# USC-SIPI REPORT #147

## Modeling and Parameter Estimation of Multidimensional Non-Gaussian Process Using Cumulants

by

**Ananthram Swami, Georgios B. Giannakis and Jerry M. Mendel**

**August 1989**

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

USC-SIPI REPORT #147

# Modeling and Parameter Estimation of Multidimensional Non-Gaussian Processes Using Cumulants

by

A. Swami, G.B. Giannakis and J.M. Mendel

August 1989

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
Powell Hall of Engineering
University Park/MC-0272
Los Angeles, CA 90089 U.S.A.

USC-SIPI Report # 147

# Modeling and Parameter Estimation Of Multidimensional Non-Gaussian Processes Using Cumulants

A. Swami, G.B. Giannakis and J.M. Mendel

August 1989

Signal and Image Processing Institute
Department of Electrical Engineering-Systems
University of Southern California
Los Angles, CA 90089-0781

A. Swami and J.M. Mendel are with the Department of Electrical Engineering-Systems, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-0781. G.B. Giannakis is with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22901. This work was performed at the University of Southern California, Los Angeles, under National Science Foundation Grant ECS-8602531 and Naval Ocean Systems Center Contract N6601-85-D-0203. The second author was partly supported by Univ. of Virginia Engr. Research Initiation Grant 6-42410, and HDL Contract 5-25227. Parts of this report were presented at ICASSP-88, New York, NY, April 1988, and at the IV IEEE ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, MN, August 1988. An earlier version of this report was submitted to the *Proceedings of the IEEE* Special Issue on Multidimensional Signal Processing, Jan 1989; revised July 1989.

## Abstract

Extending the notion of second-order correlations, we define the *cumulants* of stationary non-Gaussian random fields, and demonstrate their potential for modeling and reconstruction of multidimensional signals and systems. Cumulants and their Fourier transforms called *polyspectra* preserve complete amplitude and phase information of a multidimensional linear process, even when it is corrupted by additive colored Gaussian noise of unknown covariance function. Relying on this property, phase reconstruction algorithms are developed using polyspectra, which can be computed via a 2-D FFT-based algorithm. Additionally, consistent ARMA parameter estimators are derived for identification of linear space-invariant multidimensional models which are driven by unobservable, i.i.d., non-Gaussian random fields. Contrary to autocorrelation based multidimensional modeling approaches, when cumulants are employed, the ARMA model is allowed to be non-minimum phase, asymmetric non-causal or non-separable.

# Contents

# 1 Introduction

Multidimensional (especially 2-D) statistical models, and their phase reconstruction, are important in image processing tasks such as restoration, spectral estimation, coding, texture synthesis, and classification, [17]. Other applications include frequency-wavenumber estimation [5], biomedical signal processing, [16, 33], estimation of multidimensional harmonics in noise (simultaneous estimation of frequency and angle of arrival in array processing), [20], target detection, astronomical interferometry, [25], and wavelet estimation in 2-D and 3-D geophysical signal processing, [36].

Although the autocorrelation captures the amplitude characteristics, it fails to convey the complete phase information about the image point spread function (p.s.f.). For example, 2-D all-pass factors, and non-symmetric non-causal model parameters cannot be identified with autocorrelation statistics. Because the phase spectrum carries valuable information about edges in the image, [29], multidimensional phase reconstruction algorithms become particularly important.

The parameters of white Gaussian noise driven models may be efficiently estimated using approximate maximum likelihood and least-squares approaches, [6, 18, 19]. If the multidimensional parametric model is causal, recursive processing is possible. Since least squares estimates yield inconsistent estimates of the AR parameters of non-unilateral models, a conditional symmetric bilateral Markov model is introduced in [6, 19]. In [18], 2-D non-causal AR models driven by colored, but not necessarily MA, Gaussian noise are introduced. Parameters of causal and non-causal AR models are also estimated using 2-D linear prediction techniques, [17]. Most of these autocorrelation based parametric approaches deal with Gaussian processes; parameter estimates obtained with these methods are unique only with respect to the class of spectrally equivalent models. Thus, to ensure unique estimates, non-causal models are also constrained to be symmetric, [17, 46]; such a constraint may not be realistic for p.s.f. models encountered in exploration seismology, or, for p.s.f. models which are due to motion blurs and defocusing.

Motivated by the inability of the autocorrelation to provide complete phase information, and characterization of non-Gaussian processes, higher-order statistics, namely *cumulants*, have been recently proposed for phase reconstruction, [22, 26], and identification of non-minimum phase and non-causal 1-D parametric models (see [10]-[14], [22], [27], [31], [39]-[43], and references therein).

The *objectives* of this report are to (i) introduce cumulants of multidimensional non-Gaussian processes, (ii) develop 2-D cumulant based phase and amplitude reconstruction algorithms, and (iii) derive parameter estimation approaches which can handle general non-minimum phase and asymmetric non-causal ARMA models.

A fair amount of work employing third-order cumulants has been reported in the optics literature, [2, 24, 28]. However, optical signals are analog and deterministic, and the non-parametric reconstruction algorithms reported are only appropriate when multiple independent data records are available, as is the case in astronomical speckle interferometry, [25].

In this report, we deal with discrete-time, multidimensional non-Gaussian linear processes, and develop parametric and non-parametric modeling algorithms which apply both to the deterministic as well as to the stochastic cases. Introductory material on cumulants of 2-D non-Gaussian and linear processes, and non-parametric phase reconstruction algorithms may be found in [37].

The algorithms reported in this report expand upon results originally presented in [40]. Some of the results in this report have been recently tested in several image processing problems, such as 2-D non-minimum phase estimation [35], non-linear interactions in isotropic turbulence [23], detection of quadratic phase coupling [34], 2-D signal reconstruction [11],[8], image coding [15], image sequence analysis [38], and detection and classification [48].

The *organization* and *main contributions* of the report are as follows:

1. In Section 2, cumulants of multidimensional stationary processes, and their Fourier transforms called polyspectra, are introduced. Among other properties their phase sensitivity and immunity to additive colored Gaussian noise, are highlighted. A computationally efficient 2-D FFT based algorithm is proposed for bispectral calculations.

2. Section 3 deals with cumulants, polyspectra and polycepstra of non-Gaussian multidimensional *linear* processes.

3. Section 4 describes several non-parametric amplitude and phase reconstruction algorithms.

4. 2-D ARMA parameter estimation using cumulants is the topic of Section 5 (AR and MA models are treated as special cases). The cumulant based parameter estimation algorithms developed in Sections 5.1 and 5.2 are appropriate for identifying 2-D causal non-minimum phase ARMA models which may even include all-pass factors. A unified parameter estimation approach which allows identification of even non-symmetric non-causal ARMA models is presented in Section 5.4.

5. Consistency, order determination, and noise issues are discussed in Section 6. Finally, conclusions are drawn in Section 7.

# 2 Cumulants and Polyspectra

*Cumulants* (and their Fourier transforms, *polyspectra*), are phase-sensitive sequences, and in the 1-D context, they have been used in a variety of applications that involve linear non-Gaussian processes or non-linear Gaussian processes (see [3, 4, 11, 22, 26, 27, 39, 47, 50], and references therein). The importance of cumulants in the context of parameter estimation of linear systems stems from two important properties: 1) cumulants retain both amplitude and phase information of the linear system; and 2) cumulants of Gaussian processes are zero.

## 2.1 Definitions: Random Variables

Let $\Phi_y(v) = E\exp(jvy)$ denote the (first) characteristic function of the random variable (r.v.) $y$. Then, $M_{ky}$, the $k$-th order moment of $y$ is the coefficient of $v^k$ in the Taylor series expansion of $\Phi_y(v)$, i.e.,

$$\Phi(v) = E\{\exp(jvy)\} = 1 + \sum_{k=1}^{\infty} \frac{(jv)^k}{k!} M_{ky}$$

Similarly, the second characteristic function, $\Psi_y(v) = \ln \Phi_y(v)$ is the cumulant-generating function of the r.v., $y$, i.e.,

$$\Psi(v) = \ln\left(E\{\exp(jvy)\}\right) = \sum_{k=1}^{\infty} \frac{(jv)^k}{k!} C_{ky}$$

where $C_{ky}$ is the $k$-th order cumulant of $y$.

Let the random variables (r.v.'s) $y_1, ..., y_n$ have the joint characteristic function,

$$\Phi(v_1, ..., v_n) = E\{\exp[j(v_1 y_1 + \cdots + v_n y_n)]\}$$

Then, the mixed moment of order $k = k_1 + \cdots + k_n$ is given by

$$M_{k_1,...,k_n} := (-j)^k \frac{\partial^{k_1}}{\partial v_1^{k_1}} \cdots \frac{\partial^{k_n}}{\partial v_n^{k_n}} \Phi(v_1, ..., v_n)\Big|_{v_1 = ... = v_n = 0}$$

Similarly, the mixed cumulant of order $k = k_1 + \cdots + k_n$ is given by

$$C_{k_1,...,k_n} := (-j)^k \frac{\partial^{k_1}}{\partial v_1^{k_1}} \cdots \frac{\partial^{k_n}}{\partial v_n^{k_n}} \ln(\Phi(v_1, ..., v_n))\Big|_{v_1 = ... = v_n = 0}$$

The function $\Psi(v_1, ..., v_n) := \ln(\Phi(v_1, ..., v_n))$ is also called the second characteristic function or the cumulant generating function of the set of r.v.'s.

If $\mathbf{y} = col[y_1, ..., y_n]$ is a Gaussian random vector with mean $\mathbf{m}$ and covariance $\Sigma$, then, with $\mathbf{v} = [v_1, ..v_n]'$,

$$\begin{aligned} \Phi(\mathbf{v}) &= \exp(j\mathbf{v}^T\mathbf{m} - \frac{1}{2}\mathbf{v}^T\Sigma\mathbf{v}) \\ \Psi(\mathbf{v}) &= j\mathbf{v}^T\mathbf{m} - \frac{1}{2}\mathbf{v}^T\Sigma\mathbf{v} \end{aligned}$$

Hence, the cumulants of order greater than two of a Gaussian process are identically zero.

