2047–2052, July 1978.
[104] J. D. Scargle, "Absolute value optimization to estimate phase properties of stochastic time series," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 140–143, Jan. 1977.
[105] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York, NY: Wiley, 1980.
[106] A. N. Shiryaev, "Some problems in the spectral theory of higher-order moments, I," *Theory Prob. Appl.*, vol. 5, pp. 265–284, 1960.
[107] ——, "On conditions for eogodicity of stationary processes in terms of higher order moments," *Theory Prob. Appl.*, vol. 8, pp. 436–439, 1963.
[108] Y. G. Sinai, "On higher order spectral measures of ergodic stationary processes," *Theory Prob. Appl.*, vol. 8, pp. 429–436, 1963.
[109] D. G. Stone, "Wavelet estimation," *Proc. IEEE*, vol. 72, pp. 1394–1402, Oct. 1984.
[110] T. Subba Rao and M. M. Gabr, " A test for linearity of stationary time series," *J. Time Series Analysis* (GB), vol. 1, no. 2, pp. 145–158, 1980.
[111] T. C. Sun, "A central limit theorem for non-linear functions of normal stationary processes," *J. Math. Mech.*, vol. 12, pp. 945–978, 1963.
[112] M. Tenhoopen and P. A. Zandt, "Second-order correlations and bispectra of composite rhythms," Medical Physics Inst. Utrecht, Rep. MR382940 (Utrecht, The Netherlands), p. 7.
[113] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, 1982.
[114] L. J. Tick, "The estimation of transfer functions of quadratic systems," *Technometrics*, vol. 3, pp. 563–567, 1961.
[115] P. V. Tryon, "The bispectrum and higher-order spectra: A bibliography," *Nat. Bur. Stand.*, Washington, DC, Rep. PB81-209322.
[116] J. K. Tugnait, "Identification of nonminimum phase linear stochastic systems," in *Proc. 23rd IEEE Conf. Decision and Control*, pp. 342–347 (Las Vegas, NV, Dec. 1984).
[117] ——, "Realization and reduction of S.I.S.O. nonminimum phase stochastic systems," in *Modeling and Application of Stochastic Processes*, V. B. Desai, Ed. Boston, MA: Kluwer, 1986, ch. 8, pp. 194–214.
[118] ——, "Identification of linear stochastic systems via second and fourth order cumulant matching," *IEEE Trans. Informat. Theory*, vol. IT-33, pp. 393–407, May 1987.
[119] T. W. Tukey, "An introduction to the measurement of spectra," in *Probability and Statistics*, U. Grenander, Ed. New York, NY: Wiley, 1959, pp. 300–330.
[120] C. W. Van Atta, "Bispectral measurements in turbulence computations," in *Proc. 6th Int. Conf. on Numerical Methods in Fluid Dynamics* (in *Lecture Notes in Physics*. New York, NY: Springer-Verlag, pp. 530–536, 1978).
[121] ——, "Inertial range bispectra in turbulence," *Phys. Fluids*, vol. 22, no. 8, pp. 1440–1442, 1979.
[122] J. W. Van Ness, "Asymptotic normality of bispectral estimates," *Ann. Math. Statist.*, vol. 37, pp. 1257–1272, 1966.
[123] G. P. Weigelt and B. Wirnitzer, "Image reconstruction by the speckle-masking method," *Opt. Lett.*, vol. 8, no. 7, pp. 389–391, July 1983.
[124] B. Wells, "Voiced/unvoiced decision based on the bispectrum," in *Proc. ICASSP'85*, pp. 1589–1592 (Tampa, FL, Mar. 1985).
[125] M. B. Zado and M. Caputo, "Spectral analysis and Q of the free oscillations of the earth," *Suppl. Nouvo Cimento*, vol. 6, pp. 67–81, 1968.
[126] *IEEE Trans. Acous., Speech, Signal Processing, Special Issue on Time Delay Estimation*, vol. ASSP-29, no. 3, June 1981.

Chrysostomos L. Nikias (Member, IEEE) received the diploma in electrical and mechanical engineering from the National Technical University of Athens, Athens, Greece, in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo, in 1980 and 1982, respectively.

From 1982 to 1985, he was on the faculty of the Department of Electrical and Systems Engineering, University of Connecticut, Storrs. Since 1985 he has been with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, where he is currently an Associate Professor and Director of the Communications and Digital Signal Processing (CDSP) Laboratory. He also consults for the Maryland Institute for Emergency Medical Services Systems (MIEMSS), Baltimore, MD; ALCOA, Pittsburgh, PA; and The Gordon Institute, Wakefield, MA. His research interests lie in the areas of digital signal processing with applications, detection, and estimation theory, as well as bioengineering. He has taught extensively short courses in the United States, Canada, and Europe, covering modern spectral analysis and array processing. He is the recipient of the University of Connecticut Faculty Summer Fellowship (1982). He is a member of the Conference Board of the IEEE Acoustics, Speech, and Signal Processing (ASSP) Society, and served as Associate Editor of the IEEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING during 1985–1987. He was the Co-Chairman of the Third ASSP Workshop on Spectrum Estimation and Modeling (1986).

Dr. Nikias is a member of Tau Beta Pi, the New York Academy of Sciences, the American Association for the Advancement of Science, and the Technical Chamber of Greece.

Mysore R. Raghuveer (Member, IEEE) received the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1981, and the Ph.D. degree from the University of Connecticut, Storrs, in 1984. His research at the University of Connecticut was in the areas of multidimensional and higher order spectrum estimation.

During 1982–1984, he was actively associated with studies conducted at the Shock Trauma unit of the Maryland Institute for Emergency Medical Services Systems, Baltimore, MD, on application of spectrum estimation to biomedical signals. Since January 1985, he has been with Communication Products Division of Advanced Micro Devices Inc., Austin, TX, where he is now a member of the technical staff for ISDN design. He is the organizer and chairman of the IEEE Central Texas joint chapter of the IEEE ASSP and Communication Societies.

Dr. Raghuveer is a member of Tau Beta Pi.