Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications

JERRY M. MENDEL, FELLOW, IEEE

During the past few years there has been an increasing interest in applying higher-order statistics to a wide range of signal processing and system theory problems. These statistics are very useful in problems where either non-Gaussianity, nonminimum phase, colored noise, or nonlinearities are important and must be accounted for. More than 200 papers have already been published. These papers contain both theoretical and algorithmic results. The purpose of the present tutorial paper is twofold, namely: 1) to collect what this author believes to be some of the most useful theoretical results in one place (they are presently scattered in many papers), thereby making them readily accessible to readers for the first time (derivations are provided in the Appendix for many of the results), and, 2) to describe the applications of higher-order statistics to the identification of (possibly) nonminimum phase channels from just noisy output measurements. More than 20 new methods are summarized for the latter.

I. Introduction

During recent years higher-order statistics (spectra) have begun to find wide applicability in many diverse fields; e.g., sonar, radar, plasma physics, biomedicine, seismic data processing, image reconstruction, harmonic retrieval, time-delay estimation, adaptive filtering, array processing, and blind equalization. For overviews, see [30] and [35]; (see, also, [57]). These statistics, known as cumulants, and their associated Fourier transforms, known as polyspectra, not only reveal amplitude information about a process, but also reveal phase information. This is important, because, as is well known, second-order statistics (i.e., correlation) are phase blind.

Cumulants, on the other hand, are blind to any kind of a Gaussian process, whereas correlation is not; hence, cumulant-based signal processing methods handle colored Gaussian measurement noise automatically, whereas correlation-based methods do not. Consequently, cumulant-based methods boost signal-to-noise ratio when signals are corrupted by Gaussian measurement noise.

Higher-order statistics are applicable when we are dealing with non-Gaussian (or, possibly, nonlinear) processes, and, many real-world applications are truly non-Gaussian. In the past, due to lack of analytical tools, we have been forced to treat such applications as though they were Gaussian. With the new results that are being developed and those that are described in this paper, it should be possible to reexamine every application and/or method that has ever made use of second-order statistics, using higher-order statistics, to see if better results can be obtained.

The development of cumulants and polyspectra has paralleled the development of traditional correlation and its associated spectrum. The early works applied these higher-order statistics to different types of data sets to infer new properties about the data. With the boom in interest in spectral analysis it was quite natural for this interest to spill over into polyspectral analysis. Just as spectral analysis has split into two camps, so has polyspectral analysis, namely: nonparametric polyspectral and parametric polyspectral methods. The former are subject to the same problems that plague nonparametric spectral methods, namely, high variances and low resolution. The latter first estimate the parameters of an underlying data-generating model and then use the model to compute the polyspectrum. These models are in the classes of moving average (MA), autoregressive (AR), or autoregressive moving average (ARMA) processes.

The biggest drawback to-date to the use of either nonparametric or parametric polyspectral methods is that they require longer data lengths than do correlation-based methods. Longer data lengths are needed in order to reduce the variance associated with estimating the higher-order statistics from real data using sample-averaging techniques (see Section XI).

Mendel [29] gives an overview of parametric polyspectral results obtained prior to 1987. His paper focuses on the activities of three groups (Northeastern University, University of Southern California, and Exxon) that were generating almost all of the parametric results prior to that date. A lot of the post 1982 work was motivated by the seminal paper of Lii and Rosenblatt [24], who proved that a nonminimum phase transfer function can be restored just from output measurements using any higher-order spectra. Mendel's 1988 update [30] is organized by the type of problem studied, because in the 1987–1988 period more people joined in this activity and some of the three groups split. The problems studied are: 1) identification of MA processes, 2) identification of AR processes, 3) identification of ARMA processes, 4) order determination, 5) calculation of cumulants, 6) computation of polyspectra, 7) extensions to multichannel and multidimensional systems, and 8) applications. Since 1988, much has happened,

Manuscript received May 4, 1990; revised November 29, 1990.

The author is with the Signal and Image Processing Institute, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-0781.

IEEE Log Number 9042429.

0018-9219/91/0300-0278$01.00 © 1991 IEEE

including a June 1989 Workshop on Higher-Order Spectral Analysis, which was attended by more than 125 people, and whose Proceedings contains 50 papers; a five paper special issue of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL in January, 1990; and, an eight paper mini-special issue of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING in July, 1990.

From time to time a new approach emerges on the scene that leads to a large number of new and important theoretical results that can hopefully help us solve a wide range of problems. Such is the case with the emergence of higher-order statistics. Unfortunately, most of these results are scattered in the literature, which makes it very difficult for the newcomer to see the forest from the trees. One purpose of this tutorial paper is to collect together in one place a compendium of new theoretical results which are associated with using higher-order statistics in signal processing and system theory. Most of our results are given for one-dimensional processes; some extensions to vector processes and multichannel systems are discussed in Section IX. Proofs of many of the results are given in the Appendix. It is assumed, however, that the reader is familiar with [35].

A second purpose of this paper is to demonstrate the utility of higher-order statistics to practical problems. Space does not permit us to survey all the applications to which higher-order statistics have been or are being applied. To give the reader a feeling for the breadth of the applications, we list the titles of the application sessions at the June 1989 Workshop: identification of nonminimum phase systems, sonar and radar applications, nonlinear system analysis, harmonic retrieval and detection, multidimensional signal processing, biomedical signal analysis, and array processing. We shall survey in depth some of the applications of higher-order statistics to identification of (possibly) nonminimum phase channels from just noisy output measurements.

II. DEFINITIONS

Let $\boldsymbol{v} = \text{col}\,(v_1, v_2, \cdots, v_k)$ and $\boldsymbol{x} = \text{col}\,(x_1, x_2, \cdots, x_k)$ where $(x_1, x_2, \cdots, x_k)$ denote a collection of random variables (the material in this section is taken, for the most part, from [45]). The kth-order cumulant of these random variables is defined [41] as the coefficient of $(v_1 v_2 \cdots v_k)$ in the Taylor series expansion (provided it exists) of the cumulant-generating function

$$K(\boldsymbol{v}) = \ln \boldsymbol{E}\{\exp(j\boldsymbol{v}'\boldsymbol{x})\}. \tag{1}$$

The kth-order cumulant is therefore defined in terms of its joint moments of orders up to k. See Appendix A, (A-1), for the explicit relationship between cumulants and moments.

For zero-mean real random variables, the second-, third-, and fourth-order cumulants are given by

$$\text{cum}\,(x_1, x_2) = \boldsymbol{E}\{x_1 x_2\} \tag{2a}$$

$$\text{cum}\,(x_1, x_2, x_3) = \boldsymbol{E}\{x_1 x_2 x_3\} \tag{2b}$$

$$\begin{aligned}\text{cum}\,(x_1, x_2, x_3, x_4) = {} & \boldsymbol{E}\{x_1 x_2 x_3 x_4\} - \boldsymbol{E}\{x_1 x_2\}\,\boldsymbol{E}\{x_3 x_4\} \\ & - \boldsymbol{E}\{x_1 x_3\}\,\boldsymbol{E}\{x_2 x_4\} \\ & - \boldsymbol{E}\{x_1 x_4\}\,\boldsymbol{E}\{x_2 x_3\}.\end{aligned} \tag{2c}$$

In the case of nonzero mean real random variables [23], one replaces x_i by $x_i - \boldsymbol{E}\{x_i\}$ in these formulas. The case of complex signals is treated in Section X.

Let $\{x(t)\}$ be a zero-mean kth-order stationary random process. The kth-order cumulant of this process, denoted $C_{k,x}(\tau_1, \tau_2, \cdots, \tau_{k-1})$, is defined as the joint kth-order cumulant of the random variables $x(t), x(t+\tau_1), \cdots, x(t+\tau_{k-1})$, i.e.,

$$\begin{aligned}& C_{k,x}(\tau_1, \tau_2, \cdots, \tau_{k-1}) \\ & \quad = \text{cum}\,(x(t), x(t+\tau_1), \cdots, x(t+\tau_{k-1})).\end{aligned} \tag{3a}$$

Because of stationarity, the kth-order cumulant is only a function of the $k-1$ lags $\tau_1, \tau_2, \cdots, \tau_{k-1}$. The $\tau_1 - \tau_2 - \cdots - \tau_{k-1}$ space constitutes the domain of support for $C_{k,x}(\tau_1, \tau_2, \cdots, \tau_{k-1})$. If $\{x(t)\}$ is nonstationary, then the kth-order cumulant depends explicitly on t as well as on $\tau_1, \tau_2, \cdots, \tau_{k-1}$, and, we use the notation $C_{k,x}(t; \tau_1, \tau_2, \cdots, \tau_{k-1})$ (e.g., see Section IX).

For a zero-mean stationary random process, and, for $k = 3, 4$, the kth-order cumulant of $\{x(t)\}$ can also be defined as

$$\begin{aligned}& C_{k,x}(\tau_1, \tau_2, \cdots, \tau_{k-1}) \\ & \quad = \boldsymbol{E}\{x(\tau_1) \cdots x(\tau_{k-1})\} - \boldsymbol{E}\{g(\tau_1) \cdots g(\tau_{k-1})\}\end{aligned} \tag{3b}$$

where $\{g(t)\}$ is a Gaussian random process with the same second-order statistics as $\{x(t)\}$. Cumulants, therefore, not only display the amount of higher-order correlation, but also provide a measure of the distance of the random process from Gaussianity. Clearly, if $x(t)$ is Gaussian then the cumulants are all zero; this is not only true for $k = 3$ and 4, but for all k.

The second-, third- and fourth-order cumulants of zero-mean $x(t)$, which follow from (2) and (3a), are

$$C_{2,x}(\tau) = \boldsymbol{E}\{x(t)\,x(t+\tau)\} \tag{4a}$$

$$C_{3,x}(\tau_1, \tau_2) = \boldsymbol{E}\{x(t)\,x(t+\tau_1)\,x(t+\tau_2)\} \tag{4b}$$

$$\begin{aligned}C_{4,x}(\tau_1, \tau_2, \tau_3) = {} & \boldsymbol{E}\{x(t)\,x(t+\tau_1)\,x(t+\tau_2)\,x(t+\tau_3)\} \\ & - C_{2,x}(\tau_1)\,C_{2,x}(\tau_2 - \tau_3) \\ & - C_{2,x}(\tau_2)\,C_{2,x}(\tau_3 - \tau_1) \\ & - C_{2,x}(\tau_3)\,C_{2,x}(\tau_1 - \tau_2).\end{aligned} \tag{4c}$$

Of course, the second-order cumulant $C_{2,x}(\tau)$ is just the autocorrelation of $x(t)$. We shall use the more familiar notation for autocorrelation, namely $r_x(\tau)$, interchangeably with $C_{2,x}(\tau)$.

A 1-D slice of the kth-order cumulant is obtained by freezing $(k-2)$ of its $k-1$ indexes. Many types of 1-D slices are possible, including radial, vertical, horizontal, diagonal, and offset-diagonal. A diagonal slice is obtained by setting $\tau_i = \tau$, $i = 1, 2, \cdots, k-1$. All these 1-D slices are very useful in applications of cumulants in signal processing.

A logical question to ask is "Why do we need fourth-order cumulants, i.e., aren't third-order cumulants good enough?" If a random process is symmetrically distributed, then its third-order cumulant equals zero; hence, for such a process we must use fourth-order cumulants. For example, Laplace, Uniform, Gaussian, and Bernoulli-Gaussian distributions are symmetric, whereas Exponential, Rayleigh and k-distributions are nonsymmetric. Additionally, some processes have extremely small third-order cumulants and much larger fourth-order cumulants; hence, for such processes we would also use the latter. Finally, in some specific applications (e.g., retrieval of harmonics and cubic phase coupling) third-order cumulants equal zero (see Section X), whereas fourth-order cumulants are nonzero.

Another logical question to ask is "Why not work with higher-order moments instead of cumulants?" This question has a number of answers, some quite mathematical and others more practical. Focusing on the latter, we give the following two reasons

for preferring to work with cumulants: 1) just as the covariance function of white noise is an impulse function and its spectrum is flat, the cumulants of (higher-order) white noise are multidimensional impulse functions and the polyspectra of this noise is multidimensionally flat; and 2) the cumulant of two statistically independent random processes equals the sum of the cumulants of the individual random processes [see (6e) below], whereas the same is not true for higher-order moments. This second property lets us work with the cumulant very easily as an operator.

Assuming that $C_{k,x}(\tau_1, \tau_2, \cdots, \tau_{k-1})$ is absolutely summable, the kth-order polyspectrum is defined as the $(k-1)$-dimensional discrete-time Fourier transform of the kth-order cumulant, i.e.,

$$S_{k,x}(\omega_1, \omega_2, \cdots, \omega_{k-1}) = \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{\infty} C_{k,x}(\tau_1, \tau_2, \cdots, \tau_{k-1}) \times \exp\left[-j \sum_{i=1}^{k-1} \omega_i \tau_i\right]. \quad (5)$$

The $\omega_1 - \omega_2 - \cdots - \omega_{k-1}$ space is the domain of support for $S_{k,x}(\omega_1, \omega_2, \cdots, \omega_{k-1})$. $S_{3,x}(\omega_1, \omega_2)$ is known as the bispectrum [in many papers the notation $B_x(\omega_1, \omega_2)$ is used to denote the bispectrum], whereas $S_{4,x}(\omega_1, \omega_2, \omega_3)$ is known as the trispectrum. For continuous-time signals, (5) is replaced by the multidimensional Fourier transform.

Many symmetries [35], [43] exist in the arguments of $C_{k,x}(\tau_1, \tau_2, \cdots, \tau_{k-1})$ and $S_{k,x}(\omega_1, \omega_2, \cdots, \omega_{k-1})$ which make their calculations manageable. For example, $C_{3,x}(\tau_1, \tau_2) = C_{3,x}(\tau_2, \tau_1) = C_{3,x}(-\tau_2, \tau_1 - \tau_2) = C_{3,x}(-\tau_1, \tau_2 - \tau_1) = C_{3,x}(\tau_2 - \tau_1, -\tau_1) = C_{3,x}(\tau_1 - \tau_2, -\tau_2)$. Using these five equations, we can divide the $\tau_1 - \tau_2$ plane into six regions. Knowing the cumulants in any one of these regions, we can calculate the cumulants in the other five regions using these equations. The principal region is the first-quadrant 45 degree sector, $0 < \tau_2 \leq \tau_1$. The symmetry relationships for cumulants do not hold in the nonstationary case. For many more details about symmetry relationships for the bispectrum, see [35], [43], and, for the trispectrum, see [7].

Following are some important properties of cumulants, which are used in theoretical developments [43] (see Appendix A, Section B, for proofs).

[CP1] If λ_i, $i = 1, \cdots, k$, are constants, and x_i, $i = 1, \cdots, k$, are random variables, then

$$\text{cum}(\lambda_1 x_1, \cdots, \lambda_k x_k) = \left(\prod_{i=1}^{k} \lambda_i\right) \text{cum}(x_1, \cdots, x_k). \quad (6a)$$

[CP2] Cumulants are symmetric in their arguments, i.e.,

$$\text{cum}(x_1, \cdots, x_k) = \text{cum}(x_{i_1}, \cdots, x_{i_k}) \quad (6b)$$

where $(i_1, \cdots, i_k)$ is a permutation of $(1, \cdots, k)$.

[CP3] Cumulants are additive in their arguments, i.e.,

$$\text{cum}(x_0 + y_0, z_1, \cdots, z_k) = \text{cum}(x_0, z_1, \cdots, z_k) + \text{cum}(y_0, z_1, \cdots, z_k). \quad (6c)$$

This means that cumulants of sums equals sums of cumulants (hence, the name "cumulant")

[CP4] If α is a constant, then

$$\text{cum}(\alpha + z_1, z_2, \cdots, z_k) = \text{cum}(z_1, \cdots, z_k) \quad (6d)$$

[CP5] If the random variables $\{x_i\}$ are independent of the random variables $\{y_i\}$, $i = 1, 2, \cdots, k$, then

$$\text{cum}(x_1 + y_1, \cdots, x_k + y_k) = \text{cum}(x_1, \cdots, x_k) + \text{cum}(y_1, \cdots, y_k). \quad (6e)$$

[CP6] If a subset of the k random variables $\{x_i\}$ is independent of the rest, then

$$\text{cum}(x_1, \cdots, x_k) = 0. \quad (6f)$$

Cumulants of an independent, identically distributed random sequence are delta functions (the same is not true for joint moments), i.e., if $w(t)$ is an i.i.d. process, then $C_{k,w}(\tau_1, \tau_2, \cdots, \tau_{k-1}) = \gamma_{k,w}\delta(\tau_1)\,\delta(\tau_2)\cdots\delta(\tau_{k-1})$, where $\gamma_{k,w}$ is the kth-order cumulant of the stationary random sequence $w(n)$.

Suppose $z(n) = y(n) + v(n)$ where $y(n)$ and $v(n)$ are independent; then from [CP5]

$$C_{k,z}(\tau_1, \tau_2, \cdots, \tau_{k-1}) = C_{k,y}(\tau_1, \tau_2, \cdots, \tau_{k-1}) + C_{k,v}(\tau_1 \tau_2, \cdots, \tau_{k-1}) \quad (7)$$

If $v(n)$ is Gaussian (colored or white) and $k \geq 3$ then $C_{k,z}(\tau_1, \tau_2, \cdots, \tau_{k-1}) = C_{k,y}(\tau_1, \tau_2, \cdots, \tau_{k-1})$, whereas $C_{2,z}(\tau) = C_{2,y}(\tau) + C_{2,v}(\tau)$. This makes the higher-order statistics more robust to additive measurement noise than correlation, even if that noise is colored. In essence, cumulants can draw non-Gaussian signals out of Gaussian noise, thereby boosting their signal-to-noise ratios.

III. Cumulant-Polyspectra Formulas: Linear Systems

A familiar starting point for many problems in signal processing and system theory is the single-input single-output (SISO) linear and time-invariant (LTI) model depicted in Fig. 1, in which

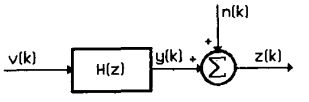


Fig. 1. Single-channel system.

$v(k)$ is white Gaussian noise with finite variance σ_v^2; $H(z)[h(k)]$ is causal and stable; $n(k)$ is also white Gaussian noise with variance σ_n^2; and, $v(k)$ and $n(k)$ are statistically independent. Letting $r(\cdot)$ and $S(\cdot)$ denote correlation and Fourier transform of correlation (i.e., spectrum), respectively, then it is well known, that (e.g., [38])

$$r_z(k) = r_y(k) + r_n(k) = \sigma_v^2 \sum_{i=0}^{\infty} h(i)\,h(i+k) + \sigma_n^2\delta(k) \quad (8)$$

$$S_z(\omega) = \sigma_v^2 |H(\omega)|^2 + \sigma_n^2 \quad (9)$$

$$r_{vz}(k) \triangleq E\{v(n)\,z(n+k)\} = \sigma_v^2 h(k). \quad (10)$$

From (9) we see that all phase information has been lost in the spectrum (or in the autocorrelation); hence, we say that correlation or spectra are phase blind.

A major generalization to (8) and (9) was established in 1955 by Bartlett [2] and in 1967 by Brillinger and Rosenblatt [5]. In this case, the system in Fig. 1 is assumed to be causal and exponentially stable, and $\{v(i)\}$ is assumed to be independent, iden-

tically distributed (i.i.d.), and non-Gaussian, i.e.,

$$C_{k,v}(\tau_1, \tau_2, \cdots, \tau_{k-1}) = \begin{cases} \gamma_{k,v} & \text{if } \tau_1 = \tau_2 = \cdots = \tau_{k-1} = 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where $\gamma_{k,v}$ denotes the kth-order cumulant of $v(i)$. Additive noise $n(k)$ is assumed to be Gaussian, but need not be white. Their generalizations to (8) and (9) are

$$C_{k,z}(\tau_1, \tau_2, \cdots, \tau_{k-1}) = \gamma_{k,v} \sum_{n=0}^{\infty} h(n)\, h(n + \tau_1) \cdots h(n + \tau_{k-1}) \quad (12)$$

and

$$S_{k,z}(\omega_1, \omega_2, \cdots, \omega_{k-1}) = \gamma_{k,v} H(\omega_1)\, H(\omega_2) \cdots H(\omega_{k-1})\, H\left(-\sum_{i=1}^{k-1} \omega_i\right). \quad (13)$$

Observe that when $k = 2$, (12) and (13) reduce to (8) [subject to the addition of $\sigma_n^2 \delta(k)$] and (9). Equations (12) and (13) have been the starting points for many nonparametric and parametric polyspectral techniques that have been developed during the past five years (e.g., [35] and [30]).

The generalization of (12) to the case of colored non-Gaussian input $v(i)$ is ([2] only considers the $k = 2, 3, 4$ cases; [5] provides results for all k)

$$C_{k,z}(\tau_1, \tau_2, \cdots, \tau_{k-1}) = \sum_{m_1} \sum_{m_2} \cdots \sum_{m_{k-1}} C_{k,v}(\tau_1 - m_1, \tau_2 - m_2, \cdots, \tau_{k-1} - m_{k-1})\, C_{k,h}(m_1, m_2, \cdots, m_{k-1}) \quad (12')$$

where

$$C_{k,h}(m_1, m_2, \cdots, m_{k-1}) = \sum_{k_1} h(k_1)\, h(k_1 + m_1) \cdots h(k_1 + m_{k-1}).$$

Observe that the right-hand side of (12') is a multidimensional convolution of $C_{k,v}(\tau_1, \tau_2, \cdots, \tau_{k-1})$ with the deterministic correlation function of $h(n)$, which we have denoted $C_{k,h}(m_1, m_2, \cdots, m_{k-1})$. Consequently, the generalization of (13) to the case of colored non-Gaussian input $v(i)$ is

$$S'_{k,z}(\omega_1, \omega_2, \cdots, \omega_{k-1}) = S_{k,v}(\omega_1, \omega_2, \cdots, \omega_{k-1})\, H(\omega_1)\, H(\omega_2) \cdots H(\omega_{k-1}) \cdot H\left(-\sum_{i=1}^{k-1} \omega_i\right). \quad (13')$$

Derivations of these results are given in Section C. of Appendix A.

Example: Suppose that $h(k)$ is the impulse response (IR) of a causal moving average (MA) system. Such a system has a finite IR, and is described by the following model.

$$y(k) = \sum_{i=0}^{q} b(i)\, v(k - i) \quad (14)$$

The MA parameters are $b(0), b(1), \cdots, b(q)$, where q is the order of the MA model, and $b(0)$ is usually assumed equal to unity [the scaling is absorbed into the statistics of $v(k)$]. It is well known that for this model $h(i) = b(i)$, $i = 0, 1, \cdots, q$; hence, when $\{v(l)\}$ is i.i.d., we find from (12) that

$$C_{3,y}(\tau_1, \tau_2) = \gamma_{3,v} \sum_{l=0}^{q} b(l)\, b(l + \tau_1)\, b(l + \tau_2). \quad (15)$$

An interesting question is "for which values of τ_1 and τ_2 is $C_{3,y}(\tau_1, \tau_2)$ nonzero?" The answer to this question is depicted in Fig. 2. The domain of support for $C_{3,y}(\tau_1 = m, \tau_2 = n)$ is the

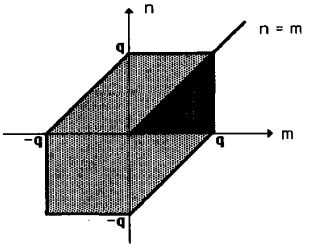


Fig. 2. The domain of support for $C_{3,y}(\tau_1, \tau_2)$ for an MA(q) system. The dark shaded triangular region is the principal region.

six-sided shaded region. This is due to the FIR nature of the MA system. The dark shaded triangular region in the first quadrant is the principal region. In the stationary case, we only have to determine third-order cumulant values in this region, $\boldsymbol{R}$, where

$$\boldsymbol{R} = \{m, n: 0 \le n \le m \le q\}. \quad (16)$$

Observe, from Fig. 2, that the third-order cumulant equals zero for either one or both of its arguments equal to $q + 1$. This suggests that it should be possible to determine the order of the MA model, q, by testing, in a statistical sense, the smallness of a third-order cumulant such as $C_{3,y}(q + 1, 0)$. See [17] to determine how to do this.

In the rest of this paper we always assume that $v(k)$ is i.i.d.

IV. Impulse Response Formulas

Equation (10) is by now a "classical" way for computing the impulse response (IR) of a SISO system; however, it requires that one have access to both the system's input and output in order to be able to compute the cross-correlation $r_{vz}(k)$. In many applications, one only has access to a noise-corrupted version of the system's output (e.g., blind equalization, reflection seismology). A natural question to ask is "Can the system's IR be determined just from output measurements?" Using output cumulants, the answer to this question is "yes."