Let $\Omega$ denote the set of all partitions of the integers $(1, ..., n)$ into $r$ groups, $1 \leq r \leq n$. Let $p = \{g_{p,1}, ..., g_{p,n_{gp}}\}$ denote a particular partition, where the $g_i$'s are the groups in the partition, and $n_{gp}$ denotes the number of groups in the partition $p$. For example, the set of partitions corresponding to $n = 3$, is given by $\{(1,2,3)\}$, $\{(1),(2,3)\}$, $\{(2),(3,1)\}$, $\{(3),(1,2)\}$, $\{(1),(2),(3)\}$. For $n = 4$, there are fifteen different partitions, and for $n = 5$, there are 52 different partitions. Then, the $n$-th order joint cumulant, with $k_i = 1, i = 1, ..., n$, may be expressed in terms of the joint moments of order less than or equal to $k$, via [3, 4]

$$C_{y_1,...,y_n} = \sum_{p \in \Omega} (-1)^{n_{gp}-1} (n_{gp} - 1)! E\{\prod_{i \in g_{p,1}} y_i\} \cdots E\{\prod_{i \in g_{p,n_{gp}}} y_i\}$$

In particular, we have

$$\begin{aligned}
C_1 &= M_1 \\
C_2 &= M_2 - M_1^2 \\
C_3 &= M_3 - 3M_2M_1 + 2M_1^3 \\
C_4 &= M_4 - 4M_3M_1 - 3M_2^2 + 12M_1^2M_2 - 6M_1^4 \\
C_5 &= M_5 - 5M_1M_4 - 10M_2M_3 + 20M_1^2M_3 + 30M_1M_2^2 - 60M_1^3M_2 + 24M_1^5
\end{aligned}$$

If $M_1 = 0$, we have $C_i = M_i$, $i = 1, 2, 3$, and

$$\begin{aligned}
C_4 &= M_4 - 3M_2^2 = M_4 - 3C_2^2 \\
C_5 &= M_5 - 10M_2M_3 = M_5 - 10C_2C_3 \\
C_6 &= M_6 - 15M_2M_4 - 10M_3^2 + 30M_2^3 \\
&= M_6 - 15C_2C_4 - 10C_3^2 - 15C_2^3
\end{aligned}$$

## 2.2 Properties

Some important properties of cumulants are, [37, Ch II]:

[**CP1.**] If $\lambda_i$, $i = 1, ..., k$ are constants, and $x_i$, $i = 1, ..., k$ are r.v.'s, then,

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

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

$$\text{cum}(x_1, ..., x_k) = \text{cum}(x_{j+1}, ..., x_k, x_1, ..., x_j)$$

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

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

[**CP4.**] If $\alpha$ is a constant, then

$$\text{cum}\,(\alpha + z_1, ..., z_k) = \text{cum}\,(z_1, ..., z_k)$$

[**CP5.**] If the r.v.'s $\{x_i\}_{i=1}^k$ are independent of the r.v.'s $\{y_i\}_{i=1}^k$, then,

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

[**CP6.**] If a subset of the r.v.'s $\{x_i\}_{i=1}^k$ is independent of the rest, then

$$\text{cum}\,(x_1, ..., x_k) = 0$$

These properties are readily proved by using the formal definition of cumulants as coefficients of the second characteristic function.

The cumulants of a random variable summarize information about the shape of the probability density or distribution function (pdf); the second-order cumulant (the variance) gives us an idea of the spread of the pdf; the third-order cumulant (also called skewness) is a measure of non-symmetry, since the skewness of a symmetric pdf is zero; the fourth-order cumulant (also called kurtosis) is a measure of tailweight. In the multivariate case, the joint cumulants are also measures of joint dependence (see property [CP5]).

## 2.3 Definitions and Properties: Random Processes

Let $y(m,n)$ be a real-valued stationary random field, and vectors $\underline{v}$ and $\underline{y}$ be defined as, $\underline{v} = [v_1, v_2, ..., v_k]^T$, and $\underline{y} = [y(m,n), y(m+i_1, n+j_1), ..., y(m+i_{k-1}, n+j_{k-1})]^T$, where $T$ denotes transposition. Then, the $k$-*th order cumulant* of the random field $y(m,n)$ is defined as the coefficient of $(v_1 v_2 ... v_k)$ in the Taylor series expansion of the cumulant-generating function, $K(\underline{v}) = \ln E\{\exp(\underline{v}^T \underline{y})\}$ . The $k$-th order cumulant of $y(m,n)$ is a function of $2(k-1)$ lag variables. The cumulants of higher-dimensional processes are defined similarly, and are functions of $d(k-1)$ lag variables, where $d$ denotes the dimensionality of the process.

In order to reduce notational complexity, we will let bold lower-case letters represent ordered $d$-tuples (or $d$-element row vectors) of the form $\mathbf{m} := (m_1, m_2, ..., m_d)$, where $d$ is the dimensionality of the process under consideration. We will let $\mathbf{0} := (0, 0, ..., 0)$, $\mathbf{1} := (1, 1, ..., 1)$; $\mathbf{m} + \mathbf{i}$ will denote the tuple $(m_1 + i_1, m_2 + i_2, ..., m_d + i_d)$; $\mathbf{m} \leq \mathbf{n}$ will denote $m_i \leq n_i$, $i = 1, ..., d$; and $i\mathbf{m}$ denotes the tuple $(im_1, im_2, ..., im_d)$; vectors will be represented by underscored letters; for 2-D random fields, such as images, $d = 2$, and $\mathbf{m} := (m_1, m_2)$, etc. With this notation, $C_{ky}(\mathbf{i}_1; ...; \mathbf{i}_{k-1})$, the $k$-th order cumulant of the $d$-D process, $y(\mathbf{n})$, at lags $(\mathbf{i}_1; ...; \mathbf{i}_{k-1})$ is the coefficient of $(v_1 ... v_k)$ in the Taylor expansion of $K(\underline{v}) = \ln E\{\exp(\underline{v}^T \underline{y})\}$ , where $\underline{y} = [y(\mathbf{m}), y(\mathbf{m} + \mathbf{i}_1), ..., y(\mathbf{m} + \mathbf{i}_{k-1})]^T$.

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

$$C_{2y}(\mathbf{i}) = E\{y(\mathbf{m})y(\mathbf{m}+\mathbf{i})\} \,, \tag{1}$$
$$C_{3y}(\mathbf{i};\mathbf{j}) = E\{y(\mathbf{m})y(\mathbf{m}+\mathbf{i})y(\mathbf{m}+\mathbf{j})\} \,, \tag{2}$$

and

$$\begin{aligned} C_{4y}(\mathbf{i};\mathbf{j};\mathbf{k}) &= E\{y(\mathbf{m})y(\mathbf{m}+\mathbf{i})y(\mathbf{m}+\mathbf{j})y(\mathbf{m}+\mathbf{k})\} \\ &- C_{2y}(\mathbf{i})C_{2y}(\mathbf{j}-\mathbf{k}) - C_{2y}(\mathbf{j})C_{2y}(\mathbf{k}-\mathbf{i}) - C_{2y}(\mathbf{k})C_{2y}(\mathbf{i}-\mathbf{j}) \ . \end{aligned} \tag{3}$$

Note that the fourth-order cumulant of a zero-mean random process equals its fourth-order moment less the fourth-order moment of a Gaussian random process with the same autocorrelation. In this report, we will be concerned only with cumulants of real processes; cumulants of complex processes are defined in [39]. The $k$-th order cumulant is a function of $(k-1)$ lag $d$-tuples, and may be expressed in terms of the joint moments of orders $k$ and less.

As a consequence of stationarity, the following symmetry properties hold:

$$\begin{aligned} C_{3y}(\mathbf{i};\mathbf{j}) &= C_{3y}(\mathbf{j};\mathbf{i}) \\ = C_{3y}(-\mathbf{j};\mathbf{i}-\mathbf{j}) &= C_{3y}(\mathbf{i}-\mathbf{j};-\mathbf{j}) \\ = C_{3y}(-\mathbf{i};\mathbf{j}-\mathbf{i}) &= C_{3y}(\mathbf{j}-\mathbf{i};-\mathbf{i}) \ . \end{aligned} \tag{4}$$

Hence, all the third-order cumulants of the stationary process, $y(i,j)$, can be computed from the cumulants in the 'wedge' $\{(\mathbf{m};\mathbf{n}) : 0 \leq m_1 \leq n_1, -\infty < m_2, n_2 < \infty\}$, which is a superset of the *non-redundant* region of support for third-order cumulants. For separable processes, the non-redundant region is defined by the 'wedge', $\{(\mathbf{m},\mathbf{n}) : \mathbf{0} \leq \mathbf{m} \leq \mathbf{n}\}$. More generally, the $k$-th order cumulant is symmetric in its $k-1$ lag $d$-tuples $\mathbf{i}_l, l = 1, ..., k-1$; there are a total of $k!$ symmetry relations of the following two forms:

$$C_{ky}(\mathbf{i}_1;\mathbf{i}_2;...;\mathbf{i}_{k-1}) = C_{ky}(\mathbf{i}_2;\mathbf{i}_1;...;\mathbf{i}_{k-1}) \quad ;$$

$$C_{ky}(\mathbf{i}_1;\mathbf{i}_2;...;\mathbf{i}_{k-1}) = C_{ky}(-\mathbf{i}_1;\mathbf{i}_2-\mathbf{i}_1;...;\mathbf{i}_{k-1}-\mathbf{i}_1) \quad .$$

Extending the results of [37] to random fields, the following important properties can be derived for cumulants[1] of multi-dimensional stochastic signals:

1. The cumulants, of orders greater than two, of a Gaussian process are identically zero.

2. The cumulants, of order greater than two, of a non-Gaussian process cannot all be identically zero. Cumulants, therefore, provide a measure of non-Gaussianity.

3. Cumulants are shift-invariant, i.e., the cumulants of $y(\mathbf{m})$ and $y(\mathbf{m}-\mathbf{n})$, where $\mathbf{n}$ is a non-random tuple, are identical[2].

[1] In general, we use the terms 'cumulants' and 'higher-order cumulants' to indicate $C_{ky}$, $k > 2$.

[2] Bartelt and Wirnitzer [1] used this property to compensate for motion blur in moving photon-limited images. They averaged the bispectra of several short-exposure images and then used a non-parametric reconstruction algorithm to estimate the image. More recently, this property was also exploited by Sadler [38]. Both papers treat the background as (colored) Gaussian noise, and do not consider the problem of object truncation or possible features in the background.

4. Cumulants, of order greater than one, are invariant to additive constants, i.e., the cumulants of $y(\mathbf{m})$ and $y(\mathbf{m})+\alpha$ ($\alpha$ non-random, fixed) are identical; thus, if the given process, $y(\mathbf{m})$, is not zero-mean, its cumulants may be computed as the cumulants of the process, $y(\mathbf{m}) - E\{y(\mathbf{m})\}$.

5. Cumulants of order greater than two, are generally not fully symmetric functions [e.g., $C_{3y}(\mathbf{i};\mathbf{j}) \neq C_{3y}(-\mathbf{i};-\mathbf{j})$], and as such, carry phase information (see also Section 3).

6. If $z(\mathbf{m}) = y(\mathbf{m})+g(\mathbf{m})$, where $y(\mathbf{m})$ is a non-Gaussian process, and $g(\mathbf{m})$ is a (colored) Gaussian process independent of $y(\mathbf{m})$, then the cumulant of $z(\mathbf{m})$ is identical to the cumulant of the signal $y(\mathbf{m})$. As we shall soon see, the cumulants of a *linear* process carry both amplitude and phase information of the linear system. Thus, *in computing the cumulants of the observed noisy process $z(\mathbf{m})$, we are transforming the observed data into a high SNR (signal to noise ratio) domain which preserves all the relevant information about the signal.*

## 2.4 Polyspectra

The *polyspectrum* of order $(k-1)$ of the $d$-D random process, $y(\mathbf{n})$, is defined as the $d(k-1)$-D Fourier transform of the $k$-th order cumulant[3]; absolute summability of the cumulant is a sufficient condition for the existence of the corresponding polyspectrum. Further, if $y(\mathbf{n})$ is a linear process, i.e., $y(\mathbf{n}) = h(\mathbf{n}) * w(\mathbf{n})$, then, the polyspectrum of $y(\mathbf{n})$ exists, if the corresponding polyspectrum of $w(\mathbf{n})$ exists, and $h(\mathbf{n})$ is absolutely summable.

In particular, the $2d$-D *bispectrum* is the $2d$-D Fourier transform of (2), and is given by

$$S_{3y}(\mathbf{u};\mathbf{v}) = \sum_{\mathbf{m},\mathbf{n}} C(\mathbf{m};\mathbf{n}) \exp\left(-j(\mathbf{u}\mathbf{m}^T + \mathbf{v}\mathbf{n}^T)\right) \tag{5}$$

where $j = \sqrt{-1}$.

The $2d$-D bispectrum is periodic in its $2d$ arguments, with period $2\pi$. The $k$-th order polyspectrum of $y(\mathbf{n})$, $S_{ky}$, is a function of $d(k-1)$ frequency variables, $\mathbf{u}_l, l = 1, ..., k-1$, where each $\mathbf{u}_l$ is a $d$-tuple, and is periodic in its $d(k-1)$ frequency variables, with period $2\pi$. In particular, symmetry properties of the bispectrum are given by,

$$\begin{aligned} S_{3y}(\mathbf{u};\mathbf{v}) &= S_{3y}(\mathbf{v};\mathbf{u}) \\ = S_{3y}(\mathbf{u};-\mathbf{u}-\mathbf{v}) &= S_{3y}(-\mathbf{u}-\mathbf{v};\mathbf{u}) \\ = S_{3y}(\mathbf{v};-\mathbf{u}-\mathbf{v}) &= S_{3y}(-\mathbf{u}-\mathbf{v};\mathbf{v}) \end{aligned} \tag{6}$$

Further, for the real-valued process $y(\mathbf{m})$, we have

$$S_{3y}(\mathbf{u};\mathbf{v}) = S^*_{3y}(-\mathbf{u};-\mathbf{v}) \ . \tag{7}$$

[3]In keeping with the notation for 1-D processes, $C_{ky}$ denotes the $k$-th order cumulant, and is a function of $(k-1)$ lag variables; its Fourier transform, denoted by $S_{ky}$, is a function of $(k-1)$ frequency variables, and is called the $k$-th order polyspectrum. Unfortunately, $S_{3y}$ and $S_{4y}$ are usually referred to as *bispectra* and *trispectra* respectively.

More generally, $k$-th order polyspectra are symmetric in their arguments, and satisfy relationships of the form

$$S_{ky}(\mathbf{u}_1;\mathbf{u}_2;...;\mathbf{u}_{k-1}) = S_{ky}(\mathbf{u}_2;\mathbf{u}_1;...;\mathbf{u}_{k-1}) = S_{ky}(\mathbf{u}_1;...;\mathbf{u}_{k-2};-\sum_{l=1}^{k-1}\mathbf{u}_l) \ .$$

The 'non-redundant' region of the bispectrum may be inferred from the symmetry properties given in (6); in this context, we must point out that that the "non-redundant" region of support given in [8] is wrong.

Analogous to 1-D cepstra, *polycepstra* are defined as inverse Fourier transforms of $\ln(S_{ky})$; the $k$-th order polycepstrum exists if the $k$-th order polyspectrum exists and is nowhere zero.

## 2.5 Cumulants versus Moments

We saw earlier that the $k$-th order cumulant is, in general, a complicated non-linear function of the $j$-th $(1 \leq j \leq k)$ order moments. What then is the justification for using cumulants rather than moments? The motivation for using cumulants rather than moments stems from the following facts: (i) cumulants, but not moments, of i.i.d. processes are delta functions; (ii) cumulants, but not moments, of linear processes, may be expressed as higher-order correlations of the p.s.f., see (14) and (15); (iii) the cumulants, but not moments, of (colored) Gaussian processes are identically zero; (iv) when a random variable is approximately normal, its higher-order cumulants, but not its higher-order moments, may be neglected; (v) cross cumulants (but not moments) of independent r.v.'s are zero; and (vi) some technical considerations regarding ergodicity and existence of polyspectra are more easily satisfied by cumulants rather than moments, [4]. However, expressions for higher-order cumulants, see (3), are more involved. Fortunately, third- and fourth-order cumulants suffice for most problems of interest. Additionally, conditional cumulants are much harder to compute than conditional moments. For example, if $A$ is a conditioning event, and $E_A$ denotes expectation with respect to the event A, we have,

$$\begin{aligned} E(XY) &= E_A\{E(XY|A)\} \\ cum(X,Y) &= E_A\{cov(X,Y|A)\} - cov_A\{E(X|A),E(Y|A)\} \ . \end{aligned}$$

For deterministic signals, it turns out that moments and moment spectra essentially play the role that cumulants and cumulant spectra do in the case of stochastic signals [39]. When a deterministic signal is observed in Gaussian noise, we must compute cumulants in order to take care of the additive noise; next, we must estimate the *signal* moments and then perform signal reconstruction. The observed signal, $z(t) = s(t) + g(t)$, is neither ergodic nor stationary; hence, computation of cumulants requires averaging over an ensemble of realizations. Assume that multiple realizations, $z_i(n) = s(n) + g_i(n)$, are available, and that only a finite segment of each realization is available, i.e., $1 \leq n \leq N$. Let $Z_i = col(z_i(n), i =$

$1, ..., N$), and let $S$ and $G_i$ be similarly defined. Then, we have $K$ observations ($i = 1, ..., K$) of the Gaussian random vector, $Z$, with unknown mean $S$. The maximum-likelihood estimate of $S$ is merely the sample mean of the $Z_i$'s (independent of the covariance matrix of $Z$). Hence, in this context, the usefulness of cumulants is rather questionable. However, assume that the signal $s(n)$ suffers an unknown shift from realization to realization; in this case, the shift-invariance property of polyspectra may be used to reconstruct the object, [1, 38].

## 2.6 Sample Estimates of Cumulants and Polyspectra

Given a finite segment of the multi-dimensional process, $y(\mathbf{n})$, $\mathbf{n} \in N_y$, sample estimates of cumulants are obtained by replacing the expectation operator by the sample average operator (the process is assumed to be ergodic); thus,

$$\hat{C}_{3y}(\mathbf{m};\mathbf{n}) = (n(N_y))^{-1} \times \sum_{\mathbf{i}} y(\mathbf{i})y(\mathbf{i}+\mathbf{m})y(\mathbf{i}+\mathbf{n}) \tag{8}$$

where $n(N_y)$ is the number of samples in the segment $N_y$. Usually, the given record is segmented into several, possibly overlapping records; sample estimates of cumulants are obtained for each segment, and then averaged across the segments to obtain the final sample cumulant estimate. The sample covariance of the estimates obtained from the various segments (if the segments are not too short) is a good estimate of the covariance of the overall estimate of the cumulant.

Sample estimates of polyspectra may be obtained by evaluating the Fourier transforms of the cumulant estimates; as in the case of correlations and power spectra, lag windows must be used to ensure consistent *bi-correlogram* estimators; windowing may also help to reduce the estimator variance, [32, Ch. 7]; *optimal* lag windows, as in the case of power spectra, are data dependent.

A modification of the *periodogram* method may also be used. As with second-order spectra (e.g., [32, p. 250]), using (5), it can be shown that the $2d$-D bispectrum may be written as

$$S_{3y}(\mathbf{u};\mathbf{v}) = E\{\, Y(\mathbf{u})Y(\mathbf{v})Y(-\mathbf{u}-\mathbf{v})\} \tag{9}$$

where $Y(\mathbf{u})$ is the $d$-D FFT of the data $y(\mathbf{n})$. A computationally efficient procedure is as follows: (i) the given data is segmented into several, possibly overlapping records, $\{y_k(\mathbf{u})\}_{k=1}^{K}$; (ii) the mean is removed from each record, $y_k(\mathbf{u})$, and its $d$-D FFT, $Y_k(\mathbf{u})$, is computed; (iii) the bispectrum of the $k$-th record is obtained as $B_k(\mathbf{u};\mathbf{v}) = Y_k(\mathbf{u})Y_k(\mathbf{v})Y_k(-\mathbf{u}-\mathbf{v})$; (iv) the bispectra are averaged to yield the final estimate $B(\mathbf{u};\mathbf{v}) = K^{-1}\sum_{k=1}^{K} B_k(\mathbf{u};\mathbf{v})$. Equation (9) is thus evaluated as a sample-average over the records. As before, a smoothing window is essential. This computational approach was suggested for the 1-D case in [21].

Generalizations of the bicorrelogram and biperiodogram methods to higher-order spectra are straightforward.

In Section 5, where parametric methods are discussed, we will see that low-variance estimates of the polyspectra of linear processes may be obtained by first estimating the transfer function of the model, followed by computation of the polyspectra.