Giannakis [11] was the first to show that the IR of a qth-order MA system can be calculated just from the system's output cumulants, as (see Appendix A, Section D, for a derivation)

$$h(k) = C_{3,y}(q, k)/C_{3,y}(-q, -q) = C_{3,y}(q, k)/C_{3,y}(q, 0), \quad k = 0, 1, \cdots, q. \quad (17)$$

Note that $C_{3,y}(-q, -q) = C_{3,y}(q, 0)$ follows from symmetry properties of the third-order cumulant. Equation (17) expresses $h(k)$ in terms of third-order cumulants. Comparable results for fourth-order cumulants are

$$h(k) = C_{4,y}(q, 0, k)/C_{4,y}(-q, -q, -q) = C_{4,y}(q, 0, k)/C_{4,y}(q, 0, 0), \quad k = 0, 1, \cdots, q. \quad (18)$$

The generalization of these results to arbitrary order cumulants is obvious (see, e.g., [49], eq. (13)).

Equation (17) is often referred to as the "$C(q, k)$ formula." Lohmann *et al.*, [26] and Lohmann and Wirnitzer [27] provide a nonrigorous derivation of the $C(q, k)$ formula for 1-D and 2-D continuous-time processes. If the system's input is not symmetrically distributed, so that $\gamma_{3,v} \neq 0$, we would use (17) to compute $h(k)$; however, if it is symmetrically distributed, so that $\gamma_{3,v} = 0$, but $\gamma_{4,v} \neq 0$, then we must use (18) (or comparable formulas that involve higher-order cumulants) to compute $h(k)$. Note that (17) and (18) only use 1-D slices of the output cumulant. Note, also, that they require exact knowledge of MA order q. They are interesting and important from a theoretical point of view, but they are impractical from an actual computation point of view. This is because, in practice, the output cumulant must be estimated, and (17) or (18) do not provide any filtering to reduce the effects of cumulant estimation errors.

A by-product of the derivation of the $C(q, k)$ formula are the following formulas for $\gamma_{3,v}$ and $\gamma_{4,v}$:

$$\gamma_{3,v} = C_{3,y}(0, 0) \Big/ \sum_{k=0}^{q} \left[C_{3,y}(q, k)/C_{3,y}(q, 0)\right]^3 \quad (19a)$$

and

$$\gamma_{4,v} = C_{4,y}(0, 0, 0) \Big/ \sum_{k=0}^{q} \left[C_{4,y}(q, 0, k)/C_{4,y}(q, 0, 0)\right]^4. \quad (19b)$$

If $h(k)$ is associated with an infinite impulse response (IIR) system, such as an autoregressive (AR) or autoregressive moving average (ARMA), then (17) and (18) do not apply. Giannakis [11] has shown that if $C_{3,y}(m_1, m_2) \approx 0$ for all $m_1 > M$ then

$$\begin{aligned} h(k) &\approx C_{3,y}(M, k)/C_{3,y}(-M, -M) \\ &= C_{3,y}(M, k)/C_{3,y}(M, 0), \qquad k = 0, 1, \cdots. \end{aligned} \quad (20)$$

A comparable result exists for $h(k)$ as a function of fourth-order (or, higher-order) cumulants.

Suppose $h(k)$ is associated with the following ARMA model:

$$\sum_{k=0}^{p} a(k)\, y(n - k) = \sum_{k=0}^{q} b(k)\, v(n - k). \quad (21)$$

The MA parameters are $b(0), b(1), \cdots, b(q)$, where q is the order of the MA part of the model; the AR parameters are $a(0)$, $a(1), \cdots, a(p)$, where p is the order of the AR part of the model; $a(0) = b(0) \triangleq 1$ [this fixes $h(0) = 1$] in order to fix the inherent scale ambiguity. Model order (p, q) is assumed to be known. (For results on determining model order (p, q) using cumulants and statistical tests, see [15], [17].) Impulse response $h(n)$ of this ARMA model satisfies the recursion

$$\sum_{k=0}^{p} a(k)\, h(n - k) = \sum_{k=0}^{q} b(k)\, \delta(n - k) = b(n), \quad n = 0, 1, \cdots, q. \quad (22)$$

Swami and Mendel [49] have developed the following formula for computing $h(n)$ under the assumption that the AR coefficients of the ARMA model are known (we discuss some methods for computing the AR coefficients in Section V) (see Appendix A, Section E, for a derivation):

$$h(n) = \left[\sum_{k=0}^{p} a(k)\, C_{m,y}(q - k, n, 0, \cdots, 0)\right] \Big/ \left[\sum_{k=0}^{p} a(k)\, C_{m,y}(q - k, 0, 0, \cdots, 0)\right], \quad n = 0, 1, \cdots. \quad (23)$$

If the ultimate goal is to compute the MA coefficients $b(1)$, $b(2), \cdots, b(q)$, then (22) reveals that we only need to know $h(0)$, $h(1), \cdots$, and $h(q)$, as well as the AR coefficients (remember, our system is assumed to be causal). Because we have assumed $a(0) = b(0) = 1$, it follows from (22) that $h(0) = 1$; hence, in reality we only need to know $h(1), h(2), \cdots, h(q)$. In addition to $C_{m,y}(q - k, 0, 0, \cdots, 0)$, (23) uses exactly q (horizontal) slices of the output cumulant to estimate the first q coefficients of the IR, namely $C_{m,y}(q - k, 1, 0, \cdots, 0)$, $C_{m,y}(q - k, 2, 0, \cdots, 0), \cdots$, and $C_{m,y}(q - k, q, 0, \cdots, 0)$. Consequently, (23) is usually referred to as the "q-slice algorithm."

Observe that when $p = 0$, in which case the ARMA model reduces to the MA model in (14), (23) reduces to Giannakis' $C(q, k)$ algorithm in (17) or (18). Just as the $C(q, k)$ algorithm is a theoretical result, and is not advocated for computational purposes, so is the q-slice algorithm. Using results found in [49] it is possible, however, to turn the q-slice algorithm into a practical computational tool. This is discussed further in Section XII-D.

V. AR Coefficients

Many ways exist for determining the AR coefficients of either the ARMA model in (21) or the following AR model

$$\sum_{k=0}^{p} a(k)\, y(n - k) = v(n) \quad (24)$$

where $a(0) = 1$. One of the most popular ways is to use the following correlation-based normal equations [3]:

$$\sum_{k=0}^{p} a(k)\, r_y(k - m) = 0, \quad \text{for } m > 0 \text{ (or } m > q \text{ in the ARMA case)}. \quad (25)$$

If AR order p is known then (25) is collected for p values of m, from which the AR coefficients $a(1), a(2), \cdots, a(p)$ can be determined.

The authors of [65], [10], [16], [45], [52], [58], and [59] have shown that the AR coefficients can also be determined using cumulants. Parzens [65] credits Akaike [66] for the "idea of extending" correlation-based normal equations to cumulants. For example, any 1-D cumulant slice satisfies the following AR recursion of maximum order p (see Appendix A, Section E, for a derivation).

$$\sum_{k=0}^{p} a(k)\, C_{m,y}(\tau - k, k_0, 0, \cdots, 0) = 0, \quad \text{for } \tau > 0 \text{ (or } \tau > q \text{ in the ARMA case)} \quad (26)$$

where k_0 is a parameter whose choice is discussed as follows.

Concatenating (26) for $\tau = 1, 2, \cdots, p + M$, (or $\tau = q + 1, q + 2, \cdots, q + p + M$ in the ARMA case) where $M \geq 0$, and k_0 is arbitrary, we obtain the cumulant-based normal equations

$$\boldsymbol{C}(k_0)\boldsymbol{a} = \boldsymbol{0} \quad (27)$$

where the entries in Toeplitz matrix $\boldsymbol{C}(k_0)$ are easily deduced, and $\boldsymbol{a} = \text{col}\,(1, a(1), \cdots, a(p-1), a(p))$. If $\boldsymbol{C}(k_0)$ has rank p then the corresponding 1-D slice (parameterized by k_0) of the mth-order cumulant is a full rank slice and the p AR coefficients can be solved for from (27). If $\boldsymbol{C}(k_0)$ does not have rank p then some of the AR coefficients cannot be solved for from (27), and those that can be solved for do not equal their true values. In essence, some poles of the AR model, and subsequently some of the AR coefficients, are invisible to (27).

Until recently, the issue of whether $\boldsymbol{C}(k_0)$ has rank p for an arbitrary k_0 was an open question (raised initially in [60] and [16], and subsequently in [45], [12], and [52]). In [45] and [52] it is shown by means of ARMA examples that 1) every 1-D cumulant slice need not be a full rank slice, and 2) a full rank cumulant slice may not exist. Swami and Mendel, as well as Giannakis, have shown that the AR coefficients of an ARMA model can be determined when $p+1$ slices of the mth-order cumulant are used. Furthermore, these cannot be arbitrary slices. Equation (26) must be concatenated for $\tau = 1, 2, \cdots, p+M$ and (at least) $k_0 = -p, \cdots, 0$ [or $\tau = q+1, q+2, \cdots, q+p+M$ and (at least) $k_0 = q-p, \cdots, q$ in the ARMA case] where $M \geq 0$. Tugnait [68] obtained a comparable result for the third-order cumulant case; however, he did not provide results for the general case of mth-order cumulants. Note that, to-date, all of the counter-examples and resulting fixes are for the problem of estimating the AR portion of an ARMA(p, q) system. The issue of whether or not more than one slice is needed to estimate the coefficients of a "truly" AR(p) system is an open question. To-date, no counter-example exists that demonstrates that it cannot be done. For reasons of "safety," people use $p+1$ slices even in the AR case.

VI. Relationships Between Second-Order and Higher-Order Statistics for Linear Systems

Let $C_{3,y}(\tau)$ denote the diagonal slice of the 3rd-order cumulant, i.e., $C_{3,y}(\tau) = C_{3,y}(\tau_1 = \tau, \tau_2 = \tau)$. Its z-transform is denoted $S_{3,y}(z)$. In [10] and [16] it is shown that (see Appendix A, Section G)

$$S_{3,y}(\omega) = (1/2\pi) \int_{-\pi}^{\pi} S_{3,y}(-\omega, \phi)\, d\phi \tag{28}$$

where $S_{3,y}(\omega, \phi)$ is the Bispectrum. $S_{3,y}(\omega)$ is referred to by Giannakis and Mendel as the "$1\frac{1}{2}$-D spectrum." It has a tomographic flavor to it.

Let

$$H_2(z) = H(z) * H(z) \tag{29}$$

where "*" denotes complex convolution in the z domain. Giannakis and Mendel have also shown that (see Appendix A, Section H, for a derivation) for the linear system in Fig. 1

$$H_2(z)\, S_y(z) = (\sigma_v^2/\gamma_{3,v})\, H(z)\, S_{3,y}(z). \tag{30}$$

This interesting equation links the usual spectrum of $y(k)$ to its $1\frac{1}{2}$-D spectrum. Its extension to higher-than third order polyspectra [10] is

$$H_{k-1}(z) S_y(z) = (\sigma_v^2/\gamma_{k,v}) H(z) S_{k,y}(z) \tag{31}$$

where

$$H_k(z) = H(z) * H(z) * \cdots * H(z) \cdot (k-1 \text{ complex convolutions}) \tag{32}$$

and $S_{k,y}(z)$ is the z transform of $C_{k,y}(\tau) = C_{k,y}(\tau_1 = \tau, \tau_2 = \tau, \cdots, \tau_{k-1} = \tau)$.

Variations of the results in (30) can be developed by working with $C_{3,y}(\tau_1 = \tau, \tau_2 = \tau + m)$ instead of $C_{3,y}(\tau_1 = \tau, \tau_2 = \tau)$ (e.g., [9], [63] and (80)), or by working with interrelationships between two 1-D cumulants slices [63].

Thus far (31) has only found widespread applicability in the MA case (e.g., [10], [16], [8], [9], and [39]; Friedlander and Porat refer to it as the "GM equation"), for which

$$H(z) = \sum_{k=0}^{q} b(k) z^{-k}, \qquad b(0) = 1. \tag{33}$$

In this case, the time-domain version of (30) is (see Appendix A, Section H)

$$\sum_{k=0}^{q} b^2(k)\, r_y(m-k) = (\sigma_v^2/\gamma_{3,v}) \sum_{k=0}^{q} b(k)\, C_{3,y}(m-k, m-k) \tag{34}$$

where $-q \leq m \leq 2q$ (work this out from Fig. 2). The counterpart of this equation for fourth-order cumulants is

$$\sum_{k=0}^{q} b^3(k)\, r_y(m-k) = (\sigma_v^2/\gamma_{4,v}) \sum_{k=0}^{q} b(k)\, C_{4,y}(m-k, m-k, m-k) \tag{35}$$

where $-q \leq m \leq 2q$. These formulas, especially (34), have been used to estimate the MA coefficients, $b(1), b(2), \cdots, b(q)$, using least-squares or adaptive algorithms. Observe, in (34) for example, that both $b(k)$ and $b^2(k)$ appear. After $b(k)$ and $b^2(k)$ are computed by concatenating (34) and using least squares (this is suboptimal, because it treats $b(k)$ and $b^2(k)$ as independent parameters, which, of course, they are not), the estimates of $b(k)$ and $b^2(k)$ are combined to provide a final estimate of $b(k)$.

It is possible (A. Swami, private communication) that the rank of the matrix in the concatenated (34) will not equal $2q+1$; hence, more than the diagonal-slice GM equation may have to be used to extract the $2q+1$ unknowns $b(1), b(2), \cdots, b(q), b^2(1), b^2(2), \cdots, b^2(q), \sigma_v^2/\gamma_{3,v}$. Exactly how many slices and which slices are needed is an open question.

When measurement noise is present, in which case $r_y(m-k) = r_z(m-k) - \sigma_n^2 \delta(m-k)$, the range on m cannot include the values $0, 1, \cdots, q$, (recall that k ranges from 0 to q) or else we would need to know σ_n^2. In this case, (34) becomes an underdetermined system of equations [63], i.e., it is a system of $2q$ equations in $2q+1$ unknowns. See the discussions after (80) for a way to handle this important case.

The use of (30) or (31) in other situations remains to be explored.

VII. Double $C(q, k)$ Formulas for Extracting ARMA Coefficients

Giannakis and Swami [18], [19] have developed a method for computing AR coefficients that is analogous to Giannakis' $C(q, k)$ formula for computing MA coefficients. They have also developed a method for computing ARMA coefficients by a double application of the $C(q, k)$ formula.

If we begin with the z transform version of (13), for $k = 3$, we find that

$$S_{3,y}(z_1, z_2) = \gamma_{3,v} H(z_1) H(z_2) H(z_1^{-1}z_2^{-1})$$
$$= \gamma_{3,v} B(z_1) B(z_2) B(z_1^{-1}z_2^{-1}) / A(z_1) A(z_2) A(z_1^{-1}z_2^{-1}) \quad (36)$$

where

$$B(z) = 1 + \sum_{i=1}^{q} b(i) z^{-i} \quad (37)$$

and

$$A(z) = 1 + \sum_{i=1}^{p} a(i) z^{-i}. \quad (38)$$

Let $D(z_1, z_2) = A(z_1) A(z_2) A(z_1^{-1}z_2^{-1})$ and multiply (36) by $D(z_1, z_2)$ to obtain the following time-domain recursive equation

$$\sum_{i=0}^{p} \sum_{j=0}^{p} d(i, j) C_{3,y}(m - i, n - j) = 0 \quad (39)$$

for (m, n) outside of the Fig. 2 six-sided shaded region (i.e., the time-domain version of $B(z_1) B(z_2) B(z_1^{-1}z_2^{-1})$ is nonzero only in the six-sided shaded region), where

$$d(i, j) = \sum_{k=0}^{p} a(k) a(k + i) a(k + j). \quad (40)$$

Observe that $d(i, j)$ in (40) looks exactly like the formula for the third-order cumulant of an MA process of order p (referring to Fig. 2, observe that $d(i, j)$ is nonzero only in the six-sided shaded region on which q is replaced by p). If we consider (39) for the case when $m = n$ (using the symmetry properties of $d(i, j)$ (see Fig. 2) and $C_{3,y}(s, t)$) then it can be rewritten as

$$\sum_{i=1}^{p} C_{3,y}(m - i, m - i)[d(i, i)/d(0, 0)] + 2 \sum_{i=1}^{p} \sum_{j=0}^{i-1} C_{3,y}(m - i, m - j)[d(i, j)/d(0, 0)] = -C_{3,y}(m, m) \quad (39')$$

for (m, m) outside of the Fig. 2 six-sided shaded region. The number of variables $[d(i, j)/d(0, 0)]$ is $p + (1 + 2 + \cdots + p) = p(p + 3)/2$.

Concatenating (39') for $m = p + 1, \cdots, p(p + 5)/2$ leads to $p(p + 5)/2 - p = p(p + 3)/2$ linear equations in the same number of unknowns. This system of equations can be expressed as $\bar{\boldsymbol{C}}\boldsymbol{d} = \bar{\boldsymbol{c}}$, where $\bar{\boldsymbol{C}}$ and $\bar{\boldsymbol{c}}$ are defined in an obvious manner, and $\boldsymbol{d}$ is the vector of lexicographically ordered $[d(i, j)/d(0, 0)]$'s. It is possible to compute $d(i, j)/d(0, 0)$ from this equation, for i, j belonging in the triangular nonredundant region enclosed by the lines $i = p$, $i = j$, and $j = 0$ (see Fig. 2). Note also that more than $p(p + 3)/2$ of the equations (39') could have been used [e.g., Giannakis and Swami [19] do not just use $m = n$]. Then $\bar{\boldsymbol{C}}\boldsymbol{d} = \bar{\boldsymbol{c}}$ would be an overdetermined system and $\boldsymbol{d}$ could be obtained using least squares.

Once the $d(i, j)$ coefficients have been determined, it is possible to extract the AR coefficients using Giannakis' $C(q, k)$ method, where q is set equal to p. To do this, (40) must be reexpressed as

$$d(i, j)/d(0, 0) = \sum_{k=0}^{p} a'(k) a'(k + i) a'(k + j) \quad (40')$$

where $a'(k) = a(k)/[d(0, 0)]^{1/3}$. After we obtain the $a'(k)$'s for $k = 0, 1, \cdots, p$, the $a(k)$'s for $k = 1, 2, \cdots, p$ can be obtained from the fact that $a(k) = a'(k)/a'(0)$ [since $a(0) = 1$, $d(0, 0)^{1/3} = a'(0)$].

This method is applicable to causal or noncausal AR models. Of course, any other cumulant-based MA parameter estimation technique could also be used to compute the scaled AR coefficients from (40').

Once the AR coefficients have been determined, we can multiply (36) by $A(z_1) A(z_2) A(z_1^{-1}z_2^{-1})$ and take inverse z transforms to obtain

$$\sum_{i_1=0}^{p} \sum_{i_2=0}^{p} \sum_{i_3=0}^{p} a(i_1) a(i_2) a(i_3) \cdot C_{3,y}(m + i_3 - i_2, n + i_3 - i_1) = b(m, n) \quad (41)$$

where

$$b(m, n) = \gamma_{3,v} \sum_{i=0}^{q} b(i) b(i + m) b(i + n). \quad (42)$$

After computing $b(m, n)$ using (41), we can use the $C(q, k)$ formula to determine the MA coefficients [i.e., the $b(j)$ from $b(j) = b(q, j)/b(q, 0)$] because $b(m, n)$ in (42) has the form of the third-order cumulant of a MA system; or, we can use any other cumulant-based MA parameter estimation technique.

To summarize, the double $C(q, k)$ algorithm is 1) estimate a sufficient number of output third-order cumulant values necessary to accomplish steps 2 and 4, 2) solve (39') for $d(i, j)/d(0, 0)$, $0 \le i \le j \le p$, 3) use the $C(q, k)$ algorithm (with q replaced by p) (or any other cumulant-based MA parameter estimation algorithm) to solve for the AR coefficients from (40') and the fact that $a(k) = a'(k)/a'(0)$, 4) compute $b(m, n)$ from (41) for $0 \le m \le n \le q$, and 5) use the $C(q, k)$ algorithm (or any other cumulant-based MA parameter estimation algorithm) to solve for the MA coefficients from (42). This algorithm is applicable to causal or noncausal ARMA processes. Its practical application requires a modified formulation so that steps 2) and 5) can be performed using least squares, and an accurate determination of AR order p.

The fact that this algorithm is applicable to noncausal models is important, because no cumulant-based technique for determining AR coefficients is guaranteed to lead to a stable AR model. In the double $C(q, k)$ method, unstable poles can be associated with the anticausal portion of the noncausal model. In this way model stability has been made a nonissue.

VIII. Bicepstral Formulas

The complex cepstrum is widely known in digital signal processing circles (e.g, [36], ch. 12). One starts with the transfer function $H(z)$, takes its logarithm, $\hat{H}(z) = \log H(z)$, and then takes the inverse z transform of $\hat{H}(z)$ to obtain the complex cepstrum $\hat{h}(n)$. When $H(z)$ is decomposed into the product of its minimum and maximum delay components, as $H_{\min}(z) H_{\max}(z)$, so that $h(n) = h_{\min}(n) * h_{\max}(n)$, then it is well known that $h_{\min}(n)$ and $h_{\max}(n)$ can be computed recursively in terms of $\hat{h}(n)$ (see (A-28) and (A-29) in Appendix A). In fact, only a finite number of elements from the infinitely long cepstral sequence are actually needed to recursively compute $h(n)$ (as long as $H(z)$ has no zeros or poles on the unit circle).

Nikias and Pan [34] and Pan and Nikias [37] have shown how similar ideas apply to the bispectrum. Beginning with the z transform of (13) for $k = 3$, namely

$$S_{3,y}(z_1, z_2) = \gamma_{3,y} H(z_1)\, H(z_2)\, H(z_1^{-1} z_2^{-1}) \qquad (43)$$

they take the logarithm of $S_{3,y}(z_1, z_2)$, $\hat{S}_{3,y}(z_1, z_2) = \log S_{3,y}(z_1, z_2)$, and then take the inverse z transform of $\hat{S}_{3,y}(z_1, z_2)$ to obtain the complex bicepstrum $\hat{S}_{3,y}(m, n)$. (Our notation is quite different from the notation used by Pan and Nikias, because we are trying to keep our notation consistent with the just described complex cepstrum notation.) Just as the complex cepstrum can be used to reconstruct $h(n)$, so can the complex bicepstrum.

The new theoretical result that is associated with the complex bicepstrum is the following bicepstrum-cumulant equation that links the complex bicepstrum with the third-order output cumulants (see Appendix A, Section I, for a derivation)

$$\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} k\hat{S}_{3,y}(k, l)\, C_{3,y}(m - k, n - l) = mC_{3,y}(m, n). \qquad (44)$$

Note that this result is true for any non-Gaussian process $\{y(i)\}$; i.e., it is not limited to processes that are derived from linear systems.

For (44) to be a practical result we must somehow truncate the double-infinite limits in (44). Pan and Nikias do an analysis of $\hat{S}_{3,y}(k, l)$ for an ARMA model. They show that $\hat{S}_{3,y}(k, l)$ has nonzero values only at: $k = l = 0$, integer values along the k and l axes, and at the intersection of these values along the 45-degree line $k = l$ (Fig. 3). Substituting the nonzero values of $\hat{S}_{3,y}(k, l)$

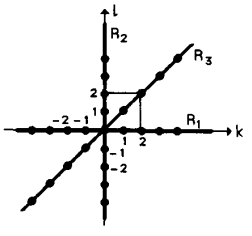


Fig. 3. The support of the complex bicepstrum is the union of the points labeled along R_1, R_2, and R_3.

into (44) (which can be expressed in terms of $\hat{h}(k)$, using (43); see, also, (A-35)), they show that (44) reduces to the following complex cepstrum-cumulant equation (see Appendix A, Section I, for a derivation)

$$\sum_{k=1}^{\infty} \left\{ k\hat{h}(k)\left[C_{3,y}(m - k, n) - C_{3,y}(m + k, n + k)\right] + k\hat{h}(-k)\left[C_{3,y}(m - k, n - k) - C_{3,y}(m + k, n)\right]\right\} = mC_{3,y}(m, n). \qquad (45)$$

Assuming that $k\hat{h}(k) \mapsto 0$ for $k > p^*$ and $k\hat{h}(-k) \mapsto 0$ for $k > q^*$, then (45) simplifies to

$$\sum_{k=1}^{p^*} \left\{ k\hat{h}(k)\left[C_{3,y}(m - k, n) - C_{3,y}(m + k, n + k)\right]\right\} + \sum_{j=1}^{q^*} \left\{ j\hat{h}(-j)\left[C_{3,y}(m - j, n - j) - C_{3,y}(m + j, n)\right]\right\} = mC_{3,y}(m, n). \qquad (46)$$

By concatenating (46) for a sufficiently large number of m and n, it is possible to solve for the p^* values of $\hat{h}(k)$ and the q^* values of $\hat{h}(-k)$. From these values it is then possible to reconstruct $h_{\min}(k)$ and $h_{\max}(k)$, using (A-28) and (A-29), after which $h(k)$ is computed as $h(k) = h_{\min}(k) * h_{\max}(k)$.