Let $y(\mathbf{n})$, $\mathbf{n} = (n_1, n_2)$, represent a 2-D process. Assume that $N^2$ samples of $y(n_1, n_2)$, $0 \leq n_i \leq N-1$, $i = 1, 2$ are available. Let $K$ denote the largest desired lag of the sample cumulants ($K \leq N-1$) along the row direction. Assume that each row is extended by adding $K$ zeros, and let $Y$ denote the lexicographic (row-ordered) representation of the image, i.e.,

$$Y(m(N+K)+n) = \begin{cases} y(m,n), & m = 0, ..., N-1; n = 0, ..., N-1 \\ 0, & \text{else} \end{cases}$$

The cumulants of the 2-D process, $y(\mathbf{n})$, may be obtained from those of the 1-D process, $Y(n)$, as follows. Assume that we want to compute the third-order cumulant $C_{3y}(\mathbf{i};\mathbf{j})$, where $|i_2| \leq K$ and $|j_2| \leq K$. Let $a = i_1(N+K) + i_2$ and $b = j_1(N+K) + j_2$. With $N' = N + K$, we have

$$\begin{aligned} \hat{C}_{3Y}(a,b) &:= \frac{1}{NN'} \sum_i Y(i) Y(i+a) Y(i+b) \\ &= \frac{1}{NN'} \sum_{m=0}^{N-1} \sum_{n=0}^{N+K-1} Y(mN'+n) Y(mN'+n+a) Y(mN'+n+b) \\ &= \frac{1}{NN'} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} Y(mN'+n) Y((m+i_1)N'+n+i_2) Y((m+j_1)N'+n+j_2) \\ &= \frac{1}{NN'} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} y(m,n) y(m+i_1, n+i_2) y(m+j_1, n+j_2) \\ &= \frac{1}{NN'} \sum_{\mathbf{n}} y(\mathbf{n}) y(\mathbf{n}+\mathbf{i}) y(\mathbf{n}+\mathbf{j}) \\ &= \frac{N}{N+K} \hat{C}_{3y}(\mathbf{i};\mathbf{j}) \end{aligned}$$

Hence,

$$C_{3y}(\mathbf{i};\mathbf{j}) = C_{3Y}(i_1(N+K)+i_2,\ j_1(N+K)+j_2) \ .$$

The non-zero cumulants in the non-redundant region of $C_{3Y}$, with $K = N-1$, correspond to the non-redundant region of $C_{3y}$. Column-ordering is useful if the maximum cumulant lag along the column direction is smaller than that along the row direction. The extension to higher-order cumulants is obvious; for example,

$$C_{4y}(\mathbf{i};\mathbf{j};\mathbf{k}) = C_{4Y}(i_1(N+K)+i_2,\ j_1(N+K)+j_2,\ k_1(N+K)+k_2) \ .$$

# 3 Cumulants of Linear Processes

In this section, we explore some general properties of the cumulants of linear processes; in particular, polyspectra may be expressed in terms of the p.s.f. transfer function; polycepstral relations are also discussed.

## 3.1 Expressions in terms of p.s.f

Let $z(\mathbf{n})$ denote the noisy output of a linear shift-invariant $d$-D model, $h(\mathbf{i})$, which is driven by a zero-mean unobservable, non-Gaussian random input, $w(\mathbf{n})$; i.e.,

$$y(\mathbf{n}) = \sum_{\mathbf{i}} h(\mathbf{i})w(\mathbf{n}-\mathbf{i}) \tag{10}$$

$$z(\mathbf{n}) = y(\mathbf{n}) + g(\mathbf{n}) \tag{11}$$

where $g(\mathbf{n})$ is colored Gaussian noise of unknown covariance, and is assumed to be independent of $w(\mathbf{n})$. For example, $h(\mathbf{i})$ in (10) may represent the p.s.f. due to the blurring mechanism of motion, atmospheric turbulence, defocussing, etc. Restoration of the degraded image demands good estimates of the p.s.f. $h(\mathbf{i})$. In reflection seismology, $h(\mathbf{i})$ represents the combined effects of the source waveform (the exciting mechanism) as well as propagation effects (internal reflections or multipath effects).

Let the transfer function of the noiseless model in (10) be denoted by $H(\mathbf{u}) = M(\mathbf{u})\exp(j\phi(\mathbf{u}))$. Since we propose the use of higher-order cumulants, we will need a corresponding whiteness assumption, i.e.,

$$\begin{aligned} C_{kw}(\mathbf{i}_1;...;\mathbf{i}_{k-1}) &= \gamma_{kw}, \text{ if } \mathbf{i}_i = ... = \mathbf{i}_{k-1} = \mathbf{0}, \\ &= 0, \text{ else.} \end{aligned} \tag{12}$$

Note that an i.i.d. process is $k$-th order white for all $k$; further, if $w(\mathbf{n})$ is non-Gaussian, there exists at least one non-zero cumulant. Notice that a weaker k-th order whiteness condition would be to assume that the input cumulant, $C_{kw}$, is an appropriate sinc function, or equivalently that the input $k$-th order polyspectrum is flat in the frequency range where the p.s.f. is non-zero.

The correlation function of the linear process, $y(\mathbf{n})$, may be expressed in terms of the p.s.f. as $(\gamma_{2w} := \sigma_w^2)$

$$C_{2y}(\mathbf{m}) = \gamma_{2w} \sum_{\mathbf{i}} h(\mathbf{i})h(\mathbf{i}+\mathbf{m}) \tag{13}$$

Using the definition of cumulants, (10) and the whiteness assumption in (12), a general expression relating the $k$-th order cumulant with the p.s.f. is obtained, as

$$C_{ky}(\mathbf{m}_1;\mathbf{m}_2;....;\mathbf{m}_{k-1}) = \gamma_{kw} \sum_{\mathbf{i}} \left[ \prod_{j=0}^{k-1} h(\mathbf{i}+\mathbf{m}_j) \right] \tag{14}$$

where $\mathbf{m}_0 = \mathbf{0}$. Thus, the $k$-th order cumulant of the linear process, $y(\mathbf{n})$ is proportional to the $k$-th order *deterministic* correlation of the p.s.f. of the linear system. From (14), we note that the $k$-th order polyspectrum is related to the p.s.f. transfer function via

$$S_{ky}(\mathbf{u}_1; ...; \mathbf{u}_{k-1}) \quad = \quad \gamma_{kw} \prod_{i=1}^{k} H(\mathbf{u}_i) \tag{15}$$

where $\mathbf{u}_k = -\sum_{i=1}^{k-1} \mathbf{u}_i$.

Given the spectrum, $S_{2y}(\mathbf{u})$, of a linear process, the p.s.f., $h(\mathbf{i})$ or $H(\mathbf{u})$, may be obtained under the minimum-phase assumption, provided the corresponding 2-D polynomial can be factored. The factorizability assumption may be relaxed by imposing other constraints on the model [e.g., assume that the model is symmetric AR, etc.]. In Sections 4 and 5, we will see that $H(\mathbf{u})$ can be recovered (to within a constant scale factor and a linear phase ambiguity) from the $k$-th order polyspectrum, $S_{ky}$, $k > 2$, without the assumptions of symmetry or factorizability!

## 3.2 Relation to Lower-Order Spectra

For linear processes, the $l$-th order polyspectrum may be estimated from the $k$-th ($k > l$) order polyspectrum in one of two ways: a) we may use one of the methods in Sections 4 and 5 to estimate the p.s.f. transfer function, $h(\mathbf{n})$ or $H(\mathbf{u})$, and then use (14) or (15) to estimate the $l$-th order cumulant and polyspectrum; or, b) we recognize the fact that specific slices of the $k$-th order polyspectrum are proportional to the $l$-th order polyspectrum (equivalently, projections in the lag domain); we discuss this below.

The $2d$-D bispectrum of the linear process in (10) is obtained from (15), with $k = 3$, as

$$\begin{aligned} S_{3y}(\mathbf{u}; \mathbf{v}) &= |S_{3y}(\mathbf{u}; \mathbf{v})| \exp\{j\psi(\mathbf{u}; \mathbf{v})\} \\ &= \gamma_{3w} H(\mathbf{u}) H(\mathbf{v}) H(-\mathbf{u} - \mathbf{v}) \; . \end{aligned} \tag{16}$$

Setting $\mathbf{v} = \mathbf{0}$ in (16) yields

$$S_{3y}(\mathbf{u}; \mathbf{0}) = \gamma_{3w} H(\mathbf{u}) H(-\mathbf{u}) H(\mathbf{0}) = \alpha S_{2y}(\mathbf{u}) \tag{17}$$

where $\alpha = \gamma_{3w} H(\mathbf{0}) / \gamma_{2w}$. Thus, the 2-D power spectral density of the p.s.f., and hence the p.s.f. amplitude, may be obtained (to within a scalar multiple) from a 1-D slice of the bispectrum. Equivalently, the correlation function may be obtained from the third-order cumulant as,

$$C_{2y}(\mathbf{m}) = \alpha^{-1} \sum_{\mathbf{i}} C_{3y}(\mathbf{m}; \mathbf{i}) \tag{18}$$

Equations (17) and (18) may be used to estimate the amplitude spectrum from the bispectrum provided $H(\mathbf{0}) \neq 0$ and $\gamma_{3w} \neq 0$. If these two conditions do not hold, then the amplitude spectrum may be estimated from one of the higher order polyspectra. The $m$-th order polyspectrum may be expressed in terms of the $k > m$-th order polyspectrum, as

$$S_{my}(\mathbf{u}_1; ...; \mathbf{u}_{m-1}) = \frac{\gamma_{mw}}{\gamma_{kw}} \frac{S_{ky}(\mathbf{u}_1; ...; \mathbf{u}_{m-1}; \mathbf{u}_m; ...; \mathbf{u}_{k-1})}{\prod_{i=m}^{k-1} H(\mathbf{u}_i)}, \quad \sum_{i=m}^{k-1} \mathbf{u}_i = \mathbf{0} \tag{19}$$

If $H(\mathbf{0}) \neq 0$, a convenient choice is $\mathbf{u}_l = \mathbf{0}$, $l = m, ..., k-1$, in which case, we obtain

$$C_{my}(\mathbf{i}_1; ...; \mathbf{i}_{m-1}) = \beta \sum_{\mathbf{i}_m} \cdots \sum_{\mathbf{i}_{k-1}} C_{ky}(\mathbf{i}_1; ...; \mathbf{i}_{m-1}; \mathbf{i}_m; ...; \mathbf{i}_{k-1}) . \tag{20}$$

where $\beta = \gamma_{mw} / \gamma_{kw} H^{k-m}(\mathbf{0})$.