The derivation of (46) relies on the bicepstral factorization given in (43). Tekalp and Erdem [56] have studied the existence of such a factorization. Their results are for the more general polyspectral factorization given in (13), and have even been given by them for 2-D as well as 1-D systems. We shall state their 1-D results for the bispectral factorization, in our own words.

Recall that we just stated that "$\hat{S}_{3,y}(k, l)$ has nonzero values only at: $k = l = 0$, integer values along the k and l axes, and at the intersection of these values along the 45-degree line $k = l$." Tekalp and Erdem prove that there exists a unique (up to a time shift and scaling factor) stable LTI system $H(z)$ such that the bispectral factorization in (43) occurs, if and only if $\hat{S}_{3,y}(k, l)$ has nonzero values only at the k and l values just stated. They also show, by means of an example, that if $\hat{S}_{3,y}(k, l)$ is nonzero off of these lines then a factorization of $\hat{S}_{3,y}(z_1, z_2)$ as in (43) may not be possible. It is very interesting that the bicepstral domain leads to a nice resolution of such an important existence question. Whether this question can be resolved in another domain remains an open question. Unfortunately, Tekalp and Erdem's results do not assume noisy measurements; hence, when measurement noise is present and sample averages of cumulants are used, so that $\hat{S}_{3,y}(k, l)$ almost certainly will be nonzero off the three lines, some sort of statistical tests need to be developed in order to test whether those values are close enough to zero to be treated as zero.

IX. Multichannel Formulas

In order to extend any of the preceding results to the multichannel case we must first define the cumulant of a vector. The material in this section is taken from Swami and Mendel [53].

If $\boldsymbol{y}(t)$ is a vector process of dimension p, i.e., $\boldsymbol{y}(t) = \text{col}\,[y_1(t), y_2(t), \cdots, y_p(t)]$, we have two choices. We could let $\boldsymbol{v}_i = \text{col}\,[v_{i1}, \cdots, v_{ip}]$, $i = 1, \cdots, k$, and $\boldsymbol{y} = [\boldsymbol{y}'(t), \boldsymbol{y}'(t + t_1), \cdots, \boldsymbol{y}'(t + t_{k-1})]$ in (1) (note that now $\boldsymbol{v}$ and $\boldsymbol{y}$ in (1) are vectors of dimension pk). Alternatively, we could define the cross-cumulants of the elements of the vectors and then gather them into a single vector or a multidimensional array, e.g.,

$$C_{y_i, y_j, y_k}(t; t_1, t_2) = E\left\{ y_i(t)\, y_j(t + t_1)\, y_k(t + t_2)\right\} \qquad (47)$$

which is the third-order cross-cumulant of $y_i(t)$, $y_j(t + t_1)$, and $y_k(t + t_2)$. Instead of collecting all of the cross-cumulants into multidimensional arrays, we find it more convenient to represent the kth-order cumulant as a p^k-element vector.

Notation: If $\boldsymbol{C}$ is a p^k-element vector (in which p refers to the dimension of $\boldsymbol{y}(t)$ and k refers to the cumulant order), we will let $\boldsymbol{C}[i_1, i_2, \cdots, i_k]$, $i_1, \cdots, i_k = 1, \cdots, p$, denote its $[(i_1 - 1)p^{k-1} + (i_2 - 1)p^{k-2} + \cdots + i_k]$th element. The p^k-element vector is treated as a k-dimensional array.

Example: Suppose $p = 2$ and $k = 2$; then, we create a $2^2 \times 1 = 4 \times 1$ second-order cumulant vector $\boldsymbol{C}$ (instead of the usual 2×2 covariance matrix), where

$$C(1, 1) = \left[(1 - 1)2^{2-1} + 1\right] \text{ element of } \boldsymbol{C} = \text{first element of } \boldsymbol{C}$$

$$C(1, 2) = \left[(1 - 1)2^{2-1} + 2\right] \text{ element of } \boldsymbol{C} = \text{second element of } \boldsymbol{C}$$

$$C(2, 1) = \left[(2 - 1)2^{2-1} + 1\right] \text{ element of } \boldsymbol{C} = \text{third element of } \boldsymbol{C}$$

$$C(2, 2) = [(2 - 1)2^{2-1} + 2] \text{ element of } \boldsymbol{C}$$

$$= \text{fourth element of } \boldsymbol{C}.$$

Hence, $\boldsymbol{C} = \text{col}\,[C(1, 1), C(1, 2), C(2, 1), C(2, 2)]$. If $p = 2$ and $k = 3$ then $\boldsymbol{C} = \text{col}\,[C(1, 1, 1), C(1, 1, 2), C(1, 2, 1), C(1, 2, 2), C(2, 1, 1), C(2, 1, 2), C(2, 2, 1), C(2, 2, 2)]$.

Definition: The cumulant of the p-element random vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$, denoted by cum $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k)$ is the p^k-element vector whose $[i_1, i_2, \cdots, i_k]$th element is given by cum $(x_{1,i_1}, x_{2,i_2}, \cdots, x_{k,i_k})$, $i_1, \cdots, i_k = 1, \cdots, p$. In particular, the kth-order cumulant of the vector process $\boldsymbol{y}(n)$, i.e., cum $(\boldsymbol{y}(n), \boldsymbol{y}(n + \tau_1), \cdots, \boldsymbol{y}(n + \tau_{k-1}))$ denoted by $\boldsymbol{C}_{k,y}(n; \tau_1, \cdots, \tau_{k-1})$, has cum $(y_{i_0}(n), y_{i_1}(n + \tau_1), \cdots, y_{i_{k-1}}(n + \tau_{k-1}))$ as its $[i_0, i_1, \cdots, i_{k-1}]$th element, $i_0, \cdots, i_{k-1} = 1, \cdots, p$. This definition is easily generalized to vectors of different dimensions.

The motivation for representing cumulants as p^k-element vectors, rather than a k-dimensional array is twofold; first, it enables us to use the usual algebra of vectors and matrices for computational purposes; second, it allows us to exploit the algebra of Kronecker products to obtain rather simple-looking expressions for the cumulants of vector processes.

The Kronecker product approach handles the multiinput multioutput (MIMO) case easily and leads to rather nice looking formulae. We will let $\otimes$ denote the Kronecker product operator. See [4] and Appendix A (Section J) for background on Kronecker products.

Swami and Mendel [50], [53] have shown that the kth-order ($k = 2, 3, 4$) cumulants of a zero-mean p-element (in general) nonstationary vector process $\boldsymbol{y}(n)$, are given by the p^k vectors, $\boldsymbol{C}_{k,y}$, as (see Appendix A, Section K)

$$\boldsymbol{C}_{2,y}(t; \tau) = \boldsymbol{E}\{\boldsymbol{y}(t) \otimes \boldsymbol{y}(t + \tau)\} \tag{48}$$

$$\boldsymbol{C}_{3,y}(t; t_1, t_2) = \boldsymbol{E}\{\boldsymbol{y}(t) \otimes \boldsymbol{y}(t + t_1) \otimes \boldsymbol{y}(t + t_2)\} \tag{49}$$

$$\begin{aligned}\boldsymbol{C}_{4,y}(t; t_1, t_2, t_3) &= \boldsymbol{E}\{\boldsymbol{y}(t) \otimes \boldsymbol{y}(t + t_1) \otimes \boldsymbol{y}(t + t_2) \otimes \boldsymbol{y}(t + t_3)\} \\ &\quad - \boldsymbol{E}\{\boldsymbol{y}(t) \otimes \boldsymbol{y}(t + t_1)\} \otimes \boldsymbol{E}\{\boldsymbol{y}(t + t_2) \otimes \boldsymbol{y}(t + t_3)\} \\ &\quad - \boldsymbol{P}_p' \boldsymbol{E}\{\boldsymbol{y}(t) \otimes \boldsymbol{y}(t + t_2)\} \\ &\quad \otimes \boldsymbol{E}\{\boldsymbol{y}(t + t_3) \otimes \boldsymbol{y}(t + t_1)\} \\ &\quad - \boldsymbol{P}_p \boldsymbol{E}\{\boldsymbol{y}(t) \otimes \boldsymbol{y}(t + t_3)\} \\ &\quad \otimes \boldsymbol{E}\{\boldsymbol{y}(t + t_1) \otimes \boldsymbol{y}(t + t_2)\}\end{aligned} \tag{50}$$

where $\boldsymbol{P}_p$ is the $p^4 \times p^4$ permutation matrix $\boldsymbol{I} \otimes \boldsymbol{U}_{p^2 \times p}$ in which the $p^3 \times p^3$ permutation matrix $\boldsymbol{U}_{p^2 \times p}$ has unity entries in elements $[(i - 1)p + k, (k - 1)p^2 + i]$, $i = 1, \cdots, p^2$ and $k = 1, \cdots, p$, and zeros elsewhere. For a formal definition of $\boldsymbol{U}_{p^2 \times p}$, see (104) in [53].

Equations (48), (49), and (50) generalize (4a), (4b), and (4c), respectively, from scalar to vector processes. Equation (50) is complicated by its last three terms, each scalar component of which must look like the last three terms of (4c).

The cumulants of random vectors satisfy properties analogous to those summarized in Section II for scalar processes (see [50], [53]), e.g., if $\boldsymbol{\Lambda}_i$, $i = 1, \cdots, k$ are constant $r \times p$ matrices, and $\boldsymbol{x}_i$, $i = 1, \cdots, k$ are p-element random vectors, then (see Appendix A, Section L, for a derivation)

$$\text{[CP1]} \quad \text{cum}\,(\boldsymbol{\Lambda}_1 \boldsymbol{x}_1, \cdots, \boldsymbol{\Lambda}_k \boldsymbol{x}_k) = [\boldsymbol{\Lambda}_1 \otimes \cdots \otimes \boldsymbol{\Lambda}_k]\,\text{cum}\,(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k). \tag{51}$$

Consider a vector process $\boldsymbol{y}(n)$ to be the output of a linear dynamical system, one whose IR matrix is $\boldsymbol{H}(n, k)$, and input is $\boldsymbol{v}(k)$, i.e.,

$$\boldsymbol{y}(n) = \sum_{k=-\infty}^{\infty} \boldsymbol{H}(n, k)\, \boldsymbol{v}(k) \tag{52}$$

where $\boldsymbol{y}(n) \in R^{n_y}$, $\boldsymbol{v}(n) \in R^{n_v}$, and $\boldsymbol{H}(n, k) \in R^{n_y} \times R^{n_v}$. Additionally, we assume that $\boldsymbol{v}(n)$ is independent of $\boldsymbol{v}(m)$, $n \neq m$; hence, its cumulants are multidimensional Kronecker delta functions, i.e.,

$$\begin{aligned}\boldsymbol{C}_{k,v}(n; \tau_1, \cdots, \tau_{k-1}) &= \text{cum}\,(\boldsymbol{v}(n), \boldsymbol{v}(n + \tau_1), \cdots, \boldsymbol{v}(n + \tau_{k-1})) \\ &= \boldsymbol{\Gamma}_{kv}(n)\, \delta(\tau_1) \cdots \delta(\tau_{k-1})\end{aligned} \tag{53}$$

where the input cumulant $\boldsymbol{\Gamma}_{kv}$ is an n_v^k-element vector. We also assume that the IR matrices $\boldsymbol{H}(n, k)$ are absolutely summable so that the output cumulants are well-defined. Then the counterpart of (12) is [50], [53] (see Appendix A, Section M, for a derivation),

$$\begin{aligned}\boldsymbol{C}_{k,y}(n; \tau_1, \cdots, \tau_{k-1}) &= \sum_{i=-\infty}^{\infty} [\boldsymbol{H}(n, i) \otimes \boldsymbol{H}(n + \tau_1, i) \\ &\quad \otimes \cdots \otimes \boldsymbol{H}(n + \tau_{k-1}, i)]\, \boldsymbol{\Gamma}_{kv}(i).\end{aligned} \tag{54}$$

Observe that the single-input single-output (SISO) result given in (12) is a special case of (54) (in the SISO case $\otimes$ becomes multiplication; in the time-invariant case, $\boldsymbol{H}(l, m) = \boldsymbol{H}(l - m)$). Use of Kronecker products has led to a beautiful generalization of (12).

Next, we assume that our MIMO system can be described by the following state-space model (SSM):

$$\boldsymbol{x}(n + 1) = \boldsymbol{\Phi}(n)\, \boldsymbol{x}(n) + \boldsymbol{B}(n)\, \boldsymbol{v}(n) \tag{55}$$

$$\boldsymbol{y}(n) = \boldsymbol{\Psi}(n)\, \boldsymbol{x}(n) + \boldsymbol{w}(n) \tag{56}$$

where $\boldsymbol{x}(n) \in R^{n_x}$, $\boldsymbol{v}(n) \in R^{n_v}$ and $\boldsymbol{w}(n) \in R^{n_y}$. Input process $\boldsymbol{v}(n)$ is zero-mean and non-Gaussian; $\boldsymbol{v}(n)$ is independent of $\boldsymbol{v}(m)$, for $n \neq m$, but the components of $\boldsymbol{v}(n)$ are not mutually independent; and, all the relevant cumulants of $\boldsymbol{v}(n)$ are finite. Measurement noise process $\boldsymbol{w}(n)$ is zero-mean and independent of $\boldsymbol{v}(n)$. The system is causal and stable. Matrices $\boldsymbol{\Phi}(n)$, $\boldsymbol{B}(n)$, and $\boldsymbol{\Psi}(n)$ and the noise statistics are assumed known. Finally, we assume that $\boldsymbol{E}\{\boldsymbol{x}(0)\} = \boldsymbol{0}$.

Swami and Mendel [50], [53] have obtained closed-form, lag-recursive, time-recursive, and simultaneous lag- and time-recursive expressions for $\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1})$, the Kronecker state cumulant vector (KSCV) of the SSM in (55). For example, they show that the KSCV can be expressed for all τ_l, $l = 1, \cdots, k - 1$, in terms of its zero-lag values, as (see Appendix A, Section N, for a derivation)

$$\begin{aligned}\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1}) &= [\boldsymbol{A}(n, n - m) \otimes \boldsymbol{A}(n + \tau_1, n - m) \\ &\quad \otimes \cdots \otimes \boldsymbol{A}(n + \tau_{k-1}, n - m)] \\ &\quad \cdot \boldsymbol{C}_{k,x}(n - m; 0, \cdots, 0)\end{aligned} \tag{57}$$

where (see discussion associated with (A-54)) $m = -\min(0, \tau_1, \cdots, \tau_{k-1})$, and

$$\boldsymbol{A}(n+i, n) = \boldsymbol{\Phi}(n+i-1)\,\boldsymbol{\Phi}(n+i-2)\cdots\boldsymbol{\Phi}(n). \tag{58}$$

This result handles the entire $(k-1)$ dimensional space, $-\infty < \tau_l < \infty$, $l = 1, \cdots, k-1$. For causal systems $\boldsymbol{C}_{k,x}(l; 0, \cdots, 0) = 0$ for $l < 0$. Observe, from (57) and the formula for m, that cumulants at positive lags at time n are expressed in terms of the zero-lag values at time n, $\boldsymbol{C}_{k,x}(n; 0, \cdots, 0)$, because for such lags $m = 0$; however, cumulants elsewhere at time n are expressed in terms of the zero-lag values at previous times $n - m$, $m > 0$.

It is well known that the mean of the state vector, $\boldsymbol{m}_x(n)$, and its covariance matrix, $\boldsymbol{P}_x(n)$, can be computed recursively in temporal variable n. For example [1]

$$\boldsymbol{P}_x(n+1) = \boldsymbol{\Phi}(n)\,\boldsymbol{P}_x(n)\,\boldsymbol{\Phi}'(n) + \boldsymbol{B}(n)\,\boldsymbol{Q}(n)\,\boldsymbol{B}'(n), \quad n \geq 0 \tag{59}$$

where $\boldsymbol{Q}(n)$ is the covariance matrix of input $\boldsymbol{v}(n)$. Equation (59) is initialized by $\boldsymbol{P}_x(0)$.

Analogous to (59), Swami and Mendel have shown that the Kronecker state-cumulant vector $\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1})$ can be computed recursively in its temporal variable, n, as follows, for positive lags, i.e., for $\tau_l \geq 0$ $(l = 1, \cdots, k-1)$ (see Appendix A, Section O, for a derivation)

$$\boldsymbol{C}_{k,x}(n+1; 0, \cdots, 0) = \boldsymbol{\Phi}_k(n)\,\boldsymbol{C}_{k,x}(n; 0, \cdots, 0) + \boldsymbol{B}_k(n)\,\boldsymbol{\Gamma}_{k,v}(n) \tag{60}$$

$$\boldsymbol{C}_{k,x}(n+1; \tau_1, \cdots, \tau_{k-1}) = \left[\otimes_{l=0}^{k-1}\boldsymbol{\Phi}(n+\tau_l)\right]\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1}) + \boldsymbol{D}(n; \tau_1, \cdots, \tau_{k-1}) \tag{61}$$

where

$$\boldsymbol{D}(n; \tau_1, \cdots, \tau_{k-1}) = \left[\otimes_{l=0}^{k-1}\boldsymbol{A}(n+\tau_l+1, n+1)\right]\boldsymbol{B}_k(n)\,\boldsymbol{\Gamma}_{k,v}(n), \tag{62}$$

$\tau_0 = 0$, and for any matrix $\boldsymbol{M}$

$$\boldsymbol{M}_k = \boldsymbol{M}_{k-1} \otimes \boldsymbol{M}, \qquad \boldsymbol{M}_1 = \boldsymbol{M}. \tag{63}$$

Observe that (60) can be used to initialize (57). Just as (59) is initialized by $\boldsymbol{P}_x(0)$, (60) is initialized by $\boldsymbol{C}_{k,x}(0; 0, \cdots, 0)$.

It turns out that KSCV $\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1})$, where at least one of the lags is negative, can be expressed in terms of cumulants at positive lags and earlier times; hence, it is not necessary to have a recursive-in-time formula for $\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1})$ for negative τ_l's. This is analogous to the well known fact that $\boldsymbol{E}\{\boldsymbol{x}(n)\,\boldsymbol{x}'(j)\}$, where $n \neq j$, can be calculated from $\boldsymbol{P}_x(n)$.

Once $\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1})$ has been computed, it is easy to compute $\boldsymbol{C}_{k,y}(n; \tau_1, \cdots, \tau_{k-1})$, i.e.,

$$\boldsymbol{C}_{k,y}(n; \tau_1, \cdots, \tau_{k-1}) = \boldsymbol{\Psi}_k(n)\,\boldsymbol{C}_{k,x}(n; \tau_1, \cdots, \tau_{k-1}). \tag{64}$$

This is analogous to the fact that once $\boldsymbol{P}_x(n)$ has been computed it is easy to compute $\boldsymbol{P}_y(n)$, i.e.,

$$\boldsymbol{P}_y(n) = \boldsymbol{\Psi}(n)\,\boldsymbol{P}_x(n)\,\boldsymbol{\Psi}'(n) + \boldsymbol{R}(n) \tag{65}$$

where $\boldsymbol{R}(n)$ is the covariance matrix of Gaussian additive measurement noise $\boldsymbol{w}(n)$. Note that because $\boldsymbol{w}(n)$ is Gaussian it does not appear in (64).

When the MIMO system in (55) and (56) is time-invariant and stationary, so that all the matrices in its description are constant, then simplifications occur in all of the preceding KSCV results. See [53] for the details.

Many of the results which have been presented in Sections IV, V, and VII have already been extended to the multichannel situation, as described in [13] and [46].

The nonrecursive and recursive cumulant formulas of this section are even useful for SISO time-invariant and stationary systems. The usual approach to calculate cumulants for such systems is to use (12), which, of course, involves the IR of the system. Note, however, that (12) contains an infinite number of terms, so that in practice some truncation error is inevitable. It seems redundant, however, to compute a system's IR when a system's internal description is given. For example, if we model a system as an AR, MA, or ARMA model, it is a simple matter to reexpress each as an equivalent state-variable model. Even in the SISO case, such a model is in terms of a state vector; hence, even for such a SISO model we will need to calculate a KSCV. Doing this enables us to compute such a system's cumulants directly in terms of its parameters rather than in terms of its IR, and without any truncation errors.

X. Harmonic Processes

There are important signal processing applications where signals are either real or complex harmonic (sinusoidal) processes. The former occur in a wide range of array processing problems, whereas the latter occur in a wide range of harmonic retrieval problems. Without getting into notation for multiple measurements, frames, etc., we shall direct our attention in this section at cumulants for the following general class of signals:

$$y(n) = \sum_{i=1}^{p} a_i(n)\,s_n(\omega_i) \tag{66}$$

where the $s_n(\cdot)$'s are signal waveshapes, ω_i's are constants, and the $a_i(n)$'s are random variables. Prasad *et al.* [40], have shown that this model describes a wide range of important signal processing applications, including retrieval of harmonics in noise, bearing estimation with linear arrays (estimation of direction of arrival), time-delay estimation for broad-band sources, echo resolution, and transient response analysis. For example, in the harmonic retrieval problem, (66) can be expressed as

$$y(n) = \sum_{i=1}^{p} \alpha_i \exp\left\{j(\omega_i n + \phi_i)\right\} \tag{67}$$

where the ϕ_i's are independent identically distributed random variables uniformly distributed over $[-\pi, \pi]$, $\omega_i \neq \omega_j$ for $i \neq j$, and the α_i's and ω_i's are constants (i.e., not random). In this case, $s_n(\omega_i) = \exp(jn\omega_i)$ and $a_i(n) = a_i = \alpha \exp(j\phi_i)$, and $y(n)$ is complex. The comparable model for real signals is

$$y(n) = \sum_{i=1}^{p} \alpha_i \cos(\omega_i n + \phi_i). \tag{68}$$

Let $a = \exp(j\phi)$, where ϕ is uniformly distributed over $[-\pi, \pi]$. Then Swami [45] has shown (see Appendix A, Section P, for derivations) the following.

1) All third-order cumulants of complex harmonic a are always zero.
2) Of the three different ways (different in the sense of which of the variables should be conjugated) to define a fourth-order cumulant of a complex harmonic only one always yields a nonzero value, i.e., cum $(a, a, a, a) = 0$, cum $(a^*, a, a, a) = 0$, but, cum $(a^*, a^*, a, a) = -1$.

Consequently, we define the fourth-order cumulant of the complex process $y(n)$, as

$$C_{4,y}(\tau_1, \tau_2, \tau_3) = \text{cum}\left(y^*(n), y^*(n+\tau_1), y(n+\tau_2), y(n+\tau_3)\right) \quad (69)$$

where the fourth-order cumulant on the right-hand side of (69) is defined as in (4c).

Note that when $y(n) = \Sigma\, a_i \exp(j\omega_i n)$, where the a_i's are random, then the fourth-order cumulant is totally independent of the temporal variable n (prove this using (69) and some cumulant properties). Note, also, that in (67) for example, if $\alpha_i = \alpha_i(n)$, in which case we would have a damped complex exponential, a different approach from the one we are now discussing is advocated, because one does not get simple closed form formulas for the cumulants (we will obtain such formulas below for the models in (66), (67), and (68)). When one believes that complex damped exponentials are present then it is advisable to use an ARMA model as the starting point rather then the summation model in (67).