If $H(0) = 0$, the amplitude spectrum may be estimated from the fourth-order spectrum via,

$$S_{2y}(\mathbf{u}) = \frac{\gamma_{2w}}{\gamma_{4w}} \frac{S_{4y}(\mathbf{u}, \mathbf{u}_0, -\mathbf{u}_0)}{S_{2y}(\mathbf{u}_0)} \tag{21}$$

for some $\mathbf{u}_0$ such that $H(\mathbf{u}_0) \neq 0$. This method of estimating amplitude spectra from higher-order spectra is particularly useful when the observations are contaminated with colored Gaussian noise. In the lag-domain this corresponds to exponentially weighted summations; i.e.,

$$C_{2y}(\mathbf{m}) \propto \sum_{\mathbf{n}_1} \sum_{\mathbf{n}_2} C_{4y}(\mathbf{m}, \mathbf{n}_1, \mathbf{n}_2) \exp(j \mathbf{u}_0 (\mathbf{n}_1 - \mathbf{n}_2)) \tag{22}$$

Although (22) holds exactly in theory, when sample estimates of $C_{4y}$ are used in practice, appropriate windowing should be used to reduce the variance of the estimated $C_{2y}$. Expressions for the asymptotic variance of the estimates obtained via (22) with weighting may be found in [12] for the 1-D case.

Similarly, the third-order cumulant may be obtained from the fifth-order cumulant via,

$$S_{3y}(\mathbf{u}_1, \mathbf{u}_2) = \frac{\gamma_{3w}}{\gamma_{5w}} \frac{S_{5y}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_0, -\mathbf{u}_0)}{S_{2y}(\mathbf{u}_0)} \tag{23}$$

$$C_{3y}(\mathbf{m}, \mathbf{n}) \propto \sum_{\mathbf{k}_1} \sum_{\mathbf{k}_2} C_{5y}(\mathbf{m}, \mathbf{n}, \mathbf{k}_1, \mathbf{k}_2) \exp(j \mathbf{u}_0 (\mathbf{k}_1 - \mathbf{k}_2)) \tag{24}$$

for some $\mathbf{u}_0$, such that $H(\mathbf{u}_0) \neq 0$.

## 3.3 Polycepstra

The frequency-domain (FD) *polycepstrum* is defined as the complex logarithm of the polyspectrum, and is given by

$$\mathcal{S}_{ky}(\mathbf{u}_1; ...; \mathbf{u}_{k-1}) := \ln\left(S_{ky}(\mathbf{u}_1; ...; \mathbf{u}_{k-1})\right) = \ln(\gamma_{kw}) + \sum_{i=1}^{k} \mathcal{H}(\mathbf{u}_i) \tag{25}$$

where $\mathcal{H}(\mathbf{u}) = \ln(H(\mathbf{u}))$ is the conventional FD cepstrum. The conventional FD cepstrum exists, if the p.s.f., $h(\mathbf{i})$, is exponentially stable and its amplitude spectrum has no zeroes on the unit d-circle; this, and the assumption, $0 < |\gamma_{kw}| < \infty$, guarantee the existence of the FD polycepstrum.

The inverse Fourier transform of the FD polycepstrum is called the time-domain (TD) polycepstrum, and will be denoted by $\mathcal{C}_{ky}(\mathbf{m}_1; ...; \mathbf{m}_{k-1})$. The TD polycepstrum is non-zero only over the 'axes': $\mathcal{R}_i : \{\mathbf{m}_l = \mathbf{0}, l = 1, ..., k-1, l \neq i\}$, $i = 1, ..., k-1$, and the 'diagonal', $\mathcal{R}_d : \{\mathbf{m}_i = ... = \mathbf{m}_{k-1}\}$. Further,

$$\begin{aligned} \mathcal{C}_{ky}(\mathbf{m}; \mathbf{0}; ...; \mathbf{0}) &= \mathcal{C}_{ky}(\mathbf{0}; \mathbf{m}; \mathbf{0}; ...; \mathbf{0}) \\ = \mathcal{C}_{ky}(\mathbf{0}; \mathbf{0}; ...; \mathbf{m}) &= \mathcal{C}_{ky}(-\mathbf{m}; -\mathbf{m}; ...; -\mathbf{m}) \\ &= \hat{h}(\mathbf{m}) + \ln(\gamma_{kw})\delta(\mathbf{m}) \end{aligned} \tag{26}$$

where $\hat{h}(\mathbf{m})$ is the $d$-D complex TD cepstrum of $h(\mathbf{m})$, and $\delta(\mathbf{m}) = 1$, if $\mathbf{m} = \mathbf{0}$, and is zero otherwise. In [44, 45], it is shown that the relationships in (26) are necessary and sufficient for a non-Gaussian process to be linear; based on this, a measure of linearity is proposed in [44][4].

Further, analogous to the convolution relation between a 1-D signal, and its complex TD cepstrum, [30, Ch. 10.4.2], and the 2-D counterpart in [9, p. 207], it is easily shown that the $k$-th order cumulant and the corresponding TD polycepstrum are related via the convolutional relationship,

$$\left( (\prod_{l=1}^{d} n_l) \mathcal{C}_{ky}(\mathbf{n}; \mathbf{m}_2; ...; \mathbf{m}_{k-1}) \right) * C_{ky}(\mathbf{n}; \mathbf{m}_2; ...; \mathbf{m}_{k-1}) = \left( (\prod_{l=1}^{d} n_l) C_{ky}(\mathbf{n}; \mathbf{m}_2; ...; \mathbf{m}_{k-1}) \right) \tag{27}$$

where $*$ denotes linear convolution in *all* the indices $\mathbf{n}, \mathbf{m}_l$, $l = 2, ..., k-1$. The bicepstral version of (27) for 1-D signals was given in [27].

For $k = 3$ and $d = 2$, the convolutional relation may be more clearly written as

$$\sum_{l_1} \sum_{l_2} l_1 l_2 \hat{h}(\mathbf{l}) [C_{3y}(\mathbf{m} - \mathbf{l}, \mathbf{n}) - C_{3y}(\mathbf{m} + \mathbf{l}, \mathbf{n} + \mathbf{l})] = m_1 m_2 C_{3y}(\mathbf{m}, \mathbf{n}) \tag{28}$$

For notational convenience, some of the algorithms in this report will be developed using 3rd-order cumulants of 2-D processes assuming that $0 < |\gamma_{3w}| < \infty$, although generalizations to higher-order cumulants and higher dimensional processes are straight-forward, and are indicated. Algorithms based on fourth-order cumulants are indicated, if the driving noise is non-Gaussian symmetric (e.g., double exponential) with $\gamma_{3w} = 0$, but $\gamma_{4w} \neq 0$.

[4]The proposed measure of linearity is the ratio of the energy of the polycepstrum along the axes and main diagonal to the energy in the entire TD polycepstrum. High values of linearity were obtained for several real images.

# 4 Non-Parametric Models

In this section, we develop three non-parametric methods for estimating the p.s.f. phase from the polyspectral phase; a polycepstral method is also briefly discussed.

Given only the noisy observations, we are interested in estimating $h(\mathbf{i})$, the impulse response of the system. We concentrate largely on estimating $\phi(\mathbf{u})$, the phase of the transfer function; however, we will also discuss methods of estimating the amplitude spectrum from slices of polyspectra; this is particularly useful when the covariance function of the additive colored Gaussian noise is unknown.

## 4.1 Single-Slice Recursive Phase Reconstruction

Equating the phases on both sides of (16), we find that the phase of the 2-D bispectrum is related to the p.s.f. phase via

$$\psi(\mathbf{u};\mathbf{v}) = \phi(\mathbf{u}) + \phi(\mathbf{v}) - \phi(\mathbf{u}+\mathbf{v}) \ . \tag{29}$$

The frequency variables in (29) are all continuous; following the steps used in the 1-D phase reconstruction algorithm of [22], we discretize the frequency range $[0, 2\pi]$. The discretized version of (29) is

$$\psi(\mathbf{m};\mathbf{n}) = \phi(\mathbf{m}) + \phi(\mathbf{n}) - \phi(\mathbf{m}+\mathbf{n}) \ ; \tag{30}$$

hence, $\psi(\mathbf{u};\mathbf{0}) = \psi(\mathbf{0};\mathbf{v}) = \phi(\mathbf{0})$. Let $\mathbf{m} = l\mathbf{n}_i + \mathbf{m}_j$, and $\mathbf{n} = \mathbf{n}_i$ in (30); we obtain

$$\psi(l\mathbf{n}_i + \mathbf{m}_j;\mathbf{n}_i) = \phi(l\mathbf{n}_i + \mathbf{m}_j) + \phi(\mathbf{n}_i) - \phi((l+1)\mathbf{n}_i + \mathbf{m}_j) \ . \tag{31}$$

Summing (31) from $l = 0$ to $l = r-1$, yields

$$\phi(r\mathbf{n}_i + \mathbf{m}_j) = \phi(\mathbf{m}_j) + r\phi(\mathbf{n}_i) - \sum_{l=0}^{r-1} \psi(l\mathbf{n}_i + \mathbf{m}_j;\mathbf{n}_i) \tag{32}$$

Now, $\phi(\mathbf{0})$ is directly obtained as $\psi(\mathbf{0};\mathbf{0}) = \phi(\mathbf{0})$; for a real valued $y(\mathbf{n})$, this must be 0 or $\pi$. Let $\mathbf{e}_l$ denote the $d$-tuple, with a 1 as its $l$-th element, and zeros elsewhere. Hence, with $\mathbf{m}_0 = \mathbf{0}$, and $\mathbf{n}_i = \mathbf{e}_i$, $i = 1, ..., d$, the phase values along the axes, $\phi(r\mathbf{e}_l)$, can be recursively computed, provided the $d$ initial values, $\phi(\mathbf{e}_l)$, $l = 1, ..., d$, are known. Then, with $\mathbf{n}_i = \mathbf{e}_i$, and $\mathbf{m}_j = s\mathbf{e}_j$, $j \neq i$, the phase values on the planes passing through the origin, $\phi(r\mathbf{e}_i + s\mathbf{e}_j)$, can be computed; this procedure can now be repeated with $\mathbf{n}_i = \mathbf{e}_i$, and $\mathbf{m}_j = r\mathbf{e}_{j_1} + s\mathbf{e}_{j_2}$, and so on, until all the phase values $\phi(\mathbf{m})$ are obtained.

The recursive procedure can be expressed in closed form as (recall, $\mathbf{m} = (m_1, ..., m_d)$),

$$\phi(\mathbf{m}) = \phi(\mathbf{0}) + \sum_{k=1}^{d} m_k \phi(\mathbf{e}_k) - \sum_{l=1}^{d} \left( \sum_{k=0}^{m_l - 1} \psi(k\mathbf{e}_l + \sum_{i=l+1}^{d} m_i \mathbf{e}_i; \mathbf{e}_l) \right) \tag{33}$$

For real signals, we must have $\phi(\frac{N}{2}\mathbf{e}_k) = n_k\pi$; choosing $\phi(\frac{N}{2}\mathbf{e}_k) = 0$ amounts to a possible shift of the p.s.f. by $n_k$ samples, parallel to one of the axes; such a linear shift ambiguity is unavoidable, since we are using only output data, [see also property (3) of Section 2.3]. Thus, the inital values are given by

$$\phi(\mathbf{e}_k) = \frac{2}{N}\left(\sum_{i=0}^{N/2-1} \psi(i\mathbf{e}_k;\mathbf{e}_k) - \phi(\mathbf{0})\right) \tag{34}$$

Thus, a closed-form expression for the system phase, modulo a linear-phase ambiguity, in terms of the bispectral phase, is given by (33), together with (34), and $\phi(\mathbf{0}) = \psi(\mathbf{0};\mathbf{0})$. The closed-form expression in (33) is readily extended to higher-order polyspectra.