Next, we state formulas for the fourth-order cumulants of the signals in (66), (67), and (68) [45], [55]. In (66), if the $a_i(n)$ are random, zero-mean and mutually independent, with fourth-order cumulant $C_{4,a_i}(\tau_1, \tau_2, \tau_3)$, then (see Appendix A, Section P, for derivations)

$$\text{cum}\left(y^*(n), y^*(n+\tau_1), y(n+\tau_2), y(n+\tau_3)\right) = \sum_{i=1}^{p} s_n^*(\omega_i)\, s_{n+\tau_1}^*(\omega_i) \cdot s_{n+\tau_2}(\omega_i)\, s_{n+\tau_3}(\omega_i)\, C_{4,a_i}(\tau_1, \tau_2, \tau_3). \quad (70)$$

For (67), the case of complex harmonics,

$$C_{4,y}(\tau_1, \tau_2, \tau_3) = -\sum_{k=1}^{p} \alpha_k^4 \exp\left\{j\omega_k(-\tau_1 + \tau_2 + \tau_3)\right\}. \quad (71)$$

Additionally,

$$C_{2,y}(\tau) = \sum_{k=1}^{p} \alpha_k^2 \exp(j\tau\omega_i). \quad (72)$$

For (68), the case of real harmonics,

$$C_{4,y}(\tau_1, \tau_2, \tau_3) = -\tfrac{1}{8} \sum_{k=1}^{p} \alpha_k^4 [\cos\omega_k(\tau_1 - \tau_2 - \tau_3) + \cos\omega_k(\tau_2 - \tau_3 - \tau_1) + \cos\omega_k(\tau_3 - \tau_1 - \tau_2)]. \quad (73)$$

Additionally,

$$C_{2,y}(\tau) = \tfrac{1}{2} \sum_{k=1}^{p} \alpha_k^2 \cos(\omega_k\tau). \quad (74)$$

We are now ready to state a major theoretical result: If $y(n)$ is a sum of p real-valued sinusoids, then the diagonal slice (or, any 1-D slice) of the fourth-order cumulant retains all of the pertinent information about the number of harmonics, their amplitudes and frequencies [45], [55]. Specifically, set $\tau_1 = \tau_2 = \tau_3$ in (73) to see that

$$C_{4,y}(\tau) = -\tfrac{3}{8} \sum_{k=1}^{p} \alpha_k^4 \cos(\omega_k\tau) \quad (75)$$

which is identical (ignoring the scale factor of $-3/4$; see (74)) with the autocorrelation of the signal

$$\bar{y}(n) = \sum_{k=1}^{p} \alpha_k^2 \cos(\omega_k n + \phi_k). \quad (76)$$

It is well known that p, ω_k, and α_k can be estimated from the output correlation in the noiseless or additive white noise case; hence, they can also be estimated from $C_{4,y}(\tau)$, which is especially useful in the additive colored noise case. This means that already existing high-resolution methods, such as the Pisarenko, MUSIC, minimum-norm, Kumeresan-Tufts, or, Toeplitz approximation (ESPRIT) methods can be applied, just about as is, by replacing correlation quantities with $C_{4,y}(\tau)$.

XI. Estimates of Cumulants

Cumulants can be calculated from channel models using: (12) or (12′) in the case of single-channel systems that are described in terms of an IR; (54) in the case of multi-channel systems that are described in terms of IR matrices; or, (57) (or the recursive equations given in Section IX) in the case of single- or multi-channel systems that are described in terms of state space models.

In many practical situations we are given data and want to calculate cumulants from the data. In Section XII we describe some applications where this must be done in order to extract useful information about non-Gaussian signals from the data. Cumulants involve expectations, and as in the case of correlations, they cannot be computed in an exact manner from real data; they must be approximated, in much the same way that correlations are approximated. Cumulants are approximated by replacing expectations by sample averages, e.g.,

$$C_{3,x}(\tau_1, \tau_2) \simeq \hat{C}_{3,x}(\tau_1, \tau_2) = (1/N_R) \sum_{t \in R} x(t)\, x(t+\tau_1)\, x(t+\tau_2) \quad (77)$$

where N_R is the number of samples in region R. A similar but more complicated equation can be obtained for $\hat{C}_{4,x}(\tau_1, \tau_2, \tau_3)$ by beginning with (4c). It will not only involve a sum of a product of four terms (analogous to (77)), but it will also involve products of sample correlations.

It is well known (e.g., [25], [41]) that exponential stability of the underlying channel model guarantees the convergence in probability of the sampled autocorrelation function to the true autocorrelation function. The convergence of sampled third-order cumulants to true third-order cumulants is studied in [44]. Essentially, if the underlying channel model is exponentially stable, input random process $v(n)$ is stationary, and its first six (eight) cumulants are absolutely summable, then the sampled third-order (fourth-order) cumulants converge in probability to the true third-order (fourth-order) cumulants.

Formulas for estimating the covariances of higher-order moment estimates can be found in [9] and [39]. Although they are quite complicated, they may be of value in certain methods where estimates of such quantities are needed.

The reason it is important to know that sample estimates of cumulants converge in probability to their true values (i.e., are consistent) is that functions of these estimates are used in many of the techniques that have been developed to solve a wide range of signal processing problems (e.g., (23)). From estimation theory (e.g., [28]), it is well known that arbitrary functions of consistent estimates are also consistent; hence, we are assured of convergence in probability when using these techniques.

It is well known (e.g., [28]) that sampled estimates of Gaussian

processes are also optimal in a maximum-likelihood sense; hence they inherit all of the properties of such estimates. Unfortunately, sampled estimates of non-Gaussian processes are not necessarily optimal in any sense; hence, it may be true that estimates other than the conventional sampled estimates provide "better" results. This is an open question.

XII. Applications

Because the field of higher-order spectra is growing at an exponential rate, there is no way we can cover all applications. The *Proceedings of the Workshop on Higher-Order Spectral Analysis* (Vail, CO, June, 1989) cover a wide range of applications, ranging from identification of nonminimum phase systems, sonar and radar, nonlinear systems, harmonic retrieval and detection, image processing, biomedical signal analysis, and array processing. For other papers that discuss the wide range of applications to which higher-order statistics have already been applied, see [30], [35].

The application we shall cover is identification of systems from just noisy output measurements. The material in this section was first presented by the author at the Vail Workshop (June 1989) as a 90 minute tutorial.

A. Identification of Systems Just from Output Measurements

Consider the single-input single-output system depicted in Fig. 1, in which: $v(k)$ is a zero-mean, non-Gaussian, independent, identically distributed random sequence; $H(z)$ is the transfer function of a stable possibly nonminimum phase system (i.e., a system some of whose zeros lie outside of the unit circle in the complex z domain), i.e.,

$$H(z) = B(z)/A(z) = \left[\sum_{j=0}^{q} b(j)z^{-j}\right] \Big/ \left[\sum_{i=0}^{p} a(i)z^{-i}\right]. \tag{78}$$

$n(k)$ is a zero-mean colored Gaussian random sequence (which can include white Gaussian noise as a special case) that is independent of $v(k)$; $a(0) = b(0) \triangleq 1$, for scaling purposes; and, orders p and q are assumed to be known. The problem is: given time-limited noisy measurements $z(k)$, $k = 1, 2, \cdots, N$, estimate $H(z)$'s parameters, $b(1), \cdots, b(q)$, and $a(1), a(2), \cdots, a(p)$. This is an output measurement identification problem. It does not include the wide range of problems where one has access to both the input and the output of the system. It suffices to say that cumulant-based techniques can also be applied to the input/output measurements identification problem when the input is non-Gaussian.

The output measurement identification problem occurs in many fields, including communications and reflection seismology. In the former, $v(k)$ is a "message," $h(k)$ is a "channel," and $y(k)$ is a "distorted message" (intersymbol interference). An accurate model of the channel is needed in order to restore the message at the receiver. This model is used in many equalization schemes. In reflection seismology $v(k)$ is the earth's "reflectivity sequence," $h(k)$ is the "seismic source wavelet," and $y(k)$ is the "seismogram." An accurate model of the source wavelet is needed in order to estimate the earth's reflectivity sequence via deconvolution.

When the numerator parameters in (78) all equal zero except for $b(0)$, we have an all-pole model, in which case we are in the realm of AR parameter estimation. When all of the denominator parameters equal zero except for $a(0)$, we are in the realm of MA parameter estimation. These two special cases have been widely studied not only for their own interest, but also because some methods for estimating ARMA parameters proceed in two steps, by first estimating the AR parameters of the ARMA model and then estimating the MA parameters of the ARMA model, making use of the just-estimated AR parameters. We shall present a brief survey of methods that use higher-order statistics for the identification of AR, MA and ARMA systems just from noisy output measurements.

As a reminder, we use higher-order statistics to solve these problems because second-order statistics are phase blind (hence, they can only give rise to minimum phase or maximum phase models, i.e., to spectrally-equivalent minimum phase (SEMP) models), and higher-order statistics are blind to additive Gaussian noise.

Ideally, we seek a model that most-likely produced the given data. To do this we need the entire probability density function (p.d.f.), because likelihood is proportional to probability [28]; but, unless the data is Gaussian or conditionally Gaussian, it is almost impossible to determine the entire p.d.f., and to subsequently obtain the most-likely model. The Gaussian assumption means our p.d.f. is completely characterized by its first two moments; hence, when we use the Gaussian model we are fitting the data by a two-moment p.d.f. If we abandon the Gaussian model, then in theory we need an infinite number of moments to best fit the data. Suppose we determine models that are based on a subset of more than two moments (or cumulants). The methods described below determine linear models that are based on such a subset. A nonlinear model may be a more likely one.

B. Identification of AR Systems

Two methods for estimating the coefficients of a causal AR model have been described in Section V. The correlation-based normal equations in (25) can be solved by a wide variety of linear algebra techniques. When more equations are collected than unknowns, so that the resulting system (known as higher-order Yule-Walker equations) is overdetermined, least-squares, total least-squares [20], or singular-value decomposition (SVD) [6] techniques can be used to solve for the AR parameters. SVD techniques can also be used to determine the AR order p.

The cumulant-based normal equations in (26) must be collected for $\tau = 1, 2, \cdots, p + M$ and (at least) $k_0 = -p, \cdots, 0$ (or $\tau = q + 1, q + 2, \cdots, q + p + M$ and (at least) $k_0 = q - p, \cdots, q$ in the ARMA case) where $M \geq 0$. Singular-value decomposition techniques are then applied to not only determine the AR coefficients, but to also estimate the AR order p. For more details, see [10], [12], [16], [49].

If a system is truly linear and we are given true cumulants then cumulant-based (as well as correlation-based) AR methods lead to a stable AR model. Unfortunately, we are never given true cumulants; we are given estimates of cumulants (see Section XI). Additionally, we are sometimes given an arbitrary time series for which we assume a linear data generating model. In both cases cumulant-based (as well as correlation-based) AR methods are not guaranteed to lead to stable AR models.

If a correlation-based method leads to an unstable AR model, the unstable poles can be reflected outside of the unit circle without affecting the correlation, because $R(z) = A^{-1}(z)\,A^{-1}(z^{-1}) = R(z^{-1})$. The result is a noncausl AR model. If a cumulant-based method leads to an unstable AR model, we cannot just reflect the unstable pole outside of the unit circle to obtain the cure. Reflection changes $S_{3,y}(z_1, z_2)$, because $S_{3,y}(z_1, z_2) =$

$A^{-1}(z_1)\ A^{-1}(z_2)\ A^{-1}(z_1^{-1}z_2^{-1})$, and $S_{3,y}(z_1^{-1}, z_2^{-1}) = A^{-1}(z_1^{-1})\ A^{-1}(z_2^{-1})\ A^{-1}(z_1z_2)$, so that $S_{3,y}(z_1^{-1}, z_2^{-1}) \neq S_{3,y}(z_1, z_2)$. The same is true for other polyspectra.

This has motivated interest in identifying noncausal AR models, i.e., models that have some of their poles inside or outside of the unit circle. In the noncausal AR case, we view $H(z)$ as follows:

$$H(z) = 1/A(z) = [1/I(z)][1/O(z^{-1})] \qquad (79)$$

where $I(z)$ is associated with the minimum phase (stable) component of $A(z)$ and $O(z)$ is associated with the maximum phase (unstable) component of $A(z)$. Of course, noncausal AR models occur naturally in certain applications, such as astronomical signal processing, spatial signal processing (e.g., blurring distortion of images), geophysical signal processing (Vibrator input), and indirectly in methods for identifying noncausal MA models [33]. Three cumulant-based approaches have been developed to identify noncausal AR models: exhaustive search, optimization, and conversion to a related problem of identifying the coefficients of an MA model.

In the exhaustive search method [62], 1) a spectrally-equivalent (SE) stable causal AR(p) model is fit to the noisy measurements using only second-order statistics, 2) the resulting $A(z)$ is factored into $n_r + n_c$ terms, where n_r is the number of real roots and n_c is the number of complex roots, 3) 2^l SE $\hat{H}(z)$'s are created, where $l = n_r + n_c/2$, by reflecting one or more of the l roots or root-pairs of $A(z)$ to reciprocal locations outside of the unit circle; let $\hat{\theta}(i)$ denote the AR parameters for the ith model, where $i = 1, 2, \cdots, 2^l$, 4) a squared-error function is evaluated between $C_{3,z}(m, n)$ and $C_{3,z}(m, n|\hat{\theta}(i))$ over an admissible range of m and n values, for each $i = 1, 2, \cdots, 2^l$, where $C_{3,z}(m, n)$ is estimated from the given data and $C_{3,z}(m, n|\hat{\theta}(i))$ is computed for the ith SE model using (12) or (57), and, 5) the "winning" model is chosen as the l^* for which $\hat{H}(z)$ has the smallest squared error. Some of $\hat{H}(z)$'s poles will lie inside the unit circle, and some will lie outside of the unit circle; hence, $\hat{H}_l{}^*(z)$ will be noncausal.

Huzii [22] developed a similar method. Steps 1 through 3 of his method are identical with Tugnait's method. Huzii then calculates the theoretical cumulant of the innovations sequence for each of the 2^l models and tests each innovations sequence for whiteness. The whitest result provides the winning AR model.

In the optimization method [62] the coefficients of the noncausal AR model, as well as the statistics of the non-Gaussian input and the Gaussian additive measurement noise are chosen to minimize an objective function. This objective function combines squared differences between correlations computed from the data and from an AR model, and cumulants computed from the data and from the same AR model. The respective differences are summed over correlation or cumulant ranges. The AR parameters are determined by minimizing this objective function using a mathematical programming algorithm. We shall describe this method in more detail in subsection D.

In the conversion method [19] we use (40′), which is the first component of the "Double $C(q, k)$" method. In effect, we have converted an AR parameter estimation problem into a related MA parameter estimation problem, since the scaled AR coefficients, which appear on the right-hand side of (40′) play the role of MA parameters. In this method, we first solve (39′) for the scaled $d(i, j)$'s; then we solve for the scaled AR parameters using any one of the MA parameter estimation techniques to be described in subsection C; finally, we solve for the unscaled AR parameters.

In summary (see Table 1), although AR coefficients can be determined using correlation-based normal equations, they can also be determined using cumulant-based normal equations. The latter are unaffected by additive Gaussian noise, either white or colored, whereas the former are seriously affected by such noise. At least $p + 1$ cumulant slices must be used to reconstruct all of the AR coefficients using the cumulant-based normal equations; hence, in practice AR parameter estimation proceeds hand-in-hand with AR order determination [17]. Unfortunately, none of the causal cumulant-based methods can guarantee recovery of a stable AR model. Fortunately, however, causal or noncausal AR models can be identified using cumulants. Viewing a model as noncausal resolves the stability dilemma.

C. Identification of MA Systems

To-date numerous methods have appeared in the literature for the identification of causal MA systems using cumulants. In order to see the forest from the trees, we will group these methods in one of the following three categories of solutions: closed-form solutions, linear algebra solutions, and optimization solutions.

1) Closed-Form Solutions: Giannakis [10] and Giannakis and Mendel [16] were the first to show that MA coefficients could be computed recursively in closed form using both correlation and the 1-D third-order diagonal slice cumulant $C_{3,y}(m, m)$. A more elegant form of this solution was provided by Swami and Mendel [48]. Unfortunately, this recursive-method-1 contains a step in which one might be dividing zero by zero, and, it does not carry over to fourth-order cumulants. Finally, it is not a very practical algorithm because it does not smooth out the effects of measurement noise. In retrospect, recursive-method-1 demonstrated the "possibility" of identifying MA systems using cumulants.

Tugnait [63] developed a different closed form solution, one that also uses both correlation and 1-D third-order cumulant slices $C_{3,y}(\tau, \tau + q)$. This recursive-method-2 does not suffer from a zero divided by zero problem, and was inspired by recursive-method-1. No version of recursive-method-2 has been presented for fourth-order cumulants, and, it also is not a very practical

Table 1 Summary of Methods for Identifying AR Systems

AR System	Method	References
Causal	Correlation-based normal equations	Box and Jenkins [3]
	Cumulant-based normal equations	Parzen [65], Akaike [66], Giannakis [10, 12], Giannakis and Mendel [16], Swami [45], Swami and Mendel [52], Tugnait [60, 68]
Noncausal	Exhaustive search	Huzii [22], Tugnait [62]
	Optimization	Tugnait [62]
	Conversion	Giannakis and Swami [19]

algorithm because it does not smooth out the effects of measurement noise.

The third closed-form solution is the $C(q, k)$ method, given in (17) for third-order cumulants, and (18) for fourth-order cumulants. These formulas can also be given for an arbitrary mth-order cumulant. Unlike recursive-methods 1 or 2, the $C(q, k)$ method is not limited to third-order cumulants. It does not suffer from a zero divided by zero problem. As pointed out in Section IV, the $C(q, k)$ method is impractical because it also does not smooth out the effects of measurement noise.

Consequently, the closed-form solutions, while interesting from a theoretical point of view, in that they demonstrate the possibility of extracting MA coefficients from just output measurements, are not recommended for numerical calculations.

2) Linear Algebra Solutions: The GM-Method, described in Section VI, is the first of three methods which we have called "Linear Algebra Methods." Each of these methods requires solution of a system of equations which are viewed to be linear in a set of parameters; hence, the name "Linear Algebra Method." In the GM-Method, (34) is treated as a linear system in both $b(k)$ and $b^2(k)$ ($k = 1, 2, \cdots, q$). Equation (34) is concatenated for appropriate values of m (e.g., in the noise-free case $m = -q, \cdots, 0, \cdots, 2q$); $b(k)$ and $b^2(k)$ ($k = 1, 2, \cdots, q$) are solved for using least-squares; and, then $b(k)$ and $b^2(k)$ ($k = 1, 2, \cdots, q$) are combined to obtain the final estimate for $b(k)$ ($k = 1, 2, \cdots, q$). For different ways to combine $b(k)$ and $b^2(k)$, see [8], [16]. Even though the GM-Method treats $b(k)$ and $b^2(k)$ as independent parameters, which they obviously are not, it seems to lead to reasonably good results, and, is a simple method to implement. Another approach would be to treat the concatenated (34) as a system of nonlinear equations in the MA coefficients and to solve this system using a mathematical programming technique. As mentioned in Section VI, when measurement noise is present (34) becomes an underdetermined system of equations.

Tugnait [63] developed another linear algebra method, which we shall refer to as the "T-Method." It is based on the following equation which also links correlations and third-order cumulants (for a derivation, see Appendix A, Section Q):

$$\sum_{k=1}^{q} b(k)\, C_{3,y}(\tau - k, \tau - k + q) - [b(q)\gamma_{3,v}/\sigma_v^2] r_y(\tau) = -C_{3,y}(\tau, \tau + q). \tag{80}$$

This equation (which uses 1-D off-diagonal cumulant slices) is concatenated for $-q \le \tau \le q$; $b(1), b(2), \cdots, b(q)$, and $b(q)\gamma_{3,v}/\sigma_v^2$ are its unknowns; the resulting system is lower triangular and is always solvable for $b(1), b(2), \cdots, b(q)$, and $b(q)\gamma_{3,v}/\sigma_v^2$. In practice, least-squares is used to solve for $b(1), b(2), \cdots, b(q)$, and $b(q)\gamma_{3,v}/\sigma_v^2$. When measurement noise is present, in which case $r_y(\tau) = r_z(\tau) - \sigma_n^2\delta(\tau)$, we do not use $\tau = 0$; in this way we do not need to know σ_n^2. Observe that (80) is truly linear in the $q + 1$ MA coefficients, whereas (34) is really nonlinear in those coefficients; however, (34) makes use of much more correlation information than does (80). In the case of noisy measurements, it is quite reasonable to combine (80) and (34), and to then solve for $b(k)$ and $b^2(k)$ using least squares or to solve for just $b(k)$ using mathematical programming.

The third and final linear algebra method for causal MA systems is the Bicepstral Method, developed by Pan and Nikias [37] and described in Section VIII. Equation (46) is linear in $\hat{h}(k)$ and $\hat{h}(-k)$. By concatenating (46) for a sufficiently large number of m and n, least-squares can be used to solve for the p^* values of $\hat{h}(k)$ and the q^* value of $\hat{h}(-k)$. From these values it is then possible to reconstruct the minimum phase and maximum phase components of $h(k)$, namely $h_{\min}(k)$ and $h_{\max}(k)$, after which $h(k)$ is computed as $h(k) = h_{\min}(k)^* h_{\max}(k)$. Of course, for an MA system $h(k) = b(k)$. This method is easily extended to determine the IR's of AR and ARMA systems, as described by Pan and Nikias. The method does not require order determination of the underlying model, because it reconstructs the system's IR, $h(k)$, rather than AR or ARMA parameters. The method also does not require a priori knowledge of the system type (i.e., MA, AR or ARMA). The lengths of $h_{\min}(k)$ and $h_{\max}(k)$ are determined by the method. Unfortunately, cepstra are not defined for band-limited signals (the logarithm of zero blows up), which may impede the use of the Bicepstral Method in some situations. The computational complexity of the method increases when the system's poles or zeros are close to the unit circle (in the AR or ARMA cases), which means that more $h(k)$ and $h(-k)$ parameters will need to be estimated, and more cumulants will have to be computed. Good estimates of p^* and q^* are required (this, in effect, is order determination). If they are severely underestimated the phase and magnitude information will not be recovered accurately. The effects of additive noise needs to be studied (none of the examples given in Pan and Nikias include this effect). A first step in this direction is the recent paper by Petropulu and Nikias [64] which provides analytical performance evaluation of the complex cepstrum and bicepstrum. It gives approximate expressions of the bias and variance of the cepstrum parameters due to the presence of additive Gaussian measurement noise.

Unlike cepstral methods that usually require phase unwrapping, which is difficult to accomplish, the Bicepstral Method does not require phase unwrapping. One explanation for this is that (46) involves both the complex cepstrum of the channel and the third-order cumulants, and, of course, cumulants contain phase information.

Finally, Nikias and Chiang [33] have developed an approximate linear algebra method for estimating the magnitude and phase of a noncausal MA model. Their method is to transform the noncausal MA model into a noncausal AR model, for which there is a two-sided system of normal equations (which can only be obtained if we know for sure that the original MA model is truly noncausal) that are then concatenated and solved for the coefficients of the noncausal AR model using least squares. From these coefficients it is then straightforward to obtain the desired estimates of the magnitude and phase of the original noncausal MA model.

3) Optimization Solutions: Lii and Rosenblatt [24] proposed the following MA exhaustive search method for determining MA coefficients: 1) use output correlation to obtain the SEMP MA process, 2) factorize the MA polynomial and obtain 2^q competing MA models by reflecting one or more minimum-phase zeros to reciprocal locations outside the unit circle, 3) compute theoretical cumulant values for the 2^q MA models (using (12) or (57)), and 4) choose the model whose cumulant values match, in a least-squares sense, the output cumulant sequence as the true model.

Lii and Rosenblatt [24] also proposed the following MA optimization-method 1 for determining MA coefficients by minimizing the sum of the squared differences between the observed cumulants and the cumulants of the proposed model.

Although Lii and Rosenblatt proposed both of these methods, they apparently never tried them out, for no simulation results are given for either method in their 1982 paper. The two methods were later extended and expanded upon by Tugnait in connection

with identifying the coefficients of an ARMA model (see subsection D).

An interesting feature of optimization-method-1 is that, because we are able to express the third- or fourth-order cumulant directly in terms of the MA parameters (e.g., see (15)), we can obtain a closed-form formula for the gradient of the squared error function, i.e., if

$$\epsilon^2 = \sum_R \left[\sum_{k=0}^{q} b(k)\, b(k+m)\, b(k+n) - \hat{C}_{3,y}(m,n) \right]^2 \tag{81}$$

denotes the squared error function for third-order cumulants (a similar expression can be formulated for fourth-order cumulants), in which R denotes the shaded triangular region shown on Fig. 2, then

$$\begin{aligned} \partial\epsilon^2/\partial b(i) &= 2 \sum_R \left[\sum_{k=0}^{q} b(k)\, b(k+m)\, b(k+n) - \hat{C}_{3,y}(m,n) \right] \\ &\quad \cdot \left[b(i+m)\, b(i+n) + b(i-m)\, b(i-m+n) \right. \\ &\quad \left. + b(i-n)\, b(i-n+m) \right] \end{aligned} \tag{82}$$

where $i = 0, 1, \cdots, q$. This equation is very useful when we use an optimization algorithm such as steepest descent to minimize (81).