It should be stressed that all these phase relations are modulo $2\pi$. An alternative form, which automatically incorporates the modulo $2\pi$ condition, is given by

$$\exp\{j\phi(\mathbf{m})\} = \exp\{j(\phi(\mathbf{0}) + \left(\sum_{k=1}^{d} m_k\phi(\mathbf{e}_k) - \sum_{k=0}^{m_l-1}\psi(k\mathbf{e}_l + \sum_{i=l+1}^{d} m_i\mathbf{e}_i;\mathbf{e}_l)\right))\} \tag{35}$$

This is an m-D extension of the 1-D algorithm in [22]. Since $y(\mathbf{m})$ is real, its Fourier transform is conjugate-symmetric. Assume that $d = 2$, and that the Fourier transform size is $N \times N$; then $\phi(m,n) = -\phi(N-m, N-n)$, $0 < m,n \leq N-1$; hence, in (33), only the $N^2/2+2$ values of $\phi(m,n)$ in the region, $\{m = 0, 0 \leq n \leq N/2\}$, $\{m = N/2, 0 \leq n \leq N/2\}$, and $\{0 < m < N/2, 0 \leq n \leq N-1\}$, need to be computed.

## 4.2 Closed-Form Phase Reconstruction

The phase relationship given in (30) may be re-written as,

$$\phi(\mathbf{m}) = \psi(\mathbf{m},\mathbf{n}) - \phi(\mathbf{n}) + \phi(\mathbf{m},\mathbf{n})$$

which leads to

$$\begin{aligned}\phi(\mathbf{m}) &= \frac{1}{N'}\sum_{\mathbf{n}}[\psi(\mathbf{m},\mathbf{n}) - \phi(\mathbf{n}) + \phi(\mathbf{m},\mathbf{n})] \\ &= \frac{1}{N'}\sum_{\mathbf{n}}\psi(\mathbf{m},\mathbf{n}) + \alpha \end{aligned} \tag{36}$$

where $N'$ is the number of distinct $\mathbf{n}$'s, and $\alpha$ is an unknown constant. Note that the closed-form solution uses all of the bispectral phase information. The obvious generalization of (36) to higher-order polyspectra holds.

## 4.3 Least-Squares Phase Reconstruction

Equation (33) suffers from the obvious drawback that it uses the bispectral phase values only along particular slices, i.e., $\psi(\mathbf{m},\mathbf{e}_l)$, $l = 1,...,d$. Improved estimates may be obtained by using all the bispectral phase values; we demonstrate the procedure for $d = 2$. In order to reduce the variance of the estimate, we may, as in the 1-D case, [2], use alternative 2-D slices to estimate $\phi(m,n)$, and then average the various estimates. Let $N = 2M$, where $(M,M)$ corresponds to the Nyquist frequency on our discretized grid, and define the $2M^2+2$ element vector $\underline{\alpha}$ by

$$\underline{\alpha} = [\mathbf{a}_0^T, ...., \mathbf{a}_M^T]^T$$

where

$$\mathbf{a}_i = [\phi(i,0), ...., \phi(i,N-1)]^T, \quad i = 1, ..., M-1 \ ,$$

and

$$\mathbf{a}_0 = [\phi(0,0), ...., \phi(0,M)]^T \ , \ \mathbf{a}_M = [\phi(M,0), ...., \phi(M,M)]^T \ ,$$

which includes all the non-redundant phase values; the phase values at the zero and Nyquist frequency samples are directly obtained from the bispectrum, as,

$$\phi(0,0) = \psi[(0,0);(0,0)]; \quad \phi(N/2,N/2) = 0.5(\psi[(N/2,N/2);(N/2,N/2)] - \phi(0,0)) \ .$$

Next, we concatenate (30) for those values of $(\mathbf{m};\mathbf{n})$ which are in the non-redundant region of the bispectrum, to obtain the linear system of equations

$$A\underline{\alpha} = \underline{\psi} \ . \tag{37}$$

where $\underline{\psi}$ is the vector of bispectral phase values resulting from (30). $A$ is a sparse matrix with at most three non-zero entries in each row (-1,1 or 2). Thus the least-squares solution to (37) can be obtained very efficiently. This algorithm uses all the values in the *non-redundant* region of the bispectrum. It may be shown that matrix $A$ has rank $2M^2$; since $\phi(0,0)$ and $\phi(M,M)$ are known, a unique (modulo linear terms) phase reconstruction is obtained from the solution of (37).

## 4.4 Recursive Amplitude and Phase Reconstruction

An alternative *recursive* algorithm for the *combined* estimation of the phase and amplitude of $H(\mathbf{u})$ may be obtained as follows. Consider the discretized version of (16): we obtain (all indices are modulo N),

$$S_{3y}(\mathbf{m};\mathbf{n}) = \gamma_{3w} H(\mathbf{m}) H(\mathbf{n}) H^*(\mathbf{m}+\mathbf{n}) \tag{38}$$

where we have assumed that the p.s.f., $h(\mathbf{i})$, is real-valued. From (38), we obtain,

$$S_{3y}(\mathbf{m};0) = \gamma_{3w} H(\mathbf{m}) H(\mathbf{0}) H^*(\mathbf{m}) \tag{39}$$

$$S_{3y}(\mathbf{m}-\mathbf{n};\mathbf{n}) = \gamma_{3w} H(\mathbf{m}-\mathbf{n}) H(\mathbf{n}) H^*(\mathbf{m}) \tag{40}$$

$$H(\mathbf{m}) = \frac{S_{3y}(\mathbf{m};\mathbf{0})}{S_{3y}(\mathbf{m}-\mathbf{n};\mathbf{n})} \, \frac{H(\mathbf{m}-\mathbf{n})}{H(\mathbf{0})} H(\mathbf{n}) \tag{41}$$

which may be used in a recursive procedure for estimating $H(\mathbf{m})$. Let $R(\mathbf{m})$ denote the region $[0,m_1] \times [0,m_2]$, but excluding the points $(m_1,m_2)$ and $(0,0)$. Then, an improved estimate of $H(\mathbf{m})$ may be obtained by averaging the estimates obtained from (41) with $\mathbf{n} \in R(\mathbf{m})$ and $\mathbf{m} - \mathbf{n} \in R(\mathbf{n})$, i.e.,

$$H(\mathbf{m}) = [(m_1+1)(m_2+1)-2]^{-1} H^{-1}(\mathbf{0}) \sum_{\mathbf{n}\in R(\mathbf{n})} \frac{S_{3y}(\mathbf{m};\mathbf{0})}{S_{3y}(\mathbf{m}-\mathbf{n};\mathbf{n})} H(\mathbf{m}-\mathbf{n}) H(\mathbf{n}) \qquad (42)$$

At the expense of some additional notational complexity, the summation may be reduced to cover only those terms that are in the non-redundant region of the bispectrum. The initial value is given by, $H(\mathbf{0}) = B^{1/3}(\mathbf{0};\mathbf{0})$, The second set of initial condition for the recursion are obtained as

$$H(\mathbf{e}_i) = |H(\mathbf{e}_i)| \exp(j\phi(\mathbf{e}_i)) = \sqrt{S_{3y}(\mathbf{e}_i;\mathbf{0})/H(\mathbf{0})} \exp(j\phi(\mathbf{e}_i)) \ , i = 1, 2 \ .$$

Estimates of the phase terms $\phi(\mathbf{e}_i)$ are obtained according to the procedure discussed in Section 4.1; the reconstructed phase is unique, modulo a linear shift. The 1-D counterpart of this was reported in [11]. By expressing (41) in polar co-ordinates, recursive expressions for the phase or the amplitude alone may be obtained. If $H(\mathbf{0})$ is zero, eq. (41) cannot be used for the recursion. In this case, (38) may be used with $\mathbf{m} = (0,1), (1,0), (1,1)$ in the $d = 2$ case; see also [2, 11, 38].

The recursive, closed-form and least-squares phase-reconstruction methods developed in the first three subsections made use of the relationship,

$$\phi(\mathbf{u};\mathbf{v}) = \phi(\mathbf{u}) + \phi(\mathbf{v}) - \phi(\mathbf{u}+\mathbf{v})$$

The log-magnitude, $M(\mathbf{u}) := \ln H(\mathbf{u})$, assuming it exists, is related to the log bispectral magnitude function via,

$$\ln |B(\mathbf{u},\mathbf{v})| = M(\mathbf{u}) + M(\mathbf{v}) + M(\mathbf{u}+\mathbf{v})$$

Hence, the three phase reconstruction methods may be easily modified to yield log amplitude estimation algorithms.

## 4.5 Polycepstral Methods

Finally, we will consider a non-parametric method which yields indirect estimates of the p.s.f. coefficients via the TD polycepstrum. This method is potentially useful since it is not recursive and, therefore, not prone to error propagation effects; however, it assumes that the amplitude spectrum has no zeros on the unit d-circle. For simplicity consider 2-D processes ($d = 2$) and bispectra, $k = 3$. The polycepstral relations in (26) and (27) simplify to

$$\sum_{\mathbf{l}} l_1 l_2 \hat{h}(\mathbf{l}) [C_{3y}(\mathbf{n}-\mathbf{l},\mathbf{m}) + C_{3y}(\mathbf{n}+\mathbf{l},\mathbf{m}+\mathbf{l})] = n_1 n_2 C_{3y}(\mathbf{n},\mathbf{m}) \qquad (43)$$

For exponentially stable p.s.f.'s, the cepstral coefficients may be assumed to be effectively non-zero only over a finite region of support. Under this truncation assumption, eq (43) leads to a system of linear equations in a finite number of unknowns, which can be solved for the $\hat{h}(\mathbf{l})$'s. The p.s.f. coefficients, $h(\mathbf{n})$'s are easily derived from the $\hat{h}(\mathbf{l})$'s using the procedure in [9]. As in the 1-D case, [27], this method does not require the explicit computation of the polycepstrum.