In the preceding optimization approach all of the data was used at one time and the MA coefficients were chosen to fit all the data simultaneously in a minimum squared-error sense. In a structured network approach Mendel and Wang [32] view each nonredundant value of $C_{3,y}(m, n)$ [or $C_{4,y}(l, m, n)$] and $r_y(l)$ as a pattern to be learned by a structured network. The structured network is trained using estimated values of $C_{3,y}(m, n)$ [or $C_{4,y}(l, m, n)$] and $r_y(l)$ as its training samples. The network has two-levels and three-layers; it belongs to the class of structure-controllable networks [21]. The weights of this network correspond to the MA parameters (this is very different from usual neural networks in which the weights usually have no physical meaning), and are trained using a steepest descent algorithm, that makes use of (82). Simulations have demonstrated that this structured network approach is very useful for simultaneous order determination and parameter estimation, i.e., if the MA order is overspecified then the overspecified parameters do converge to zero (in a probabilistic sense) via this training procedure. A similar behavior should be obtained using MA optimization-method 1.

Friedlander and Porat [9] developed the following MA optimization-method 2 for estimating moving-average coefficients: 1) establish a vector, s, of second- and third-order statistics, 2) estimate s as $\hat{s} = s_N(y_N)$, where y_N is a vector of N samples of the system's output, 3) estimate the asymptotic covariance of s, Σ, as $\hat{\Sigma}(y_N)$, 4) obtain an initial estimate of the MA parameters b, b_0, 5) compute the gradient of s evaluated at b_i, 6) let

$$V(x) = [s(x) - \hat{s}]' \hat{\Sigma}^{-1} [s(x) - \hat{s}] \tag{83}$$

then

$$\hat{b} = \arg\min_{x \in B} V(x). \tag{84}$$

The actual minimization is accomplished using a Newton-Gauss algorithm which uses the gradient of s, and 7) iterate on steps 5) and 6) until $|V(b_{i+1}) - V(b_i)| < \epsilon$. Steps 1 and 2 are straightforward. Step 3 is quite involved. Formulas for the elements of Σ have been computed in [39]. Friedlander and Porat use the GM-Method to provide b_0. If $\hat{\Sigma}$ is not positive definite, an eigenvalue decomposition must be performed to convert s into a lower-dimensional vector ξ. Step 5 is actually performed for ξ rather than s; it is in this step where the MA nature of the model is exploited. Noise-free examples demonstrate that this MA optimization method outperforms the GM-Method. No results have been provided however for the case of noisy measurements.

This entire procedure is not limited to MA models. It can also be applied to ARMA models, but to do so requires knowing an effective MA order (i.e., length of the ARMA IR) to estimate the covariance of third-order cumulants. It then uses the residual time-series method (see subsection D).

Note that (83) and (84) differ from the mathematical formulation of MA optimization-method 1 only in the introduction of the term $\hat{\Sigma}^{-1}$. This term down-weights the errors in which one does not have high confidence.

In summary (see Table 2), we have described ten methods for estimating the coefficients in an MA(q) model. Prior to these methods nothing really new had occurred regarding the solution to this problem in a very long time.

D. Identification of ARMA Systems

To-date, several methods have been reported in the literature for identifying the coefficients of an ARMA model from just noisy output measurements.

Table 2 Summary of Methods for Identifying MA Systems

Class of Methods	Method	References
Closed-Form Solutions	Recursive-1	Giannakis [10], Giannakis and Mendel [16], Swami and Mendel [48]
	Recursive-2	Tugnait [63]
	$C(q, k)$	Giannakis [11]
Linear Algebra Solutions	GM	Giannakis [10], Giannakis and Mendel [16], Friedlander and Porat [8, 9], Porat and Friedlander [39]
	T	Tugnait [63]
	Bicepstral	Pan and Nikias [37], Nikias and Pan [34]
Optimization Solutions	MA Exhaustive search	Lii and Rosenblatt [24]
	MA Optimization-1	Lii and Rosenblatt [24]
	Structured network	Mendel and Wang [32]
	MA Optimization-2	Friedlander and Porat [9], Porat and Friedlander [39]

Tugnait [58] extended the exhaustive search technique, first proposed by Lii and Rosenblatt [24], to the identification of ARMA coefficients. In this ARMA exhaustive search method, we 1) use prediction-error identification techniques to determine a SEMP model; this leads to a model with p poles and $p-1$ zeros, 2) form the 2^l possible sets of SE zeros, where $l = n_r + n_c/2$, by reflecting one or more minimum-phase zeros to reciprocal locations outside of the unit circle; this leads to 2^l SE models, $M_1, M_2, \cdots, M_{2^l}$, that all have the same poles but different zeros, 3) choose M^* as

$$M^* = \underset{M_i \forall i}{\text{Arg min}} \sum_{t_1=-L}^{0} \sum_{t_2=-L}^{0} \cdot \sum_{t_3=-L}^{0} \left[C_{4,y}(t_1, t_2, t_3 | M_i) - \hat{C}_{4,y}(t_1, t_2, t_3) \right]^2. \quad (85)$$

Cumulant estimate $\hat{C}_{4,y}(t_1, t_2, t_3)$ is obtained from the data, whereas $C_{4,y}(t_1, t_2, t_3 | M_i)$ is computed for the ith model M_i. The latter can be computed from (12) or (57). This ARMA exhaustive search method is, as noted by Tugnait, blind to all-pass factors because only correlation information is used in step 1.

Tugnait [61] has also extended the optimization technique, first proposed by Lii and Rosenblatt [24] to the identification of ARMA coefficients. In this ARMA optimization method, we 1) use sample statistics to compute sampled estimates of both correlations and fourth-order cumulants, 2) fix the ARMA order at p and let $\boldsymbol{\theta} = \text{col}\,(a, \cdots, a_p, b_1, \cdots, b_{p-1}, \sigma_n^2, \sigma_v^2, \gamma_{4,v})$, and, 3) choose $\boldsymbol{\theta}^*$ as

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\text{Arg min}} \left\{ \frac{1}{2} \sum_{\tau=-L_1}^{0} \left[\hat{r}_y(\tau) - r_y(\tau | \boldsymbol{\theta}) \right]^2 + \lambda/2 \sum_{t_1=-L_2}^{0} \sum_{t_2=-t_1}^{0} \sum_{t_3=-t_2}^{0} \cdot \left[\hat{C}_{4,y}(t_1, t_2, t_3) - C_{4,y}(t_1, t_2, t_3 | \boldsymbol{\theta}) \right]^2. \quad (86)$$

This is now a mathematical programming problem in which $\boldsymbol{\theta}$ is chosen by minimizing a weighted combination of squared errors between second- and fourth-order statistics. If the system's input is not symmetrically distributed, so that its third-order cumulant is nonzero, then (86) can be modified by replacing fourth-order cumulants with third-order cumulants. A formula for λ is given in [61]. This ARMA optimization method is not blind to all-pass factors.

Giannakis [10] and Giannakis and Mendel [16] developed the following three step residual time-series method for estimating ARMA coefficients: 1) estimate the AR coefficients, 2) compute the residual time series, and 3) estimate the MA parameters from the residual time series. The AR parameters can be estimated using any one of the methods that we have described in subsection B, although Giannakis and Mendel emphasize the cumulant-based normal equations. The so-called "residual time series," denoted $\hat{y}(n)$, equals $y(n) - \hat{y}(n)$, where

$$\hat{y}(n) = -\sum_{k=1}^{p} \hat{a}(k)\, y(n-k). \quad (87)$$

Beginning with the ARMA model in (21), letting $\tilde{a}(k) = a(k) - \hat{a}(k)$, and using (87), it is straightforward to rearrange (21) as

$$\hat{y}(n) = \sum_{k=0}^{q} b(k)\, v(n-k) - \sum_{k=1}^{p} \tilde{a}(k)\, y(n-k). \quad (88)$$

If $\hat{a}(k) = a(k)$, then

$$\hat{y}(n) = \sum_{k=0}^{q} b(k)\, v(n-k) \quad (89)$$

in which case the residual satisfies an MA(q) model. Generally speaking, $\hat{a}(k) \neq a(k)$, in which case the second sum on the right-hand side of (88) is present. Unfortunately, this term is a doubly stochastic process (because $\tilde{a}(k)$ depends on $y(j)$, $j = 1, \cdots, n$, so that $\tilde{a}(k)\, y(n-k)$ is a product of two dependent random sequences) that (to-date) defies analysis. The customary approach, therefore, is to assume that $\tilde{a}(k) = 0$.

Any of the MA coefficient methods that were described in subsection C can be used to estimate the MA coefficients of (89), where $\hat{y}(n)$, which is constructed as $y(n) - \hat{y}(n)$, plays the role of the MA system's measurement. Giannakis and Mendel emphasize the GM-Method for doing this.

Swami [45] and Swami and Mendel [49] have developed the following two-slice algorithm for determining MA coefficients in an ARMA model: 1) determine the AR coefficients using any one of the methods described in subsection B, and 2) determine the MA coefficients, as well as the noise statistics, using two 1-D cumulant slices. The starting point for the implementation of step 2 is (A-16) and (A-17). Analyzing the right-hand side of (A-17) for the $t = 0$ and $t = 1$ slices of $f_m(t; \tau)$ reveals that $f_m(0; \tau)$ is a function of (note that, because $h(l)$ is causal, the first nonzero term occurs when $j = \tau$) $\gamma_{m,v} b(\tau), \cdots, \gamma_{m,v} b(q), h(1), \cdots, h(q-\tau)$, whereas $f_m(1; \tau)$ is a function of $\gamma_{m,v} b(\tau), \cdots, \gamma_{m,v} b(q), h(1), \cdots, h(q-\tau+1)$. As usual, $h(0)$ is normalized to be unity. Setting $\tau = q$ in both $f_m(0; \tau)$ and $f_m(1; \tau)$, using the specific structure of the right-hand side of (A-17), lets us compute $\gamma_{m,v} b(q)$ and $h(1)$, in that order. Setting $\tau = q - 1$ in both $f_m(0; \tau)$ and $f_m(1; \tau)$, and again using the specific structure of the right-hand side of (A-17), lets us next compute $\gamma_{m,v} b(q-1)$ and $h(2)$, in that order. The complete procedure for step 2 (when q is even) is (2a) iterate $f_m(0; \tau)$ and $f_m(1; \tau)$ for $\tau = q, q-1, \cdots, q/2$ to obtain $\gamma_{m,v} b(q), h(1), \gamma_{m,v} b(q-1), h(2), \cdots, \gamma_{m,v} b(q/2)$ and $h(q/2+1)$; (2b) solve (22) for $b(0), b(1), \cdots, b(q/2-1)$; (2c) evaluate (22) at $k = q/2$ to obtain $b(q/2)$; (2d) having both $\gamma_{m,v} b(q/2)$ and $b(q/2)$, solve for $\gamma_{m,v}$; and, (2e) finally, having $\gamma_{m,v}$, solve for $b(q), b(q-1), \cdots, b(q/2+1)$ from $\gamma_{m,v} b(q), \gamma_{m,v} b(q-1), \cdots, \gamma_{m,v} b(q/2+1)$. A very similar procedure exists for q odd.

The novel feature of this two-slice algorithm is that it weaves together MA and IR estimation. It does not require calculation of a residual time series. It is also possible to weave together second- and higher-order statistics to solve for the MA coefficients. The $t = 0$ and $t = 1$ slices may even come from cumulants of different orders. Unfortunately, because of the recursive nature of this algorithm it is prone to error propogation. In essence, this algorithm is a generalization of the recursive algorithms for MA systems to ARMA systems.

Swami [45] and Swami and Mendel [49] have also developed the following q-slice algorithm for determining MA coefficients in an ARMA model: 1) determine the AR coefficients by any one of the methods described earlier in subsection B, 2) determine the first q IR coefficients using q 1-D cumulant slices, and 3) determine the MA coefficients using (22). Equation (23) is the basis for step 2. The q-slice algorithm does not introduce errors due to the computation of a residual time series, and it is not recursive so that it is not prone to error propogation. In fact, it is possible to recast (23) so that it can be combined with the cumulant-based normal equations in (27) (which, of course, can be

used for step 1), i.e., (23) can be reexpressed as (use the version of (23) given in (A-22), along with (A-16))

$$\sum_{k=1}^{p} a(k)\, C_{m,y}(q-k, n, 0, \cdots, 0) - f_m(0; q)\, h(n) = -C_{m,y}(q, n, 0, \cdots, 0) \quad (90)$$

for $n = 0, 1, \cdots, Q \geq q$. The concatenation of (27) and (90) leads to an overdetermined lower-triangular system of equations in the p unknown AR coefficients, $a(k)$, and Q unknown scaled IR coefficients, $f_m(0; q)\, h(n)$. Note that, because $h(0) = 1$, $f_m(0; q)$ can be solved for from this system of equations, so that the unscaled IR coefficients can also be solved for. Simulations indicate that the q-slice algorithm in which AR and IR coefficients are solved for simultaneously, using numerically robust techniques, such as singular value decomposition [20] and total least-squares [20], gives very good results.

In Section VII we described the Double $C(q, k)$ Algorithm, developed by Giannakis and Swami [18], [19] for estimating ARMA coefficients. Recall that this method is applicable to noncausal as well as causal ARMA models. The key point of this algorithm is that it reduces the ARMA parameter estimation problem to two MA parameter estimation problems. In Section VII we said that each of these MA problems could be solved by a direct application of the $C(q, k)$ algorithm that was described in Section IV; hence, the name "Double $C(q, k)$ Algorithm." Now that we have learned that there are (at least) ten new cumulant-based methods for estimating MA coefficients, let us modify the statement of the Double $C(q, k)$ algorithm (perhaps, we should refer to the following version as a "Double MA Algorithm") to: 1) estimate a sufficient number of output third-order cumulant values necessary to accomplish steps 2 and 4, 2) solve (39′) for $d(i, j)/d(0, 0)$, $0 \leq i \leq j \leq p$, 3) use any one of the MA algorithms described in subsection C (with q replaced by p), and the fact that $a(k) = a'(k)/a'(0)$, $k = 1, 2, \cdots, p$, to estimate the AR coefficients from (40′), 4) compute $b(m, n)$ from (41) for $0 \leq m \leq n \leq q$, and 5) use any one of the MA algorithms described in subsection C to estimate the MA coefficients from (42).

All of the preceding methods model $H(z)$ directly as a (possibly) nonminimum phase system, $H_{NMP}(z)$. Another possibility is to model $H(z)$ as a cascade of all-pass and minimum phase systems, $H_{AP}(z)$ and $H_{MP}(z)$, respectively. In essence, $H_{AP}(z)$ alters the phase of $H_{MP}(z)$ to produce $H_{NMP}(z) = H_{AP}(z)\, H_{MP}(z)$; it does not alter the spectrum of $H_{MP}(z)$. $H_{AP}(z)\, H_{MP}(z)$ is spectrally equivalent to $H_{NMP}(z)$. Giannakis [10] and Giannakis and Mendel [14] have developed a minimum-phase all-pass decomposition method for estimating the coefficients of an ARMA model. Correlation and a standard stochastic realization algorithm are used to first obtain $H_{MP}(z)$. Then $H_{AP}(z)$ is obtained. Because the steps of this algorithm are a bit involved, we leave its details to the reader.

In summary (see Table 3), we have presented 7 methods for estimating the coefficients in an ARMA(p, q) model. Some methods determine all of the ARMA coefficients simultaneously, whereas other methods first determine AR coefficients and then the MA coefficients. Of course, if the system's input is known, as in some control system applications, or testing procedures, then other "higher-order" techniques can be used to exploit all of this knowledge. These techniques will be (theoretically) immune to additive colored Gaussian noise, whereas (10) is not.

E. Examples

A multitude of simulated examples can be found in many of the references for the AR, MA and ARMA identification methods. These examples have certain things in common, namely: 1) results are often presented for different data lengths to demonstrate that, as longer data lengths are used, we get better results, "better" in the sense of lower parameter estimation error standard deviations, 2) results are presented for nonminimum phase channels, and often for channels that include an all-pass factor, and, 3) results are often presented for additive colored Gaussian measurement noise.

Showing results for different data lengths confirms the fact that longer data lengths are needed for cumulant-based estimates as compared to correlation-based estimates, and that the variances of both estimates are always limited by signal-to-noise ratio and the bandwidth of the channel. When cumulant-based parameter estimates are compared against correlation-based estimates, as they invariably are, then looking at nonminimum phase channels, especially those with all-pass factors, and additive colored Gaussian noise, could be viewed as stacking the deck for the cumulant-based methods. The correlation-based results should be viewed as baseline results; i.e., results that were obtainable before the development of the cumulant-based methods, and which were known to be poor, but about which nothing much could be done. The cumulant-based results demonstrate how much can be done when higher-order statistics are used on these same problems; hence, it is not stacking the deck to test cumulant-based methods in those situations where it is known ahead of time that they must do better than correlation-based results.

In this subsection we present two ARMA identification examples, merely to illustrate what can be obtained using some cumulant-based methods.

Example 1: As in [16], the channel is an ARMA(2, 2) with poles at $z = -0.5$ and 0.8, and zeros at $z = -2$ and 1.25; hence, this is a second-order all-pass filter, a filter that is totally invisible to output correlation-based methods. The filter's input was a non-Gaussian one-sided i.i.d. exponentially distributed random sequence, with $\sigma_v^2 = 1$ and $\gamma_{3,v} = 2$. 1024 Gaussian white noisy measurements were used, for which signal-to-noise ratio equals 100. The SNR in our examples is defined as SNR $= E\{y^2(i)\}/E\{n^2(i)\}$. A Monte Carlo simulation (different from the one in [16]) was conducted for 30 runs. The residual time series method (RTS) was used in which the AR coefficients were

Table 3 Summary of Methods for Identifying ARMA Systems

Method	References
ARMA exhaustive search	Tugnait [58]
ARMA optimization	Tugnait [61]
Residual time-series	Giannakis [10], Giannakis and Mendel [16]
Two-slice	Swami [45], Swami and Mendel [49]
q-slice	Swami [45], Swami and Mendel [49]
Double $C(q, k)$	Giannakis and Swami [18, 19]
Minimum-phase all-pass	Giannakis [10], Giannakis and Mendel [14]

Table 4 Estimated All-Pass Parameters

Method	$a(1)$	$a(2)$	$b(1)$	$b(2)$
True	−0.3	−0.4	0.75	−2.5
RTS	−0.2758 (0.0454)	−0.3875 (0.0512)	0.8251 (0.0882)	−2.2301 (0.1801)
DC(q, k)	−0.2784 (0.0512)	−0.3772 (0.0627)	0.8324 (0.0682)	−2.2000 (0.1509)

estimated using AR-based normal equations and the MA coefficients were estimated using the GM Method. Table 4 summarizes the true model parameters as well as average values for the estimates of the parameters and their associated standard deviations. The resulting approximation to the all-pass channel is quite good.

Results obtained from applying the Double $C(q, k)$(DC(q, k)) algorithm to the data are also summarized in Table 4. These results were obtained by applying a structured network training algorithm, as described in [32], to first determine the AR parameters and to then determine the MA parameters. For the former parameters a network was trained, using a gradient algorithm, on the patterns $d(1, 0)/d(0, 0)$, $d(1, 1)/d(0, 0)$, $d(2, 0)/d(0, 0)$, $d(2, 1)/d(0, 0)$, $d(2, 2)/d(0, 0)$; $d(1, 0)/d(0, 0)$, $\cdots$. For the latter parameters another network was trained, again using a gradient algorithm, but this time on the patterns $b(0, 0)$, $b(1, 0)$, $b(1, 1)$, $b(2, 0)$, $b(2, 1)$, $b(2, 2)$; $b(0, 0)$, $b(1, 0)$, $\cdots$. Again, the resulting approximation to the all-pass channel is quite good.

Convergence (in a probabilistic sense) of the training procedure to the true parameter values is depicted in Fig. 4. The true and estimated impulse responses for the all-pass filter are depicted in Fig. 5. A comparable plot for the RTS results could not be distinguished from the DC(q, k) results; hence, we do not show it here.

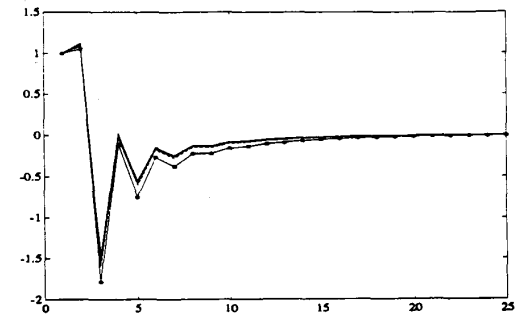


Fig. 5. True (line with circles) versus estimated impulse rsponses (30 realization).

Example 2: The channel is the following ARMA(3, 2) with an all-pass factor [49]:

$$H(z) = [1 - 2.95z^{-1} + 1.90z^{-2}]/[1 - 1.30z^{-1} + 1.05z^{-2} - 0.325z^{-3}] = (z - 2)z(z - 0.95)/(z - 0.5)(z^2 - 0.8z + 0.65]. \quad (91)$$

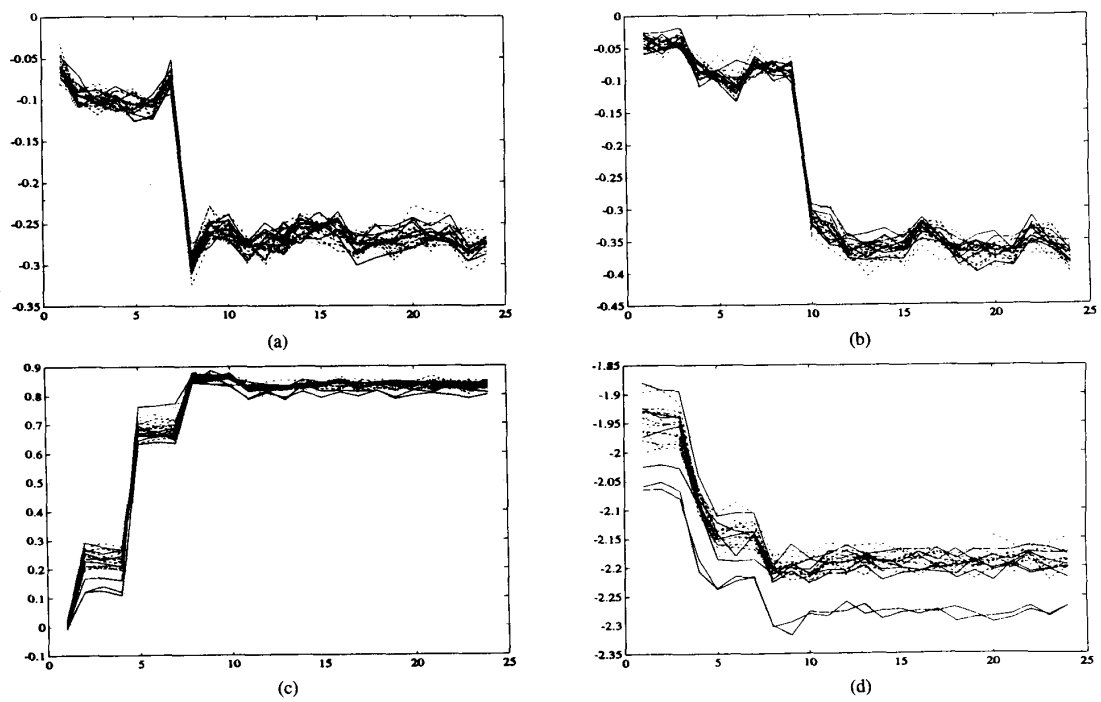


Fig. 4. Estimates of ARMA parameters: (a) a(1), (b) a(2), (c) b(1), and (d) b(2). The horizontal axis corresponds to the number of training patterns, whereas the vertical axis corresponds to the value of the parameter.

As in Example 1, the filter's input was a non-Gaussian one-sided i.i.d. exponentially distributed random sequence, with $\sigma_v^2 = 1$ and $\gamma_{3,v} = 2.2048$ Gaussian white noisy measurements were used, for which signal-to-noise ratio equaled 20. A Monte Carlo simulation was conducted for 30 runs. Sample averages of the third-order cumulant $C_{3,z}(m, k_0)$ were computed for $k_0 = -1, 0, 1, 2$ and $|m| \leq 12$; p and q were assumed known. Parameters were estimated by the following four methods.

Method RTS: ARMA parameters were estimated using the residual time-series method.

Method Qs2: AR parameters were estimated using SVD and total least-squares; then IR parameters, and subsequently the MA parameters, were estimated using the q-slice algorithm.

Method Qs: AR and IR parameters were estimated simultaneously using least-squares in the q-slice method. MA parameters were then estimated from (22).

Method Qst: Same as Method Qs, except total least-squares was used instead of least-squares.

Tables 5 and 6 summarize the results for the AR and MA parameter estimates, respectively. For this example, the Qs2 and Qst methods give the best AR coefficient estimates, whereas the Qst method gives the best MA coefficient estimates. A point that this example illustrates is that it is very important to use numerically robust methods for solving a system of overdetermined equations. The combination of SVD and total least-squares is quite powerful and does, indeed, improve results markedly.

Table 5 AR Parameter Estimates

Method	$a(1)$	$a(2)$	$a(3)$
True	−1.3	1.05	−0.325
RTS	−1.2148 (0.1838)	0.9732 (0.1851)	−0.2632 (0.1570)
Qs2	−1.2907 (0.1380)	1.0395 (0.1366)	−0.3182 (0.1147)
Qs	−0.7946 (0.1527)	0.6358 (0.1255)	0.0437 (0.1130)
Qst	−1.3261 (0.1927)	1.0701 (0.1810)	−0.3497 (0.1644)

Table 6 MA Parameter Estimates

Method	$b(1)$	$b(2)$
True	−2.95	1.90
RTS	−2.6808 (2.5098)	1.6058 (1.5808)
Qs2	−3.0493 (2.7199)	2.2277 (2.6669)
Qs	−0.9452 (0.8032)	0.1735 (0.6240)
Qst	−2.8646 (2.1578)	1.9737 (2.3018)

XIII. Conclusions

The main purpose of this paper has been to collect a large number of relatively new theoretical results for higher-order statistics in one place so that the reader will have access to all of them. Extensions of many of the results in Sections IV, V, and VII to two-dimensional systems can be found in [47]. A second purpose of this paper has been to illustrate the application of higher-order statistics to the broad range of problems associated with identifying an unknown, possibly nonminimum-phase channel from just noisy output measurements. A multitude of new methods has been developed using higher-order statistics.

No doubt this author has inadvertantly omitted some other "important" new theoretical results and some other identification methods. This was not done intentionally; the field is emerging very rapidly. Very worthy of mention is the interesting phase-coupling discrimination results obtained in [42] and [45]. The former showed that quadratic phase coupling can be detected by the bispectrum. The latter showed that cubic phase coupling can be detected by the trispectrum, and, that the trispectrum is "blind" to quadratic phase coupling, whereas the bispectrum is "blind" to cubic phase coupling.

This paper has focused exclusively on higher-order statistics for linear models because such models fall within the author's area of expertise. See [35] for references that deal with applications of higher-order spectra to nonlinear models, including works by Brillinger, Hinich, Kim and Powers, and others.

It is hoped that the reader will hve a better picture of some of what has been happening in the emerging field of higher-order statistics as a result of this paper. As of the writing of this paper, more than 200 papers have already been published, mostly in conference proceedings (see, e.g., [35] or [30]). This represents an exponential growth, during the past few years.

Much work remains to be done, including testing of the many algorithms that have been developed across many applications on real data, extending results to new and interesting processes (e.g., multiplicative processes, point processes, hidden Markov processes, etc.) and applications, developing more accurate ways to estimate third- and fourth-order cumulants from data, exploiting parallelism in cumulant-based methods to develop "fast" algorithms, and, implementing cumulant-based algorithms in VLSI.

A reviewer of this paper asked the question "How could the reader choose a (nonminimum phase identification) technique without first trying them all?" This question is open to debate even for the multitude of correlation (spectrum)-based techniques that have been developed during the past 30 years. No precise answer is known at this time for the cumulant (polyspectral)-based techniques. Our limited experience suggests the use of either an optimization-based method or a linear algebra-based method; but, the latter must be accompanied by the use of SVD and total least squares. Establishing performance bounds for many of the methods remains to be done.

If non-Gaussianity, nonminimum phase, colored noise effects, or nonlinearities are important to you, then higher-order statistics offer a new set of tools, methodologies and algorithms for handling some or all of these effects. We are in no way advocating the abandonment of second-order statistics; we are, however, advocating the use of more than second-order statistics in the face of the preceding effects. As stated in the Introduction, "with the new results that are being developed and those that are described in this paper, it should be possible to reexamine every application and/or method that has ever made use of second-order statistics, using higher-order statistics, to see if better results can be obtained."

Appendix A
Derivations and Relationships

A. Relationships Between Cumulants and Moments

Let $\boldsymbol{x}$ denote a collection of random variables, i.e., $\boldsymbol{x} = \text{col}\,(x_1, x_2, \cdots, x_k)$, and $I_x = \{1, 2, \cdots, k\}$ denote the set of indices of the components of $\boldsymbol{x}$. If $I \subseteq I_x$, then $\boldsymbol{x}_I$ is the vector consisting of those components of $\boldsymbol{x}$ whose indices belong to I. We denote the simple moment and cumulant of the subvector $\boldsymbol{x}_I$ of the vector $\boldsymbol{x}$ as $m_x(I)$ (i.e., $m_x(I)$ is the expectation of the product of the elements in $\boldsymbol{x}_I$) and $C_x(I)$. The "partition" of the set I is the unordered collection of nonintersecting nonempty sets I_p such that $\cup_p I_p = I$. For example, the set of partitions corresponding to $k = 3$ is $\{(1, 2, 3)\}$, $\{(1), (2, 3)\}$, $\{(2), (1, 3)\}$, $\{(3), (1, 2)\}$, $\{(1), (2), (3)\}$.

The moment-to-cumulant (i.e., M-C) formula is [23]:

$$C_x(I) = \sum_{\cup_{p=1}^{q} I_p = I} (-1)^{q-1}(q-1)! \prod_{p=1}^{q} m_x(I_p) \quad \text{(A-1)}$$

where $\cup_{p=1}^{q} I_p = I$ denotes summation over all partitions of set I. In the preceding example, $q = 1$ for $\{(1, 2, 3)\}$, $q = 2$ for $\{(1), (2, 3)\}$, $\{(2), (1, 3)\}$, and $\{(3), (1, 2)\}$, and, $q = 3$ for $\{(1), (2), (3)\}$. The cumulant-to-moment (i.e., C-M) equation is

$$m_x(I) = \sum_{\cup_{p=1}^{q} I_p = I} \prod_{p=1}^{q} C_x(I_p). \quad \text{(A-2)}$$

An example which illustrates the use of (A-1) and (A-2) for $I = \{1, 2, 3, 4\}$ is given in Table A-1. In its bottom row, Σ means add all preceding rows to obtain either cum (x_1, x_2, x_3, x_4), using (A-1),or $E\{x_1x_2x_3x_4\}$, using (A-2).

B. Properties of Cumulants of Scalar Processes

Property [CP1]: Let $y = \text{col}\,(\lambda_1 x_1, \cdots, \lambda_k x_k)$ and $x = \text{col}\,(x_1, \cdots, x_k)$. Note that (see Section A) $I_x = I_y$. From (A-1), we see that

$$C_y(I_y) = \sum_{\cup_{p=1}^{q} I_p = I_y} (-1)^{q-1}(q-1)! \prod_{p=1}^{q} m_y(I_p) \quad \text{(A-3)}$$

where (see, for example, Table A-1)

$$\prod_{p=1}^{q} m_y(I_p) = \left(\prod_{p=1}^{k} \lambda_p\right)\left(\prod_{p=1}^{q} m_x(I_p)\right). \quad \text{(A-4)}$$

Consequently,

$$C_y(I_y) = \left(\prod_{p=1}^{k} \lambda_p\right) C_x(I_x) \quad \text{(A-5)}$$

which is (6a).

Property [CP2]: Referring to (A-1), since the partition of the set I_x is an unordered collection of nonintersecting nonempty sets of I_p such that $\cup_{p=1}^{q} I_p = I_x$, the order in the cumulant's argument is irrelevant to the value of the cumulant. As a result, cumulants are symmetric in their arguments.

Property [CP3]: Let $x = \text{col}\,(u_1 + v_1, x_2, \cdots, x_k)$, where $u = \text{col}\,(u_1, x_2, \cdots, x_k)$, and $v = \text{col}\,(v_1, x_2, \cdots, x_k)$. Observe that (because $m_x(I_i)$ is the expectation of the product of the elements in I_i and $u_1 + v_1$ appears only raised to the unity power)

$$\prod_{i=1}^{q} m_x(I_i) = \prod_{i=1}^{q} m_u(I_i) + \prod_{i=1}^{q} m_v(I_i). \quad \text{(A-6)}$$

Substitute (A-6) into (A-1) to obtain the result in (6c).

Property [CP4]: Let $y = \text{col}\,(\alpha + x_1, x_2, \cdots, x_k)$; then, from (1), we see that

$$\begin{aligned} K(v) &= \ln E\{\exp(jv'y)\} \\ &= \ln E\{\exp(j[v_1(\alpha + x_1) + v_2x_2 + \cdots + v_kx_k])\} \\ &= \ln E\{\exp(jv_1\alpha)\} \\ &\quad + \ln E\{\exp(jv_1x_1 + \cdots + jv_kx_k)\}. \end{aligned} \quad \text{(A-7)}$$

According to the paragraph which precedes (1), we know that

$$\text{cum}\,(\alpha + x_1, x_2, \cdots, x_k) = \frac{1}{k!}[\partial^k/\partial v_1 \partial v_2 \cdots \partial v_k]K(v)\big|_{v=0} \quad \text{(A-8)}$$

but, from (A-7), we see that

$$\begin{aligned} &\frac{1}{k!}[\partial^k/\partial v_1\partial v_2 \cdots \partial v_k]K(v)\big|_{v=0} \\ &= \frac{1}{k!}[\partial^k/\partial v_1\partial v_2 \cdots \partial v_k]\big[\ln E\{\exp(jv_1\alpha)\} \\ &\quad + \ln E\{\exp(jv_1x_1 + \cdots + jv_kx_k)\}\big]\big|_{v=0} \\ &= \frac{1}{k!}[\partial^k/\partial v_1\partial v_2 \cdots \partial v_k] \\ &\quad \cdot \ln E\{\exp(jv_1x_1 + \cdots + jv_kx_k)\}\big]\big|_{v=0} \\ &= \text{cum}\,(x_1, x_2, \cdots, x_k) \end{aligned}$$

which is (6d).

Property [CP5]: Let $z = \text{col}\,(x_1 + y_1, \cdots, x_k + y_k) = x + y$, where $x = \text{col}\,(x_1, \cdots, x_k)$ and $y = \text{col}\,(y_1, \cdots, y_k)$.

Table A-1 Calculations of Fourth-Order Cumulants in Terms of Moments and Vice-Versa

I_1	I_2	I_3	I_4	q	M-C Equation $(-1)^{q-1}(q-1)!\prod_{p=1}^{q} m_x(I_p)$	C-M Equation $\prod_{p=1}^{q} C_x(I_p)$
1	2	3	4	4	$-6E\{x_1\}E\{x_2\}E\{x_3\}E\{x_4\}$	$C(x_1)C(x_2)C(x_3)C(x_4)$
1, 2	3	4		3	$2E\{x_1x_2\}E\{x_3\}E\{x_4\}$	$C(x_1, x_2)C(x_3)C(x_4)$
1, 3	2	4		3	$2E\{x_1x_3\}E\{x_2\}E\{x_4\}$	$C(x_1, x_3)C(x_2)C(x_4)$
1, 4	2	3		3	$2E\{x_1x_4\}E\{x_2\}E\{x_3\}$	$C(x_1, x_4)C(x_2)C(x_3)$
2, 3	1	4		3	$2E\{x_2x_3\}E\{x_1\}E\{x_4\}$	$C(x_2, x_3)C(x_1)C(x_4)$
2, 4	1	3		3	$2E\{x_2x_4\}E\{x_1\}E\{x_3\}$	$C(x_2, x_4)C(x_1)C(x_3)$
3, 4	1	2		3	$2E\{x_3x_4\}E\{x_1\}E\{x_2\}$	$C(x_3, x_4)C(x_1)C(x_2)$
1, 2	3, 4			2	$-E\{x_1x_2\}E\{x_3x_4\}$	$C(x_1, x_2)C(x_3, x_4)$
1, 3	2, 4			2	$-E(x_1x_3)E(x_2x_4)$	$C(x_1, x_3)C(x_2, x_4)$
1, 4	2, 3			2	$-E\{x_1x_4\}E\{x_2x_3\}$	$C(x_1, x_4)C(x_2, x_3)$
1, 2, 3	4			2	$-E\{x_1x_2x_3\}E\{x_4\}$	$C(x_1, x_2, x_3)C(x_4)$
1, 2, 4	3			2	$-E\{x_1x_2x_4\}E\{x_3\}$	$C(x_1, x_2, x_4)C(x_3)$
1, 3, 4	2			2	$-E\{x_1x_3x_4\}E\{x_2\}$	$C(x_1, x_3, x_4)C(x_2)$
2, 3, 4	1			2	$-E\{x_2x_3x_4\}E\{x_1\}$	$C(x_2, x_3, x_4)C(x_1)$
1, 2, 3, 4				1	$E\{x_1x_2x_3x_4\}$	$C_1\{x_1, x_2, x_3, x_4\}$
				Σ	cum (x_1, x_2, x_3, x_4)	$E\{x_1x_2x_3x_4\}$

Using the independence of the $\{x_i\}$ and $\{y_i\}$, it follows that

$$\begin{aligned} K_z(\boldsymbol{v}) &= \ln E\{\exp[jv_1(x_1+y_1)+\cdots+jv_k(x_k+y_k)]\} \\ &= \ln E\{\exp(jv_1x_1+\cdots+jv_kx_k)\} \\ &\quad + \ln E\{\exp(jv_1y_1+\cdots+jv_ky_k)\} \\ &= K_x(\boldsymbol{v})+K_y(\boldsymbol{v}) \end{aligned} \tag{A-9}$$

from which the result in (6e) follows directly.

Property [CP6]: Assume that $(x_1, \cdots, x_i)$ is independent of $(x_{i+1}, \cdots, x_k)$; hence,

$$\begin{aligned} K(\boldsymbol{v}) &= \ln E\{\exp(jv_1x_1+\cdots+jv_ix_i)\} \\ &\quad + \ln E\{\exp(jv_{i+1}x_{i+1}+\cdots+jv_kx_k)\}. \end{aligned} \tag{A-10}$$

Now,

$$\begin{aligned} &\frac{1}{k!}[\partial^k/\partial v_1\partial v_2\cdots\partial v_k]K(\boldsymbol{v})\big|_{v=0} \\ &\quad = \frac{1}{k!}[\partial^k/\partial v_1\partial v_2\cdots\partial v_k] \\ &\qquad \cdot [\ln E\{\exp(jv_1x_1+\cdots+jv_ix_i)\} \\ &\qquad + \ln E\{\exp(jv_{i+1}x_{i+1}+\cdots+jv_kx_k)\}]\big|_{v=0} = 0 \end{aligned}$$

which is (6f).

C. Cumulant and Polyspectrum for Single Channel Systems

Here we derive the Bartlett-Brillinger-Rosenblatt formulas in (12) and (13). We also derive (12′). Because $n(k)$ is assumed to be Gaussian, the kth-order cumulant of $z(k)$ equals the kth-order cumulant of $y(k)$, where

$$y(k) = \sum_{i=-\infty}^{\infty} v(i)\, h(k-i). \tag{A-11}$$

Note that the following derivation is expedited by working with the more general form in (A-11) where i ranges from $-\infty$ to ∞, instead of the form associated with a causal IR for which i ranges from 0 to k. Changes of variables then do not change the ranges of the summation. Consequently,

$$\begin{aligned} &C_{k,z}(\tau_1, \cdots, \tau_{k-1}) \\ &\quad = C_{k,y}(\tau_1, \cdots, \tau_{k-1}) \\ &\quad = \operatorname{cum}(y(l), y(l+\tau_1), \cdots, y(l+\tau_{k-1})) \\ &\quad = \operatorname{cum}\Big[\sum_{i_0} v(i_0)\, h(l-i_0), \sum_{i_1} v(i_1) \\ &\qquad \cdot h(l-i_1+\tau_1), \cdots, \\ &\qquad \sum_{i_{k-1}} v(i_{k-1})\, h(l-i_{k-1}+\tau_{k-1})\Big] \\ &\quad = \sum_{i_0}\sum_{i_1}\cdots\sum_{i_{k-1}} \operatorname{cum}[v(i_0)\, h(l-i_0), \\ &\qquad v(i_1)\, h(l-i_1+\tau_1), \cdots, \\ &\qquad v(i_{k-1})\, h(l-i_{k-1}+\tau_{k-1})] \\ &\quad = \sum_{i_0}\sum_{i_1}\cdots\sum_{i_{k-1}} h(l-i_0)\, h(l-i_1+\tau_1)\cdots \\ &\qquad h(l-i_{k-1}+\tau_{k-1}) \\ &\qquad \cdot \operatorname{cum}[v(i_0), v(i_1), \cdots, v(i_{k-1})] \end{aligned} \tag{*}$$

To arrive at the third line of this derivation we have used cumulant property [CP3]; and, to arrive at the fourth line we have used cumulant property [CP1].

In the case of a white noise input, (*) simplifies considerably to

$$\begin{aligned} &C_{k,z}(\tau_1, \cdots, \tau_{k-1}) \\ &\quad = \gamma_{k,v}\sum_{i_0} h(l-i_0)\, h(l-i_0+\tau_1)\cdots h(l-i_0+\tau_{k-1}) \\ &\quad = \gamma_{k,v}\sum_{n=0}^{\infty} h(n)\, h(n+\tau_1)\cdots h(n+\tau_{k-1}) \end{aligned}$$

which is (12). To arrive at the first line we have used (11), noting that

$$\begin{aligned} &\operatorname{cum}[v(i_0), v(i_1), \cdots, v(i_{k-1})] \\ &\quad = \operatorname{cum}[v(i_0), v(i_0+i_1-i_0), \cdots, v(i_0+i_{k-1}-i_0)] \\ &\quad = C_{k,v}(i_1-i_0, \cdots, i_{k-1}-i_0) \\ &\quad = \gamma_{k,v} \text{ only if } i_1 = i_0, \cdots, i_{k-1} = i_0 \end{aligned}$$

and, to arrive at the second line we have made a simple substitution of variables and invoked the stationarity of $y(k)$ and the causality of $h(k)$. The former tells us that $C_{k,y}(\tau_1, \cdots, \tau_{k-1}) = C_{k,z}(\tau_1, \cdots, \tau_{k-1})$ will not depend upon time k; the latter tells us that $h(k) = 0$ for $k < 0$.

In the case of a colored noise input we first make the substitution of variables: $j_0 = l - i_0, j_1 = l - i_1 + \tau_1, \cdots, j_{k-1} = l - i_{k-1} + \tau_{k-1}$, so that (*) becomes

$$\begin{aligned} &C_{k,z}(\tau_1, \cdots, \tau_{k-1}) \\ &\quad = \sum_{j_0}\sum_{j_1}\cdots\sum_{j_{k-1}} h(j_0)\, h(j_1)\cdots h(j_{k-1}) \\ &\qquad \cdot \operatorname{cum}[v(l-j_0), v(l+\tau_1-j_1), \cdots, \\ &\qquad v(l+\tau_{k-1}-j_{k-1})]. \end{aligned}$$

Using the stationarity of $v(l)$, this last equation can be expressed as

$$\begin{aligned} &C_{k,z}(\tau_1, \cdots, \tau_{k-1}) \\ &\quad = \sum_{j_0}\sum_{j_1}\cdots\sum_{j_{k-1}} h(j_0)\, h(j_1)\cdots h(j_{k-1}) \\ &\qquad \cdot C_{k,v}(j_0-j_1+\tau_1, j_0-j_2+\tau_2, \cdots, \\ &\qquad j_0-j_{k-1}+\tau_{k-1}). \end{aligned}$$

Finally, making a second transformation of variables, $m_1 = j_1 - j_0, m_2 = j_2 - j_0, \cdots, m_{k-1} = j_{k-1} - j_0$, we obtain the result in (12′) (where we have replaced j_0 by k_1).

The polyspectrum in (13) is easily obtained by taking the $(k-1)$-dimensional Fourier transform of (12):

$$\begin{aligned} &S_{k,z}(\omega_1, \cdots, \omega_{k-1}) \\ &\quad = \sum_{\tau_1=-\infty}^{\infty}\cdots\sum_{\tau_{k-1}=-\infty}^{\infty} C_{k,z}(\tau_1, \cdots, \tau_{k-1}) \\ &\qquad \cdot \exp[-j(\omega_1\tau_1+\cdots+\omega_{k-1}\tau_{k-1})]. \end{aligned} \tag{A-12}$$

Substitute (12) into (A-12) to obtain (13).

D. The $C(q, k)$ Formula

Here, as in [11], we derive the $C(q, k)$ formula for third-order cumulants, leaving the comparable derivation for fourth-order

cumulants to the reader. We begin with (12) for $k = 3$, i.e.,

$$C_{3,z}(\tau_1, \tau_2) = \gamma_{3,v} \sum_{n=0}^{\infty} h(n)\, h(n + \tau_1)\, h(n + \tau_2) \quad \text{(A-13)}$$

in which $h(0) = 1$, for normalization purposes. Set $\tau_1 = q$ and $\tau_2 = k$ in (A-13), and use the fact that for an MA(q) system $h(j) = 0\ \forall\, j > q$, to see that

$$C_{3,z}(q, k) = \gamma_{3,v} h(q)\, h(k). \quad \text{(A-14)}$$

Next, set $\tau_1 = q$ and $\tau_2 = 0$ in (A-13), to see that

$$C_{3,z}(q, 0) = \gamma_{3,v} h(q). \quad \text{(A-15)}$$

Dividing (A-14) by (A-15) we obtain the $C(q, k)$ formula given in (17).

To obtain (19a) for $\gamma_{3,v}$, set $\tau_1 = \tau_2 = 0$ in (A-13), and then use (17) for $h(k)$.

E. q-Slice Algorithm

To begin, we define the scalar function $f_m(t; \tau)$ as

$$f_m(t; \tau) = \sum_{k=0}^{p} a(k)\, C_{m,y}(\tau - k, t, 0, \cdots, 0). \quad \text{(A-16)}$$

If the AR coefficients are known, and the cumulants are computed from the data, then these functions can be computed. Next we show, as in [45], [49], that the right-hand side of (A-16) can be expressed in terms of the MA coefficients and the system's impulse response, i.e.,

$$f_m(t; \tau) = \gamma_{m,v} \sum_{j=0}^{q} h^{(m-2)}(j - \tau)\, h(j - \tau + t)\, b(j). \quad \text{(A-17)}$$

To obtain (A-17), we begin by substituting (12) into the right-hand side of (A-16), i.e.,

$$\sum_{k=0}^{p} a(k)\, C_{m,y}(\tau - k, t, 0, \cdots, 0) = \gamma_{m,v} \sum_{i=0}^{\infty} h^{(m-2)}(i)\, h(i + t) \sum_{k=0}^{p} a(k)\, h(i + \tau - k). \quad \text{(A-18)}$$

From (22), we recognize that the sum on the far right of (A-18) equals $b(i + \tau)$; hence, (A-18) becomes

$$\sum_{k=0}^{p} a(k)\, C_{m,y}(\tau - k, t, 0, \cdots, 0) = \gamma_{m,v} \sum_{i=0}^{\infty} h^{(m-2)}(i)\, h(i + t)\, b(i + \tau). \quad \text{(A-19)}$$

Finally, to obtain (A-17) from (A-19): substitute the right-hand side of (A-19) into (A-16), letting $i + \tau = j$; truncate the upper limit in the summation from ∞ to q, because $b(j) = 0$ for $\forall\, j > q$; and, extend the lower limit in the summation from $j = \tau$ to $j = 0$, because $h(l)$ is causal (so that $h(-\tau) = h(-\tau + 1) = \cdots = h(-1) = 0$).

From (A-17) and the causality of $h(l)$, it is straightforward to show that

$$f_m(0; q) = \gamma_{m,v} b(q) \quad \text{(A-20)}$$

and

$$f_m(t; q) = \gamma_{m,v} b(q)\, h(t). \quad \text{(A-21)}$$

Consequently,

$$h(t) = f_m(t; q)/f_m(0; q), \qquad t = 0, 1, \cdots, q. \quad \text{(A-22)}$$

Substituting (A-16) into this result we obtain the q-slice algorithm in (23).

F. Cumulant-Based Normal Equations

Here, as in [45], we derive (26). Starting with (A-16) and (A-17), it follows that

$$\sum_{k=0}^{p} a(k)\, C_{m,y}(\tau - k, k_0, 0, \cdots, 0) = \gamma_{m,v} \sum_{j=0}^{q} h^{(m-2)}(j - \tau)\, h(j - \tau + k_0)\, b(j). \quad \text{(A-23)}$$

From the causality of $h(l)$, it follows that the right-hand side of (A-23) is zero for $\tau > q$; hence, the result in (26). Of course, in the AR case $q = 0$, so that in this case the right-hand side of (A-23) is zero for $\tau > 0$.