The third-order cumulant of a 2-D image is a 4-D function; in order to reduce the computational burden, we can also reconstruct the image from its 1-D projections. The 2-D image may be reconstructed via the Radon transform from the reconstructed projections; in this case, 1-D reconstruction methods are directly applicable [28].

# 5 Parametric Models

Two-dimensional ARMA models have been extensively used in absorption spectroscopy, astronomical data processing and in various image processing tasks, but parameter estimation techniques have been largely restricted to the case of Gaussian excitation [17, 18], with the consequent restrictions such as symmetric non-causal models. In our model, we assume that the input excitation $w(\mathbf{m})$ is non-Gaussian and $k$-th order white; however, we do not place any restrictions on the p.s.f.

Our 2-D ARMA model is given by

$$\sum_{\mathbf{k}\in\mathcal{N}_a} a(\mathbf{k})\, y(\mathbf{n}-\mathbf{k}) = \sum_{\mathbf{k}\in\mathcal{N}_b} b(\mathbf{k})\, w(\mathbf{n}-\mathbf{k}) \tag{44}$$

where $\mathcal{N}_a$ and $\mathcal{N}_b$ denote the regions of support ('neighborhoods') of the AR and MA parameters; we assume rectangular neighborhoods, i.e., for $d=2$, $\mathcal{N}_a = [-M_1^a, M_2^a] \times [-N_1^a, N_2^a]$, and $\mathcal{N}_b = [-M_1^b, M_2^b] \times [-N_1^b, N_2^b]$. Since ARMA modeling based on output data is insensitive to time shifts, without loss of generality, we may appropriately shift indices to obtain, in the $d=2$ case, for example, support regions defined over $\mathcal{N}_a = [0, p1] \times [0, p2]$, and $\mathcal{N}_b = [0, q1] \times [0, q2]$[5]. The observed noisy process is given by

$$z(\mathbf{n}) = y(\mathbf{n}) + g(\mathbf{n}) \tag{45}$$

where $z(\mathbf{n})$ is the observed series, and $g(\mathbf{n})$ is colored Gaussian noise, which is independent of the signal $y(\mathbf{n})$. The system is assumed to be exponentially stable. Since the $k$-th ($k>2$) cumulants of a Gaussian process vanish, $C_{kz}(\mathbf{m}_1, ..., \mathbf{m}_{k-1}) = C_{ky}(\mathbf{m}_1, ..., \mathbf{m}_{k-1})$, $k>2$. Further, if bispectra are used, then the output noise, $g(\mathbf{n})$, may be colored noise of unknown distribution (not necessarily Gaussian) and spectral density (not necessarily i.i.d.) so long as its third-order cumulant is zero (e.g., $g(\mathbf{n})$ has a symmetric p.d.f.).

The driving noise for the AR part of our model is a colored MA process, as in [46], where essentially the noiseless case is considered; since only correlation information is used, the four-quadrant decomposition algorithm in [46] needs additional phase assumptions (such as zero-phase) to guarantee identifiability; furthermore, since symmetric p.s.f.'s are assumed in [46], our model is more general. Symmetric AR models driven by i.i.d. noise are also considered in [6, 19], where iterative methods for approximate maximum-likelihood (ML) solutions are derived; the ARMA parameter estimation algorithms in this section do not assume that the p.d.f. of the input noise is known; hence, compared to ML methods, they will be statistically inefficient; when the input p.d.f. is known, the techniques in this Section may be used to obtain good initial estimates for an iterative ML algorithm.

[5]Shifting of the indices does not convert a stable non-causal model to a stable causal model; index shifting is done only for notational convenience

## 5.1 Causal ARMA Models: AR Estimation

If the model represented by (44) is strictly (quarter-plane) causal, then, multiplying (44) by $y(\mathbf{n}-\mathbf{m})$ and taking expectations yields the following recursion for the autocorrelation,

$$\sum_{\mathbf{l}\in\mathcal{N}_a} a(\mathbf{l})C_{2y}(\mathbf{l}-\mathbf{m}) = 0, \;\; \mathbf{m} \notin \mathcal{N}_b \; . \tag{46}$$

Similarly, multiplying (44) by $y(\mathbf{n}-\mathbf{i})y(\mathbf{n}-\mathbf{j})$ and taking expectations yields the following recursion for third-order cumulants

$$\sum_{\mathbf{l}\in\mathcal{N}_a} a(\mathbf{l})C_{3y}(\mathbf{l}-\mathbf{i};\mathbf{l}-\mathbf{j}) = 0, \;\; \mathbf{i} \notin \mathcal{N}_b \text{ or } \mathbf{j} \notin \mathcal{N}_b. \tag{47}$$

The recursion for $k$-th order cumulants are similarly obtained as

$$\sum_{\mathbf{l}\in\mathcal{N}_a} a(\mathbf{l})C_{ky}(\mathbf{l}-\mathbf{i}_1;\mathbf{l}-\mathbf{i}_2;...;\mathbf{l}-\mathbf{i}_{k-1}) = 0, \;\; \mathbf{i}_j \notin \mathcal{N}_b, \text{ for some } j \in [1, k-1]. \tag{48}$$

As in the 1-D case, both the autocorrelation and the cumulants satisfy the AR recursion. Using (46) and/or (48), the $a(\mathbf{l})$'s can be obtained as the solution to an overdetermined system of linear equations. Optimal weighted least-squares and minimum-variance algorithms may be developed along the lines of [31] where the 1-D case is treated. The cumulant-based solution is appropriate if the additive noise is colored Gaussian, or if the transfer function has inherent all-pass factors. Additionally, if a multi-dimensional AR bispectrum estimator is desired (e.g., for detection of phase-coupling, or retrieval of harmonics in colored Gaussian noise, [27, 42]), the AR parameters should be estimated via (48), and not via (46). As in the pure correlation-based (1-D or 2-D) case, the solution is not guaranteed to be stable (roots within the unit $d$-circle); in practice, models of increasing orders are fitted to find a stable one.

## 5.2 Causal ARMA Models: MA Estimation

In this subsection, we describe two cumulant based methods for estimating the MA parameters. The first uses the cumulants of the residual (i.e., AR compensated) multidimensional sequence, $\tilde{z}(\mathbf{n}) = \hat{a}(\mathbf{n}) * z(\mathbf{n})$, whereas the second method directly uses the output cumulants.

In image processing applications, typically, a least-squares linear prediction approach is used to estimate the 'AR' parameters of the model [17]

$$\sum_{\mathbf{i}\in\mathcal{N}_a} a(\mathbf{i})y(\mathbf{m}-\mathbf{i}) = \tilde{g}(\mathbf{m}) \tag{49}$$

where $\mathcal{N}_a$ is the neighborhood set and the colored noise term $\tilde{g}(\mathbf{m})$ is the residual error in the least squares fit, rather than the driving noise of an AR model. For causal AR models ($\mathcal{N}_a$ in the first quadrant), our approach also yields an MA characterization of the colored residual error, if $\tilde{z}$ is replaced by $\tilde{g}$. In the 2-D case, ARMA modeling is a good approximation, but not as general as the models considered in [18].

### 5.2.1 The Residual C(q,k) Method

We compute the residual (i.e., AR compensated) time series, using the estimated AR parameters,

$$\begin{aligned} \tilde{z}(\mathbf{m}) &= \sum_{\mathbf{i}\in\mathcal{N}_a} \hat{a}(\mathbf{i})z(\mathbf{m}-\mathbf{i}) \\ &= \sum_{\mathbf{i}\in\mathcal{N}_b} b(\mathbf{i})w(\mathbf{m}-\mathbf{i}) + \tilde{g}(\mathbf{m}) \end{aligned} \tag{50}$$

where the second equality follows if $\hat{a} = a$, and $\tilde{g}(\mathbf{m})$ is now colored Gaussian noise (or at any rate, given our modeling assumptions on the output noise, its third-order cumulant vanishes).

The third-order cumulant of the residual MA time-series is given by

$$C_{3\tilde{z}}(\mathbf{m};\mathbf{n}) = \gamma_{3w} \sum_{\mathbf{i}\in\mathcal{N}_b} b(\mathbf{i})b(\mathbf{i}+\mathbf{m})b(\mathbf{i}+\mathbf{n}) \tag{51}$$

which follows directly from (50), since $\tilde{g}(\mathbf{n})$ is Gaussian, and $w(\mathbf{n})$ is third-order white. Evaluating (51) with $\mathbf{m} = \mathbf{q} = (q1, q2, ..., qd)$, we obtain

$$C_{3\tilde{z}}(\mathbf{q};\mathbf{n}) = \gamma_{3w} b(\mathbf{0})b(\mathbf{q})\, b(\mathbf{n}) \tag{52}$$

And assuming $b(\mathbf{0}) = 1$, we have

$$b(\mathbf{n}) = \frac{C_{3\tilde{z}}(\mathbf{q};\mathbf{n})}{C_{3\tilde{z}}(\mathbf{q};\mathbf{0})} \tag{53}$$

Equation (53) provides a closed-form solution for the MA parameters in terms of the third-order cumulants. Once the ARMA parameters have been obtained as described here, a parametric phase estimate of the model is easily computed. Finally, the 3-rd order cumulant of the input, $\gamma_{3w}$, can also be obtained from (52), as

$$\gamma_{3w} = \frac{C_{3\tilde{z}}(\mathbf{q};\mathbf{0})}{b(\mathbf{q})} \tag{54}$$

See [10] for corresponding results in the 1-D case. The cumulant of the residual time series may be computed from the output cumulants, and the estimated AR parameters, without explicitly computing the residual series itself. Thus, from (50) and the definition of $C_{3y}$, we obtain

$$C_{3\tilde{z}}(\mathbf{m};\mathbf{n}) = \sum_{\mathbf{i};\mathbf{j}\in\mathcal{N}_\alpha} \hat{\alpha}(\mathbf{i},\mathbf{j})C_{3z}(\mathbf{m}-\mathbf{i};\mathbf{n}-\mathbf{j}) \tag{55}$$

where

$$\hat{\alpha}(\mathbf{i};\mathbf{j}) = \sum_{\mathbf{k}\in\mathcal{N}_a} \hat{a}(\mathbf{k})\, \hat{a}(\mathbf{k}+\mathbf{i})\, \hat{a}(\mathbf{k}+\mathbf{j}) \tag{56}$$

and $\mathcal{N}_\alpha$ is the region of support of $\hat{\alpha}(\mathbf{i},\mathbf{j})$. Comparing (56) with (14), we see that $\hat{\alpha}(\mathbf{i};\mathbf{j})$ behaves like the third-order cumulant of a MA process; consequently, it enjoys all of the symmetry properties of third-order cumulants (see Section 3); these symmetry properties may be used to simplify the computation of the $\hat{\alpha}(\mathbf{i};\mathbf{j})$'s and the evaluation of the double convolution in (55).