G. $1\frac{1}{2}$-D Spectrum

As in [10], [16], set $\tau_1 = \tau_2 = \tau$ in (12), to see that

$$C_{3,y}(\tau) = \gamma_{3,v} \sum_{n=0}^{\infty} h(n)\, h^2(n + \tau). \quad \text{(A-24)}$$

In the z transform domain, this equation can be expressed [36] as

$$S_{3,y}(z) = \gamma_{3,v} H(z^{-1})[H(z)^* H(z)] = (\gamma_{3,v}/2\pi j)\, H(z^{-1}) \int_{\Gamma} H(u)\, H(zu^{-1}) u^{-1}\, du \quad \text{(A-25)}$$

where * denotes complex convolution, and Γ is a counterclockwise contour within the region of convergence of $H(z)$. In the frequency domain, $z = \exp(j\omega_1)$ and $u = \exp(j\omega_2)$, so that (A-25) becomes

$$S_{3,y}(\omega_1) = (\gamma_{3,v}/2\pi) \int_{-\pi}^{\pi} H(e^{-j\omega_1})\, H(e^{j\omega_2})\, H(e^{j(\omega_1 - \omega_2)})\, d\omega_2. \quad \text{(A-26)}$$

Replace ω_1 by $-\omega_1$ and set $\omega_2 = \phi$ in (13), to see that

$$S_{3,y}(-\omega_1, \phi) = \gamma_{3,v} H(-\omega_1)\, H(\phi)\, H(\omega_1 - \phi). \quad \text{(A-27)}$$

Substituting (A-27) into (A-26), we obtain the formula for the $1\frac{1}{2}$-D spectrum given in (28).

H. Spectrum/$1\frac{1}{2}$-D Spectrum Equation and the GM Equation

From (A-25), (as in [10], [16]), we know that $S_{3,y}(z) = \gamma_{3,v} H(z^{-1})\, (H(z)^* H(z))$. We also know from (9), that for a noise-free signal, $S_y(z) = \sigma_v^2 H(z)\, H(z^{-1})$. Eliminating $H(z^{-1})$ between these two equations, and letting $H_2(z) = H(z)^* H(z)$, we obtain the Spectrum/$1\frac{1}{2}$-D Spectrum Equation in (30).

The GM Equation, given in (34), is obtained by recognizing that, for the MA system in (33), $H_2(z)$ is simply the z transform of $b^2(k)$; hence, (34) is simply the inverse z transform of (30).

I. Bicepstral Formulas

For the sake of completeness, we state (12.77) and (12.78) from Oppenheim and Schafer [36], since they are the ones that

can be used to compute $h_{\min}(n)$ and $h_{\max}(n)$, respectively:

$$h_{\min}(n) = \begin{cases} 0, & n < 0 \\ e^{\hat{h}(0)}, & n = 0 \\ \hat{h}(n) + \sum_{k=0}^{n-1} \frac{k}{n} \hat{h}(k)\, h_{\min}(n-k), & n > 0 \end{cases} \quad \text{(A-28)}$$

and

$$h_{\max}(n) = \begin{cases} 0, & n > 0 \\ 1, & n = 0 \\ \hat{h}(n) + \sum_{k=n+1}^{0} \frac{k}{n} \hat{h}(k)\, h_{\max}(n-k), & n < 0. \end{cases} \quad \text{(A-29)}$$

Next, as in [37], we derive the bicepstrum-cumulant equation (44). Starting with the fact that $\hat{S}_{3,y}(z_1, z_2) = \log S_{3,y}(z_1, z_2)$, it follows that

$$\partial \hat{S}_{3,y}(z_1, z_2)/\partial z_1 = \left[1/S_{3,y}(z_1, z_2)\right]\left[\partial S_{3,y}(z_1, z_2)/\partial z_1\right]. \quad \text{(A-30)}$$

Multiplying both sides of this equation by z_1, we find that

$$S_{3,y}(z_1, z_2)\, z_1\left[\partial \hat{S}_{3,y}(z_1, z_2)/\partial z_1\right] = z_1\left[\partial S_{3,y}(z_1, z_2)/\partial z_1\right]. \quad \text{(A-31)}$$

Taking the 2-D inverse z transform of this equation, using the fact, for example, that the inverse z transform of $z_1[\partial \hat{S}_{3,y}(z_1, z_2)/\partial z_1]$ is $-k\hat{S}_{3,y}(k, l)$, we obtain the bicepstrum-cumulant equation in (44), i.e.,

$$\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} k\hat{S}_{3,y}(k, l)\, C_{3,y}(m-k, n-l) = mC_{3,y}(m, n). \quad \text{(44)}$$

In order to derive the complex cepstrum-cumulant equation (45), we must first evaluate $\hat{S}_{3,y}(k, l)$. Although Pan and Nikias [37] do this for an ARMA model, it is possible to do this directly from (43) without having to introduce a specific parametric model for $h(i)$. From (43), and the definition of $\hat{S}_{3,y}(z_1, z_2)$, we see that

$$\hat{S}_{3,y}(z_1, z_2) = \log S_{3,y}(z_1, z_2) = \log \gamma_{3,v} + \log H(z_1) + \log H(z_2) + \log H(z_1^{-1}z_2^{-1}) \quad \text{(A-32)}$$

hence

$$\hat{S}_{3,y}(k, l) = (\log \gamma_{3,v})\delta(k, l) + \hat{h}(k)\,\delta(l) + \hat{h}(l)\,\delta(k) + \mathcal{Z}_2^{-1}\left\{\hat{H}(z_1^{-1}z_2^{-1})\right\} \quad \text{(A-33)}$$

where $\mathcal{Z}_2^{-1}\{\cdot\}$ denotes the 2-D inverse z transform of $(\cdot)$.

As an aside, note that the 2-D z transform of $\hat{h}(-k)\,\delta(l-k)$ is

$$\mathcal{Z}_2\left\{\hat{h}(-k)\,\delta(l-k)\right\} = \sum_k \sum_l \hat{h}(-k)\,\delta(l-k) z_1^{-k} z_2^{-l} = \sum_k \hat{h}(-k) z_1^{-k} z_2^{-k} = \hat{H}(z_1^{-1}z_2^{-1}). \quad \text{(A-34)}$$

Consequently, (A-33) becomes

$$\hat{S}_{3,y}(k, l) = (\log \gamma_{3,v})\delta(k, l) + \hat{h}(k)\,\delta(l) + \hat{h}(l)\,\delta(k) + \hat{h}(-k)\,\delta(l-k). \quad \text{(A-35)}$$

Clearly $\hat{S}_{3,y}(k, l)$ has nonzero values only at $k = l = 0$, integer values along the k and l axes, and at the intersection of these values along the 45-degree line $k = l$.

Equation (45) is obtained by substituting (A-35) into (44) and simplifying. This is done in three stages: 1) along the k-axis for which $l = 0$; 2) along the l-axis for which $k = 0$; and, 3) along the line $k = l$. The details are straightforward and are left to the reader.

J. Facts about Kronecker Products

Here we collect a number of facts about Kronecker products that are useful in the derivations of the multichannel formulas given in Section IX [4]. The Kronecker product of a $(p \times q)$ matrix $\boldsymbol{A} = \{a_{ij}\}$ and an $(m \times n)$ matrix $\boldsymbol{B} = \{b_{ij}\}$ is the $(pm \times qn)$ matrix $\{a_{ij}\boldsymbol{B}\}$, denoted $\boldsymbol{A} \otimes \boldsymbol{B}$, i.e., $\boldsymbol{A} \otimes \boldsymbol{B} = \{a_{ij}\boldsymbol{B}\}$. Five useful facts about Kronecker products are

$$[\text{KP1}]: (\boldsymbol{A} + \boldsymbol{B}) \otimes (\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{A} \otimes \boldsymbol{D} + \boldsymbol{B} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{D} \quad \text{(A-36)}$$

$$[\text{KP2}]: (\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD} \quad \text{(A-37)}$$

$$[\text{KP3}]: (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes (\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C} \otimes \boldsymbol{D} \quad \text{(A-38)}$$

$$[\text{KP4}]: (\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1} \quad \text{(A-39)}$$

$$[\text{KP5}]: (\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'. \quad \text{(A-40)}$$

Direct consequences of (A-36) and (A-37) are the facts that

$$\left[\sum_{i=1}^{m} \boldsymbol{A}_i\right] \otimes \left[\sum_{j=1}^{n} \boldsymbol{B}_j\right] = \sum_{i=1}^{m} \sum_{j=1}^{n} \boldsymbol{A}_i \otimes \boldsymbol{B}_j \quad \text{(A-41)}$$

$$\prod_{i=1}^{n} (\boldsymbol{A}_i \otimes \boldsymbol{B}_i) = \left(\prod_{i=1}^{n} \boldsymbol{A}_i\right) \otimes \left(\prod_{i=1}^{n} \boldsymbol{B}_i\right) \quad \text{(A-42)}$$

and

$$[\otimes_{i=1}^{n} \boldsymbol{A}_i][\otimes_{i=1}^{n} \boldsymbol{B}_i] = \otimes_{i=1}^{n} \boldsymbol{A}_i \boldsymbol{B}_i. \quad \text{(A-43)}$$

K. Cumulants of Vector Processes

Proofs of (48)–(50) can be found in [53]. The proofs rely on showing that each element on the right-hand side of these equations is a legitimate cross-cumulant (e.g., as in (47) for the third-order case). The proof for $\boldsymbol{C}_{4,y}(t; t_1, t_2, t_3)$ is a bit tedious because the fourth-order cumulant contains three products of second-order terms (e.g., see (4c)). Note that for scalar processes, $\otimes$ = multiplication and $\boldsymbol{P}_p = 1$, in which case all these vector cumulant definitions reduce to their scalar counterparts.

L. Properties of Cumulants of Vector Processes

Here, as in [53], we derive (51) because it is a heavily used property in later derivations. Let $\boldsymbol{y}_i = \boldsymbol{\Lambda}_i \boldsymbol{x}_i$, $i = 1, 2, \cdots, k$ and let $\boldsymbol{C} = \text{cum}(\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_k)$. Additionally, let $\lambda_{i;j,l}$ denote the (j, l) element of $\boldsymbol{\Lambda}_i$. Then

$$C(i_1, \cdots, i_k) = \text{cum}\,(y_{1,i_1}, \cdots, y_{k,i_k})$$

$$= \text{cum}\left(\sum_{j_1=1}^{p} \lambda_{1;i_1,j_1} x_{1,j_1}, \cdots, \sum_{j_k=1}^{p} \lambda_{k;i_k,j_k} x_{k,j_k}\right)$$

$$= \sum_{j_1=1}^{p} \cdots \sum_{j_k=1}^{p} \lambda_{1;i_1,j_1} \cdots \lambda_{k;i_k,j_k}\,\text{cum}\,(x_{1,j_1}, \cdots, x_{k,j_k}) \quad \text{(A-44)}$$

where we have used cumulant properties [CP1] and [CP3] for scalar random variables in going from the first line to the second line of (A-44).

Next we expand the right-hand side of (51) to show that we get the same result as in (A-44). From the definition of Kronecker products, we have

$$(\boldsymbol{\Lambda}_1 \otimes \cdots \otimes \boldsymbol{\Lambda}_k)\,\text{cum}\,(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k)[i_1, \cdots, i_k]$$

$$= \sum_{j_1=1}^{p} \cdots \sum_{j_k=1}^{p} \lambda_{1;i_1,j_1} \cdots \lambda_{k;i_k,j_k}\,\text{cum}\,(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k)[j_1, \cdots, j_k]$$

$$= \sum_{j_1=1}^{p} \cdots \sum_{j_k=1}^{p} \lambda_{1;i_1,j_1} \cdots \lambda_{k;i_k,j_k}\,\text{cum}\,(x_{1,j_1}, \cdots, x_{k,j_k}) \quad \text{(A-45)}$$

where the last equality follows from the definition of cum $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k)$ given in Section IX. Comparing (A-45) and (A-44), we see that they are indeed the same; hence, (51) is valid.

M. Cumulant for Multichannel System

Here, as in [53], we derive the multichannel Bartlett-Brillinger-Rosenblatt formula, given in (54). Our derivation is for a time-invariant system, for which $\boldsymbol{H}(n, k) = \boldsymbol{H}(n - k)$; the extension to a time-varying system follows in a straightforward manner from our derivation.

From (52) and cumulant properties [CP3] and [CP1] (both generalized to vectors), we see that

$$\boldsymbol{C}_{k,y}(\tau_1, \cdots, \tau_{k-1})$$

$$= \text{cum}\,(\boldsymbol{y}(t), \boldsymbol{y}(t + \tau_1), \cdots, \boldsymbol{y}(t + \tau_{k-1}))$$

$$= \text{cum}\left(\sum_{u_0} \boldsymbol{H}(t - u_0)\,\boldsymbol{v}(u_0), \sum_{u_1} \boldsymbol{H}(t + \tau_1 - u_1)\,\boldsymbol{v}(u_1), \cdots, \sum_{u_{k-1}} \boldsymbol{H}(t + \tau_{k-1} - u_{k-1})\,\boldsymbol{v}(u_{k-1})\right)$$

$$= \sum_{u_0} \cdots \sum_{u_{k-1}} \text{cum}\,(\boldsymbol{H}(t - u_0)\,\boldsymbol{v}(u_0), \cdots, \boldsymbol{H}(t + \tau_{k-1} - u_{k-1})\,\boldsymbol{v}(u_{k-1}))$$

$$= \sum_{u_0} \cdots \sum_{u_{k-1}} [\boldsymbol{H}(t - u_0) \otimes \cdots \otimes \boldsymbol{H}(t + \tau_{k-1} - u_{k-1})] \cdot \text{cum}\,(\boldsymbol{v}(u_0), \cdots, \boldsymbol{v}(u_{k-1})). \quad \text{(A-46)}$$

Making use of (53), we find

$$\boldsymbol{C}_{k,y}(\tau_1, \cdots, \tau_{k-1}) = \sum_{u_0} \cdots \sum_{u_{k-1}} [\boldsymbol{H}(t - u_0) \otimes \cdots \otimes \boldsymbol{H}(t + \tau_{k-1} - u_{k-1})] \cdot \boldsymbol{\Gamma}_{kv}\delta(u_0 - u_1) \cdots \delta(u_0 - u_{k-1}) \quad \text{(A-47)}$$

but $\delta(u_0 - u_j) = 1$ if $u_j = u_0$ and $\delta(u_0 - u_j) = 0$ if $u_j \neq u_0$; hence,

$$\boldsymbol{C}_{k,y}(\tau_1, \cdots, \tau_{k-1}) = \sum_{u_0} [\boldsymbol{H}(t - u_0) \otimes \cdots \otimes \boldsymbol{H}(t + \tau_{k-1} - u_0)]\boldsymbol{\Gamma}_{kv} = \sum_{i} [\boldsymbol{H}(i) \otimes \cdots \otimes \boldsymbol{H}(i + \tau_{k-1})]\boldsymbol{\Gamma}_{kv} \quad \text{(A-48)}$$

which is the time-invariant version of (54).

N. Calculation of the Kronecker State Cumulant Vector

The solution to state equation (55), is, as is well known (e.g., [1], [28]),

$$\boldsymbol{x}(n) = \sum_{k=0}^{n-1} \boldsymbol{A}(n, k + 1)\,\boldsymbol{B}(k)\,\boldsymbol{v}(k) + \boldsymbol{A}(n, 0)\,\boldsymbol{x}(0) \quad \text{(A-49)}$$

where

$$\boldsymbol{A}(n + i, n) = \boldsymbol{\Phi}(n + i - 1) \cdots \boldsymbol{\Phi}(n), \qquad \boldsymbol{A}(n, n) = \boldsymbol{I}. \quad \text{(A-50)}$$

Initial state vector $\boldsymbol{x}(0)$ is assumed to be independent of input $\boldsymbol{v}(n)$. By cumulant property [CP5], the cumulant of the state vector $\boldsymbol{x}(n)$ is the sum of the cumulants due separately to the input and initial state. As in [53], we use the subscripts *zis* (zero initial state) and *zin* (zero input) to represent these two terms.

When the input is zero, the state vector is given by $\boldsymbol{x}(n) = \boldsymbol{A}(n, 0)\,\boldsymbol{x}(0)$, $n \geq 0$; hence,

$$\text{cum}_{\text{zin}}\,(\boldsymbol{x}(n), \boldsymbol{x}(n + i_1), \cdots, \boldsymbol{x}(n + i_{k-1})) \triangleq \boldsymbol{C}_{k,x|\text{zin}}(n; i_1, \cdots, i_{k-1}) = \text{cum}\,(\boldsymbol{A}(n, 0)\,\boldsymbol{x}(0), \boldsymbol{A}(n + i_1, 0)\,\boldsymbol{x}(0), \cdots, \boldsymbol{A}(n + i_{k-1}, 0)\,\boldsymbol{x}(0)) = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, 0)]\boldsymbol{C}_{k,x}(0; 0, \cdots, 0) \quad \text{(A-51)}$$

where we have used (51), and, by definition, $i_0 = 0$. Our goal is to express $\boldsymbol{C}_{k,x|\text{zin}}(n; i_1, \cdots, i_{k-1})$ as a function of $\boldsymbol{C}_{k,x|\text{zin}}(n - m; 0, \cdots, 0)$ (because, as we shall see below, $\boldsymbol{C}_{k,x|\text{zis}}(n; i_1, \cdots, i_{k-1})$ is also expressed in terms of a cumulant whose arguments are $(n - m; 0, \cdots, 0)$), where

$$\boldsymbol{C}_{k,x|\text{zin}}(n - m; 0, \cdots, 0) = \text{cum}_{\text{zin}}\,(\boldsymbol{x}(n - m), \cdots, \boldsymbol{x}(n - m)) = \text{cum}\,(\boldsymbol{A}(n - m, 0)\,\boldsymbol{x}(0), \cdots, \boldsymbol{A}(n - m, 0)\,\boldsymbol{x}(0)) = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n - m, 0)]\boldsymbol{C}_{k,x}(0; 0, \cdots, 0) \quad \text{(A-52)}$$

where we have again used (51) and $\otimes_{l=0}^{k-1}\boldsymbol{A}(n - m, 0)$ denotes the Kronecker product of $\boldsymbol{A}(n - m, 0)$ with itself $k - 1$ times.

From (A-50) it is easy to show that

$$\boldsymbol{A}(n + i_l, 0) = \boldsymbol{A}(n + i_l, n - m)\,\boldsymbol{A}(n - m, 0) \quad \text{(A-53)}$$

as long as $n - m$ is to the left of $n + i_l$ on the time axis. Because there are a product of k $\boldsymbol{A}(n + i_l, 0)$'s in (A-51), for us to replace all of these quantities by (A-53), in which there is a common "m", we must choose $n - m$ so that it is to the left of $n + i_1$, $n + i_2$, $\cdots$, $n + i_{k-1}$, where the i_l's can be positive or negative.

One choice of m that accomplishes this is

$$m = -\min(0, i_1, \cdots, i_{k-1}). \tag{A-54}$$

Now substitute (A-53) into (A-51), to see that

$$\begin{aligned}
&\boldsymbol{C}_{k,x|\text{zin}}(n; i_1, \cdots, i_{k-1}) \\
&\quad = \otimes_{l=0}^{k-1}[\boldsymbol{A}(n + i_l, n - m)\,\boldsymbol{A}(n - m, 0)] \\
&\qquad \cdot \boldsymbol{C}_{k,x}(0; 0, \cdots, 0) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n - m)][\otimes_{l=0}^{k-1}\boldsymbol{A}(n - m, 0)] \\
&\qquad \cdot \boldsymbol{C}_{k,x}(0; 0, \cdots, 0) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n - m)][\boldsymbol{C}_{k,x|\text{zin}}(n - m; 0, \cdots, 0)
\end{aligned} \tag{A-55}$$

where we have used (A-43) and (A-52).

The contribution to the Kronecker state cumulant vector due to the input $\boldsymbol{v}(n)$ is obtained by comparing (A-49) (when $\boldsymbol{x}(0) \triangleq \boldsymbol{0}$) with (52) and then using (54). More specifically, set $\boldsymbol{H}(n, k) = \boldsymbol{A}(n, k + 1)\,\boldsymbol{B}(k)$, where $k \le n - 1$, to see that ($i_0 = 0$)

$$\begin{aligned}
&\boldsymbol{C}_{k,x|\text{zis}}(n; i_1, \cdots, i_{k-1}) \\
&\quad = \sum_j [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, j + 1)\,\boldsymbol{B}(j)]\boldsymbol{\Gamma}_{kv}(j) \\
&\quad = \sum_j [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, j + 1)]\,\boldsymbol{B}_k(j)\boldsymbol{\Gamma}_{kv}(j) \\
&\quad = \sum_j [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n - m) \\
&\qquad \cdot \boldsymbol{A}(n - m, j + 1)]\boldsymbol{B}_k(j)\boldsymbol{\Gamma}_{kv}(j) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n - m)] \\
&\qquad \cdot \sum_j [\otimes_{l=0}^{k-1}\boldsymbol{A}(n - m, j + 1)]\,\boldsymbol{B}_k(j)\boldsymbol{\Gamma}_{kv}(j) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n - m)]\boldsymbol{C}_{k,x|\text{zis}}(n - m; 0, \cdots, 0)
\end{aligned} \tag{A-56, A-57}$$

where $\boldsymbol{B}_k(j)$ is short for $\boldsymbol{B}(j) \otimes \boldsymbol{B}(j) \otimes \cdots \otimes \boldsymbol{B}(j)$ (i.e., the Kronecker product of k $\boldsymbol{B}(j)$'s), and we have used [KP2] and (A-56).

Finally, from (A-55) and (A-57), we find that

$$\begin{aligned}
&\boldsymbol{C}_{k,x}(n; i_1, \cdots, i_{k-1}) \\
&\quad = \boldsymbol{C}_{k,x|\text{zin}}(n; i_1, \cdots, i_{k-1}) + \boldsymbol{C}_{k,x|\text{zis}}(n; i_1, \cdots, i_{k-1}) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n - m)][\boldsymbol{C}_{k,x|\text{zin}}(n - m; 0, \cdots, 0) \\
&\qquad + \boldsymbol{C}_{k,x|\text{zis}}(n - m; 0, \cdots, 0)] \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n - m)]\boldsymbol{C}_{k,x}(n - m; 0, \cdots, 0)
\end{aligned} \tag{A-58}$$

which is (57).

O. Recursive-in-Time Calculation of the Kronecker State Cumulant Vector

Here, as in [53], we derive (60) and (61). To begin we consider the zero-lag case of (60). From the definition of cumulants of vector processes, we have

$$\begin{aligned}
&\boldsymbol{C}_{k,x}(n + 1; 0, \cdots, 0) \\
&\quad = \text{cum}\,(\boldsymbol{x}(n + 1), \cdots, \boldsymbol{x}(n + 1)) \\
&\quad = \text{cum}\,(\boldsymbol{\Phi}(n)\,\boldsymbol{x}(n) + \boldsymbol{B}(n)\,\boldsymbol{v}(n), \cdots, \\
&\qquad \boldsymbol{\Phi}(n)\,\boldsymbol{x}(n) + \boldsymbol{B}(n)\,\boldsymbol{v}(n)) \\
&\quad = \text{cum}\,(\boldsymbol{\Phi}(n)\,\boldsymbol{x}(n), \cdots, \boldsymbol{\Phi}(n)\,\boldsymbol{x}(n)) \\
&\qquad + \text{cum}\,(\boldsymbol{B}(n)\,\boldsymbol{v}(n), \cdots, \boldsymbol{B}(n)\,\boldsymbol{v}(n)) \\
&\quad = \boldsymbol{\Phi}_k(n)\,\text{cum}\,(\boldsymbol{x}(n), \cdots, \boldsymbol{x}(n)) \\
&\qquad + \boldsymbol{B}_k(n)\,\text{cum}\,(\boldsymbol{v}(n), \cdots, \boldsymbol{v}(n)) \\
&\quad = \boldsymbol{\Phi}_k(n)\,\boldsymbol{C}_{k,x}(n; 0, \cdots, 0) + \boldsymbol{B}_k(n)\,\boldsymbol{\Gamma}_{k,v}(n)
\end{aligned} \tag{A-59}$$

where we have used (55), [CP5], and [CP1].

Next, we establish the temporal recursion in (61) for the nonzero positive lag case. From (A-58) and (A-59) and the fact that $m = 0$ (see (A-54)), we see that

$$\begin{aligned}
&\boldsymbol{C}_{k,x}(n + 1; i_1, \cdots, i_{k-1}) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + 1 + i_l, n + 1)]\boldsymbol{C}_{k,x}(n + 1; 0, \cdots, 0) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + 1 + i_l, n + 1)] \\
&\qquad \cdot [\boldsymbol{\Phi}_k(n)\,\boldsymbol{C}_{k,x}(n; 0, \cdots, 0) + \boldsymbol{B}_k(n)\,\boldsymbol{\Gamma}_{k,v}(n)] \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + 1 + i_l, n + 1)\,\boldsymbol{\Phi}(n)] \\
&\qquad \cdot \boldsymbol{C}_{k,x}(n; 0, \cdots, 0) + \boldsymbol{D}(i_1, \cdots, i_{k-1}) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{A}(n + 1 + i_l, n)]\boldsymbol{C}_{k,x}(n; 0, \cdots, 0) \\
&\qquad + \boldsymbol{D}(i_1, \cdots, i_{k-1}) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{\Phi}(n + i_l)\,\boldsymbol{A}(n + i_l, n)]\boldsymbol{C}_{k,x}(n; 0, \cdots, 0) \\
&\qquad + \boldsymbol{D}(i_1, \cdots, i_{k-1}) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{\Phi}(n + i_l)][\otimes_{l=0}^{k-1}\boldsymbol{A}(n + i_l, n)] \\
&\qquad \cdot \boldsymbol{C}_{k,x}(n; 0, \cdots, 0) + \boldsymbol{D}(i_1, \cdots, i_{k-1}) \\
&\quad = [\otimes_{l=0}^{k-1}\boldsymbol{\Phi}(n + i_l)]\boldsymbol{C}_{k,x}(n; i_1, \cdots, i_{k-1}) \\
&\qquad + \boldsymbol{D}(i_1, \cdots, i_{k-1})
\end{aligned} \tag{A-60}$$

where we have used [KP2], the definition of $\boldsymbol{D}(i_1, \cdots, i_{k-1})$ in (62), the facts that $\boldsymbol{A}(n + i, n) = \boldsymbol{A}(n + i, n + 1)\,\boldsymbol{\Phi}(n)$ and $\boldsymbol{A}(n + i, n) = \boldsymbol{\Phi}(n + i - 1)\,\boldsymbol{A}(n + i - 1, n)$, $i > 0$ (both of which follow from (A-50)) and (A-58) for ($m = 0$).

P. Cumulants of Harmonic Processes

If ϕ_i are independent and identically distributed over $[-\pi, \pi]$, then the following results are easily established (e.g., [38]):

$$E\{\exp(jm\phi_i)\} = 0 \tag{A-61a}$$

$$E\{\exp[j(\phi_i + \phi_j)]\} = 0 \tag{A-61b}$$

$$E\{\exp[j(\phi_i - \phi_j)]\} = \delta_{i,j} \tag{A-61c}$$

$$E\{\exp[j(\phi_i \pm \phi_k \pm \phi_l)]\} = 0 \tag{A-61d}$$

$$E\{\exp[j(\pm\phi_{i_0} + \phi_{i_1} + \phi_{i_2} + \phi_{i_3})]\} = 0 \tag{A-61e}$$

$$\begin{aligned}
&E\{\exp[j(-\phi_{i_0} - \phi_{i_1} + \phi_{i_2} + \phi_{i_3})]\} \\
&\quad = \{1, \text{ if } i_0 = i_1 = i_2 = i_3 \text{ or } (i_0 = i_2) \ne (i_1 = i_3) \\
&\qquad \text{or } (i_0 = i_3) \ne (i_1 = i_2); 0, \quad \text{otherwise}\}.
\end{aligned} \tag{A-61f}$$

Making use of (A-61a)–(A-61c), it is straightforward to show, as in [45], that if $a = \exp(j\phi)$, then

$$E\{a\} = 0 \tag{A-62a}$$

$$\text{cum}(a, a) = E\{a^2\} = 0 \tag{A-62b}$$

$$\text{cum}(a^*, a) = E\{|a|^2\} = 1 \tag{A-62c}$$

$$\text{cum}(a, a, a) = E\{a^3\} = 0 \tag{A-62d}$$

$$\text{cum}(a^*, a, a) = E\{a^*a^2\} = 0 \tag{A-62e}$$

$$\text{cum}(a, a, a, a) = E\{a^4\} - 3E\{a^2\}E\{a^2\} = 0 \tag{A-62f}$$

$$\text{cum}(a^*, a, a, a) = E\{a^*a^3\} - 3E\{|a|^2\}E\{a^2\} = 0 \tag{A-62g}$$

$$\text{cum}(a^*, a^*, a, a) = E\{|a|^4\} - |E\{a^2\}|^2 - 2E\{|a|^2\}E\{|a|^2\} = -1. \tag{A-62h}$$

This demonstrates that the third-order cumulant of a is identically zero, and that of the three ways to define the fourth-order cumulant of a only (A-62h) yields nonzero real values.

Next, as in [55], we derive (70). Substituting (66) into (69), and using [CP3] and [CP1], we find that

$$\begin{aligned}
&\text{cum}\left(y^*(n), y^*(n+\tau_1), y(n+\tau_2), y(n+\tau_3)\right) \\
&= \text{cum}\Big(\sum_i a_i^*(n)s_n^*(\omega_i), \\
&\quad \sum_j a_j^*(n+\tau_1)s_{n+\tau_1}^*(\omega_j), \sum_k a_k(n+\tau_2)s_{n+\tau_2}(\omega_k), \\
&\quad \sum_l a_l(n+\tau_3)s_{n+\tau_3}(\omega_l)\Big) \\
&= \sum_i \sum_j \sum_k \sum_l s_n^*(\omega_i)s_{n+\tau_1}^*(\omega_j)s_{n+\tau_2}(\omega_k)s_{n+\tau_3}(\omega_l) \\
&\quad \cdot \text{cum}\left(a_i^*(n), a_j^*(n+\tau_1), a_k(n+\tau_2), a_l(n+\tau_3)\right).
\end{aligned} \tag{A-63}$$

Because the $a_i(n)$ are mutually independent, cum $(a_i^*(n), a_j^*(n+\tau_1), a_k(n+\tau_2), a_l(n+\tau_3)) = 0$, unless $i = j = k = l$, in which case (A-63) reduces to (70).

Next, as in [55], we derive (71). Equation (71) is easily obtained from (70). Just below (67) we showed how (67) is the special case of (66), when $s_n(\omega_i) = \exp(jn\omega_i)$ and $a_i(n) = a_i = \alpha_i \exp(j\phi_i)$; hence, in (70)

$$\begin{aligned}
&s_n^*(\omega_i)s_{n+\tau_1}^*(\omega_i)s_{n+\tau_2}(\omega_i)\, s_{n+\tau_3} \\
&\quad = \exp\left[j(-\tau_1 + \tau_2 + \tau_3)\omega_i\right]
\end{aligned} \tag{A-64}$$

and

$$\begin{aligned}
C_{4,a_i}(\tau_1, \tau_2, \tau_3) &= \text{cum}\left(a_i^*(n), a_i^*(n+\tau_1), a_i(n+\tau_2), a_i(n+\tau_3)\right) \\
&= \text{cum}\left(\alpha_i^* \exp(-j\phi_i), \alpha_i^* \exp(-j\phi_i), \alpha_i \exp(j\phi_i), \alpha_i \exp(j\phi_i)\right) \\
&= \alpha_i^4\, \text{cum}\left(\exp(-j\phi_i), \exp(-j\phi_i), \exp(j\phi_i), \exp(j\phi_i)\right) \\
&= -\alpha_i^4
\end{aligned} \tag{A-65}$$

where we have used [CP1] and (A-62h). Substituting (A-64) and (A-65) into (70), we obtain the desired result in (71).

The derivation of (72) is so simple (and should be well known) that we leave it to the reader. We conclude this section with the derivation of (73), as in [55]. Not too surprisingly, our approach will be to reexpress (68) as

$$y(n) = \sum_{i=1}^{p} \tfrac{1}{2}\alpha_i\left\{\exp\left[j(n\omega_i + \phi_i)\right] + \exp\left[-j(n\omega_i + \phi_i)\right]\right\} \tag{A-66}$$

and to then compute the fourth-order cumulant of p pairs of complex exponentials. The following result, which is proved in [45], [55], permits us to do this: Let $s = \exp(j\phi)$, where ϕ is uniformly distributed over $[-\pi, \pi]$, and let $a_l (l = 0, 1, 2, 3)$ be constants; then,

$$\begin{aligned}
&\text{cum}\left(a_0 s + a_0^* s^*, a_1 s + a_1^* s^*, a_2 s + a_2^* s^*, a_3 s + a_3^* s^*\right) \\
&\quad = -2\,\text{Re}\left(a_0 a_1 a_2^* a_3^* + a_0 a_1^* a_2^* a_3 + a_0 a_1^* a_2 a_3^*\right).
\end{aligned} \tag{A-67}$$

Since the ϕ_i's in (A-66) are independent, the pairs are independent of each other; hence, the cumulant of $y(n)$ is the sum of the cumulants due to each of the pairs. Consequently, it suffices to consider a single pair at frequency $\pm\omega$ and amplitude α. Let $a_i = \alpha_i \exp[j\omega(n + \tau_i)]/2$, $i = 0, 1, 2, 3$, with $\tau_0 = 0$, and $s = \exp(j\phi_i)$. Then, for example, $-2\,\text{Re}\,(a_0 a_1 a_2^* a_3^*) = -\alpha_i^4 \cos(\tau_i - \tau_2 - \tau_3)/8$, which is the first term on the right-hand-side of (73). The remaining two terms on the right-hand-side of (73) are obtained by evaluating the corresponding terms in (A-67).

Q. Derivation of Generalized GM-Equation and (80)

Tugnait [63] and Friedlander and Porat [9] have generalized the GM equation from a diagonal slice result to a 1-D slice result. Setting $\tau_1 = \tau$ and $\tau_2 = \tau + m$ in (12) (for $k = 3$), they show that

$$C_{3,y}(\tau, \tau + m) = \gamma_{3,v} \sum_{n=0}^{\infty} h(n)h(n+\tau)h(n+\tau+m). \tag{A-68}$$

In the z transform domain, this equation can be expressed as

$$S_{3,y}(z; m) = \gamma_{3,v} H(z^{-1})\left[H(z) * z^m H(z)\right] \tag{A-69}$$

which should be compared with (A-25). As in the derivation of (30), we now compare (A-69) with the spectrum equation $S_y(z) = \sigma_v^2 H(z)H(z^{-1})$ to conclude that

$$\left[H(z) * z^m H(z)\right]S_y(z) = (\sigma_v^2/\gamma_{3,v})H(z)S_{3,y}(z; m). \tag{A-70}$$

Clearly, when $m = 0$, (A-70) reduces to (30). The generalized GM equation is obtained by recognizing that, for the MA system in (33), $H(z) * z^m H(z)$ is the z transform of $b(k)b(k+m)$; hence, the inverse z transform of (A-70) for an MA(q) system is

$$\begin{aligned}
&\sum_{k=0}^{q} b(k)b(k+m)r_y(\tau - k) \\
&\quad = (\sigma_v^2/\gamma_{3,v}) \sum_{k=0}^{q} b(k)C_{3,y}(\tau - k, \tau - k + m).
\end{aligned} \tag{A-71}$$

This reduces to the GM equation in (34) when $m = 0$.

Equation (80), obtained by Tugnait [63], is obtained by setting $m = q$ in (A-71). Note that only one term survives on the left-hand side of (A-71), namely $b(0)b(q)r_y(\tau) = b(q)r_y(\tau)$, since

$b(0) \triangleq 1$. An extension of (80) to fourth-order cumulants also appears in the journal version of [63].

Note, also, that (A-69) can be ratioed for $m = m_1$ and $m = m_2$. Doing this eliminates $\gamma_{3,v}H(z^{-1})$ and leads to another interesting relationship, developed by Tugnait in [63].

Acknowledgment

This paper is an outgrowth of material in [30] and [31], and a 90 minute tutorial presented by the author at the 1989 Vail Workshop on Higher-Order Spectral Analysis. The author wishes to acknowledge the contributions of L. Wang, for the simulation results in Example 1 of Section XII; C. Y. Ngo for the example in Table A-1 and details of the proofs of properties of cumulants of scalar processes in Appendix A; and, A. Swami for a very careful reading of the paper which led to quite a few modifications, and to our many fruitful discussions on a wide range of topics related to the contents of this paper.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[2] M. S. Bartlett, *An Introduction to Stochastic Processes*. London, U.K.: Cambridge Univ. Press, 1955.

[3] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden-Day, 1970.

[4] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 772–781, 1978.

[5] D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of kth-order spectra," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, pp. 189–232, 1967.

[6] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–938, 1982.

[7] J. W. Dalle Molle and M. J. Hinich, "The trispectrum," in *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, CO, 1989, pp. 68–72.

[8] B. Friedlander and B. Porat, "Adaptive IIR algorithms based on high-order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 485–495, 1989.

[9] ——, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from high-order moments," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 27–35, 1990.

[10] G. B. Giannakis, "Signal processing using higher-order statistics," Ph.D. dissertation, Dep. Elec. Eng., Univ. Southern California, Los Angeles, CA, 1987.

[11] ——, "Cumulants: A powerful tool in signal processing," *Proc. IEEE*, vol. 75, pp. 1333–1334, 1987.

[12] ——, "On the identifiability of non-Gaussian ARMA models using cumulants," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 18–26, 1990.

[13] G. B. Giannakis, Y. Inouye, and J. M. Mendel, "Cumulant based identification of multichannel moving-average models," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 783–787, July 1989.

[14] G. B. Giannakis and J. M. Mendel, "Stochastic realization of non-minimum phase systems," in *Proc. Amer. Control Conf.*, vol. 2, Seattle, WA, 1986, pp. 1254–1259.

[15] ——, "ARMA order determination via higher-order statistics," presented at the MTNS Conf., Phoenix, AZ, 1987.

[16] ——, "Identification of non-minimum phase systems using higher-order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 360–377, 1989.

[17] ——, "Cumulant-based order determination of non-Gaussian ARMA models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1411–1423, 1990.

[18] G. B. Giannakis and A. Swami, "New results on state-space and input-output identification of non-Gaussian processes using cumulants," in *Proc. SPIE-87, Advanced Algorithms and Architectures for Signal Processing II*, San Diego, CA, 1987, pp. 199–204.

[19] ——, "On estimating non-causal non-minimum phase ARMA models of non-Gaussian processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 478–495, 1990.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The John Hopkins Univ. Press, 1983.

[21] H. P. Graf, L. D. Jackel, and W. E. Hubbard, "VLSI implementation of a neural network model," *Computer*, pp. 41–49, 1988.

[22] M. Huzii, "Estimation of coefficients of an AR process by using higher-order moment," *J. Time Series Anal.*, vol. 2, pp. 87–93, 1981.

[23] V. P. Leonov and A. N. Shiryaev, "On a method of calculation of semi-invariants," *Theory Prob. Appl.*, vol. 4, pp. 319–328, 1959.

[24] K-S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," in *The Annals of Statistics*, vol. 10, pp. 1195–1208, 1982.

[25] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: The MIT Press, 1983.

[26] A. W. Lohmann, G. Weigelt, and B. Wirnitzer, "Speckle masking in astronomy: Triple correlation theory and applications," *Appl. Opt.*, vol. 22, pp. 4028–4037, 1983.

[27] A. W. Lohmann and B. Wirnitzer, "Triple correlations," *Proc. IEEE*, vol. 72, pp. 889–901, 1984.

[28] J. M. Mendel, *Lessons in Digital Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[29] ——, "Use of higher-order statistics in signal processing and system theory: A short perspective," presented at *8th Int'l. Symp. on MTNS*, Phoenix, AZ, 1987; in *Linear Circuits, Systems and Signal Processing: Theory and Applications.*, C. I. Byrnes, C. F. Martin, and R. E. Saeks, Eds. New York, NY: North Holland, 1988, pp. 217–224.

[30] ——, "Use of higher-order statistics in signal processing and system theory: An update," in *Proc. SPIE Conf. on Advanced Algorithms and Architectures for Signal Processing III*, San Diego, CA, 1988, pp. 126–144.

[31] ——, "A compendium of new theoretical results associated with using higher-order statistics in signal processing and system theory," in *Proc. of Indo-U.S. Workshop on Spectral Analysis in One or Two Dimensions*, New Delhi, India, Nov. 1989.

[32] J. M. Mendel and L. Wang, "Identification of moving average systems using higher-order statistics and learning," in *Neural Networks: A Dynamical Approach to Machine Intelligence*, B. Kosko, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1990; see also, "Cumulant-based parameter estimation using stuctured networks," *IEEE Trans. Neural Networks*, vol. 2, pp. 73–83, Jan. 1991.

[33] C. L. Nikias and H. Chiang, "Higher-order spectrum estimation via noncausal autoregressive modeling and deconvolution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1911–1913, Dec. 1988.

[34] C. L. Nikias and R. Pan, "Non-minimum phase system identification via cepstrum modeling of higher-order moments," in *Proc. ICASSP-87*, Dallas, TX, 1987, pp. 980–983.

[35] C. L. Nikias and M. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, pp. 869–891, 1987.

[36] A. V. Oppenheim and R. W. Shafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[37] R. Pan and C. L. Nikias, "The complex cepstrum of higher-order moments," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 186–205, 1988.

[38] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1984.

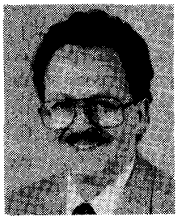
[39] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *Int. J. Adaptive Control and Signal Processing*, vol. 3, pp. 191–229, 1989.

[40] S. Prasad, R. T. Williams, A. K. Mahalanabis, and L. H. Sibul, "A transform-based covariance differencing approach for some classes of parameter estimation problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 631–641, 1988.

[41] M. B. Priestley, *Spectral Analysis and Time Series*. London, UK: Academic Press, 1981.

[42] M. Raghuveer and C. L. Nikias, "Bispectrum estimation: A parametric approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 1213–1230, 1985.

[43] M. Rosenblatt, *Stationary Sequences and Random Processes.* Boston, MA: Birkhauser, 1985.
[44] M. Rosenblatt and J. W. Van Ness, "Estimation of the bispectrum," *Ann. Math. Stat.*, pp. 420–436, 1965.
[45] A. Swami, "System identification using cumulants," Ph.D. dissertation, USC SIPI Report 140, Dep. Elec. Eng.-Syst., Univ. Southern California, Los Angeles, CA, 1988.
[46] A. Swami, G. B. Giannakis, and J. M. Mendel, "A unified approach to modeling multichannel ARMA processes," in *Proc. ICASSP-89*, Glascow, Scotland, 1989, pp. 2182–2185.
[47] ——, "Modeling of multi-dimensional non-Gaussian processes using cumulants," *J. Multidimensional Signals and Systems*, vol. 1, pp. 1–37, Mar. 1990.
[48] A. Swami and J. M. Mendel, "Closed-form recursive estimation of MA coefficient using autocorrelations and third-order cumulants," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1794–1795, Nov. 1989.
[49] ——, "ARMA parameter estimation using only output cumulants," in *Proc. IV IEEE ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, MN, 1988, pp. 193–198; also, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, July 1990, pp. 1257–1265.
[50] ——, "Computation of cumulants of ARMA processes," in *Proc. ICASSP-89*, Glascow, Scotland, 1989, pp. 2318–2321.
[51] ——, "AR identifiability using cumulants," in *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, CO, 1989, pp. 1–6.
[52] ——, "Identifiability of the parameters of an ARMA process using cumulants," *IEEE Trans. Automat. Contr.*, 1991.
[53] ——, "Time and lag recursive computation of cumulants from a state space model," USC-SIPI Rep. 121, Los Angeles, CA, 1988; also, in *IEEE Trans. Automat. Contr.*, vol. 35, pp. 4–17, 1990.
[54] ——, "Cumulant-based approach to the harmonic retrieval problem," in *Proc. ICASSP-88*, New York, NY, 1988, pp. 2264–2267.
[55] ——, "Cumulant-based approach to the harmonic retrieval and related problems," *IEEE Trans. Acoust., Speech, Signal Processing*, 1991.
[56] M. A. Tekalp and A. T. Erdem, "Higher-order spectrum factorization in one and two dimensions with applications in signal modeling and nonminimum phase system identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1537–1549, Oct. 1989.
[57] P. V. Tryon, "The bispectrum and higher-order spectra: A bibliography," NBS Tech. Note 1036, National Bureau of Standards, Boulder, CO, 1981.
[58] J. Tugnait, "Identification of nonminimum phase linear stochastic systems," *Automatica*, vol. 22, pp. 454–464, 1986.
[59] ——, "Order reduction of SISO nonminimum phase stochastic systems," *IEEE Trans. Automat. Contr.*, vol. 31, pp. 623–632, July 1986.
[60] ——, "Realization and reduction of SISO nonminimum-phase parametric models of non-Gaussian linear processes: modeling, parameter estimation and system identification," presented at *MTNS Conf.*, Phoenix, AZ, 1987.
[61] ——, "Identification of linear stochastic systems via second—and fourth—order cumulant matching," *IEEE Trans. Inform. Theory*, vol. 33, pp. 393–407, 1987.
[62] ——, "Fitting noncausal AR signal plus noise models to noisy non-Gaussian linear processes," *IEEE Trans. Automat. Contr.*, vol. 32, pp. 547–552, 1987.
[63] ——, "Approaches to FIR system identification with noisy data using higher-order statistics," in *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, CO, 1989, pp. 13–18; also, in *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1307–1317, July 1990.
[64] A. P. Petropulu and C. L. Nikias, "The complex cepstrum and bicepstrum: Analytic performance evaluation in the presence of Gaussian noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1246–1256, July 1990.
[65] E. Parzen, "Time series analysis for models of signal plus white noise," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 233–257.
[66] H. Akaike, "Note on higher order spectra," *Annals of the Institute of Statistical Mathematics*. Tokyo, vol. 18, 1966, pp. 123–126.
[67] ——, "On the use of non-Gaussian processes in the identification of a linear dynamic system," Research Memo 3, Institute of Statistical Mathematics, Tokyo, 1966.
[68] J. K. Tugnait, "Recovering the poles from third-order cumulants of system output," *IEEE Trans. on Automat. Conr.*, vol. 34, pp. 1085–1089, 1989.



Jerry M. Mendel (Fellow, IEEE) received the B.S. degree in mechanical engineering and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1959, 1960, and 1963, respectively.

His experience includes teaching courses in electrical engineering at the Polytechnic Institute of Brooklyn, from 1960 to 1963, and he has also had various consulting positions. From July 1963 to January 1974 he was with McDonnell Douglas Astronautics Company. Currently he is Professor and Chairman of Electrical Engineering-Systems at the University of Southern California in Los Angeles. He teaches courses in estimation theory and deconvolution, and was Director of the USC Geo-Signal Processing Program (1980–1983). He has published over 230 technical papers and is author of the monographs *Maximum-Likelihood Deconvolution: A Journey into Model-Based Signal Processing* (Springer-Verlag, 1990) and *Optimal Seismic Deconvolution: An Estimation-Based Approach* (Academic Press, 1983), the texts *Lessons in Digital Estimation Theory* (Prentice-Hall, 1987), and *Discrete Techniques of Parameter Estimation: The Equation Error Formulation* (Dekker, 1973), and is co-editor (with K. S. Fu (deceased)) of *Adaptive, Learning and Pattern Recognition Systems* (Academic Press, 1970). He is also author of the IEEE Individual Learning Program, *Kalman Filtering, and Other Digital Estimation Techniques*. He served as Editor of the IEEE Control Systems Society's IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and is on the Editorial Board of the PROCEEDINGS OF THE IEEE.

Dr. Mendel is a Distinguished Member of the IEEE Control Systems Society, a member of the IEEE Geoscience and Remote Sensing Society, the IEEE Signal Processing Society, the Society of Exploration Geophysicists, the European Association for Signal Processing, Tau Beta Pi and Pi Tau Sigma, and a registered Professional Control Systems Engineer in California. He was President of the IEEE Control Systems Society in 1986. He received the SEG 1976 Outstanding Presentation Award for a paper on the application of Kalman Filtering to deconvolution, the 1983 Best Transactions Paper Award for a paper on maximum-likelihood deconvolution in the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, a Phi Kappa Phi book award for his research monograph on seismic deconvolution, a 1985 Burlington Northern Faculty Achievement Award, and a 1984 IEEE Centennial Medal.