The algorithm described above uses the residual cumulants at a fixed lag, namely $\mathbf{q} = (q1, q2, ..., qd)$. Other shift vectors can be used to compute the $b(\mathbf{i})$'s, and the corresponding estimates can be averaged to reduce the variance of the final MA parameter estimates. Further, the algorithm extends directly to higher-order cumulants, i.e.,

$$b(\mathbf{n}) = \frac{C_{k\tilde{z}}(\mathbf{q}; ...; \mathbf{q}; \mathbf{n})}{C_{k\tilde{z}}(\mathbf{q}; ...; \mathbf{q}; \mathbf{0})}, \quad \gamma_{kw} = \frac{C_{k\tilde{z}}(\mathbf{q}; \mathbf{0}; ...; \mathbf{0})}{b(\mathbf{q})} \tag{57}$$

where residual cumulants may be computed via

$$C_{k\tilde{z}}(\mathbf{n}_1; ...; \mathbf{n}_{k-1}) = \sum_{\mathbf{i}_1, ..., \mathbf{i}_{k-1} \in \mathcal{N}_\alpha} \hat{\alpha}(\mathbf{i}_1; ...; \mathbf{i}_{k-1}) C_{kz}(\mathbf{n}_1 - \mathbf{i}_1; ...; \mathbf{n}_{k-1} - \mathbf{i}_{k-1}) \tag{58}$$

and,

$$\hat{\alpha}(\mathbf{i}_1; ...; \mathbf{i}_{k-1}) = \sum_{\mathbf{n} \in \mathcal{N}_a} (\prod_{l=0}^{k-1} \hat{a}(\mathbf{n} + \mathbf{i}_{k-1})) \tag{59}$$

where $\mathbf{i}_0 = \mathbf{0}$.

### 5.2.2 The q-slice algorithm

The residual C(q,k) algorithm requires the computation of the cumulants of the residual 2-D sequence; an alternative method that directly uses the output cumulants will now be developed [43]. We will use $l$-th order cumulants in the following. Let

$$\mathrm{C}_l(\mathbf{i};\mathbf{j}) \equiv C_{ly}(\mathbf{0}; ...; \mathbf{0}; \mathbf{i}; \mathbf{j}) \tag{60}$$
$$= \sum_{\mathbf{k}} h^{l-2}(\mathbf{k}) h(\mathbf{k}+\mathbf{i}) h(\mathbf{k}+\mathbf{j}) \tag{61}$$

Define the function $f_l$ by

$$f_l(\mathbf{i};\mathbf{j}) \equiv \sum_{\mathbf{k} \in \mathcal{N}_a} a(\mathbf{k}) \mathrm{C}_m(\mathbf{i}; \mathbf{j} - \mathbf{k}) \tag{62}$$

Substituting for $\mathrm{C}_l$ from (61) into (62) yields,

$$f_l(\mathbf{i};\mathbf{j}) = \sum_{\mathbf{k}} h^{l-2}(\mathbf{k}) h(\mathbf{k}+\mathbf{i}) b(\mathbf{k}+\mathbf{j}) \tag{63}$$

where we have interchanged summations to obtain the last equality. Letting $\mathbf{j} = \mathbf{q} = (q1, q2, ..., qd)$ in (63) yields,

$$f_m(\mathbf{i}; \mathbf{q}) = b(\mathbf{q}) h(\mathbf{i}) \tag{64}$$

Assuming $h(\mathbf{0}) = 1$, we obtain from (64), $f_m(\mathbf{0};\mathbf{q}) = b(\mathbf{q})$; hence,

$$h(\mathbf{i}) = f_m(\mathbf{i};\mathbf{q})/f_m(\mathbf{0};\mathbf{q}) \tag{65}$$

Letting $\mathbf{i}$ take all the values in $\mathcal{N}_b$ yields all the $h(\mathbf{i})$'s; the $b(\mathbf{i})$'s are obtained via

$$\sum_{\mathbf{k}} a(\mathbf{k})h(\mathbf{i}-\mathbf{k}) = b(\mathbf{i}) \tag{66}$$

This 'q-slice' algorithm [39, 43] yields the impulse response coefficients, rather than the MA parameters directly. An alternative '2-slice' algorithm that uses 2-slices of the output cumulant (possibly of different orders) is described in detail in [39, 43], and is described in Appendix A for 2-D models. We note that this algorithm is not restricted to (quarter-plane) causal models. Additionally, this algorithm involves only one summation, as opposed to the $k$ summations in the method presented earlier.

#### 5.2.3 The Residual RC Algorithms

For a MA process (such as the AR compensated residual time-series), we have

$$C_{ky}(\mathbf{i}_1,...,\mathbf{i}_{k-1}) = \sum_{\mathbf{i}\in\mathcal{N}_b} b(\mathbf{i})b(\mathbf{i}+\mathbf{i}_1)\cdots b(\mathbf{i}+\mathbf{i}_{k-1}) \tag{67}$$

For simplicity assume that $\mathcal{N}_b = [0, q1]^d$, $b(\mathbf{0}) = 1$, and $b(\mathbf{q}) \neq 0$. Then, the following relationships between the second and third-order cumulants are easily obtained (see [49] for the 1-D case):

$$\begin{aligned}
\sum_{\mathbf{k}} b(\mathbf{k})b(\mathbf{k}+\mathbf{m})C_{2y}(\mathbf{k}-\mathbf{n}) &= \epsilon\sum_{\mathbf{k}} b(\mathbf{k})C_{3y}(\mathbf{k}-\mathbf{n},\mathbf{m}) &(68)\\
\sum_{\mathbf{k}} b(\mathbf{k})b(\mathbf{k}+\mathbf{m}_1)C_{3y}(\mathbf{k}-\mathbf{n},\mathbf{m}_0) &= \sum_{\mathbf{k}} b(\mathbf{k})b(\mathbf{k}+\mathbf{m}_0)C_{3y}(\mathbf{k}-\mathbf{n},\mathbf{m}_1) &(69)\\
b(\mathbf{0})b(\mathbf{q})C_{2y}(\mathbf{n}) &= \epsilon\sum_{\mathbf{k}} b(\mathbf{k})C_{3y}(\mathbf{k}-\mathbf{n},\mathbf{q}) &(70)\\
C_{3y}(\mathbf{n},\mathbf{q}) &= b(\mathbf{q})C_{3y}(\mathbf{0},\mathbf{n}) - \sum_{\mathbf{i}\neq\mathbf{0}} b^2(\mathbf{i})C_{3y}(\mathbf{i}+\mathbf{n},\mathbf{q}) &(71)
\end{aligned}$$

A variety of MA parameter estimation algorithms may be developed from these relationships. For example, (68) may be viewed as a set of linear equations in the unknowns, $\epsilon b(\mathbf{k})$ and $b(\mathbf{k})b(\mathbf{k}+\mathbf{m})$; and (69) may be viewed as a set of linear equations in the unknowns, $b(\mathbf{k})b(\mathbf{k}+\mathbf{m}_0)$ and $b(\mathbf{k})b(\mathbf{k}+\mathbf{m}_1)$. Equation (70) yields a system of linear equations in the $b(\mathbf{k})$'s; the $b(\mathbf{k})$'s may also be estimated recursively. Finally, (71) leads to recursive and least-squares estimates for the $b^2(\mathbf{i})$'s. Since (68) and (69) involve the correlation, they may be used in the additive white noise case by avoiding the zero-lag term, or in the colored Gaussian noise case, by estimating correlations through higher-order cumulants, as discsussed in Section 3.2.

The above relationships may be generalized; for example,

$$\sum_{\mathbf{k}} b(\mathbf{k})b(\mathbf{k}+\mathbf{m}_1)C_{3y}(\mathbf{k}-\mathbf{m},\mathbf{k}-\mathbf{n}) = \epsilon\sum_{\mathbf{i}} b(\mathbf{i})C_{4y}(\mathbf{i}-\mathbf{n},\mathbf{i}-\mathbf{m},\mathbf{m}_1) \quad (72)$$

$$b(\mathbf{0})b(\mathbf{m})b(\mathbf{q})C_{2y}(\mathbf{n}) = \sum_{\mathbf{k}} b(\mathbf{k})C_{4y}(\mathbf{k}-\mathbf{n},\mathbf{m},\mathbf{q}) \quad (73)$$

$$\sum_{\mathbf{i}} b(\mathbf{i})C_{4y}(\mathbf{i}+\mathbf{n},\mathbf{m},\mathbf{q}) = \epsilon' C_{3y}(m,n)$$

$$\sum_{\mathbf{i}} b^3(\mathbf{i})C_{4y}(\mathbf{i}+\mathbf{m},\mathbf{n},\mathbf{q}) = b(\mathbf{q})b(\mathbf{n})C_{4y}(\mathbf{0},\mathbf{0},\mathbf{j}+\mathbf{m})$$

$$\sum_{\mathbf{i}} b^3(\mathbf{i})C_{4y}(i+m,q,q) = b^2(q)C_{4y}(0,0,j+m) \quad (74)$$

Another interesting relationship may be obtained as follows,

$$C_{3y}(\mathbf{i};\mathbf{q}) = \gamma_{3w} b(\mathbf{0})b(\mathbf{q})b(\mathbf{i})$$

$$b(\mathbf{j})C_{3y}(\mathbf{i};\mathbf{q}) = b(\mathbf{i})C_{3y}(\mathbf{j};\mathbf{q}) \quad (75)$$

$$C_{4y}(\mathbf{i},\mathbf{j},\mathbf{q}) = \gamma_{4w} b(\mathbf{0})b(\mathbf{q})b(\mathbf{i})b(\mathbf{j})$$

$$b(\mathbf{j})C_{4y}(\mathbf{i},\mathbf{k},\mathbf{q}) = b(\mathbf{i})C_{4y}(\mathbf{j},\mathbf{k},\mathbf{q}) \quad (76)$$

Equation (76) leads to an over-determined system of linear equations in the $b(\mathbf{i})$'s [7]. The relationships in (68)-(76) and resultant MA parameter estimation algorithms are readily extended to higher-order cumulants.

## 5.3 Adaptive Estimation

For causal cumulant-based image modeling, the adaptive double-lattice structures developed in [39, 41] for 1-D causal models are readily extended to the 2-D case.

## 5.4 Non-causal Models

In deriving (46) and (47), we had to assume that the system was causal. However, ARMA models used in image-processing tasks are usually non-causal ($\mathcal{N}_a$ arbitrary), [18]. Further, non-causal models find applications in image coding, modeling, representation and interpolation [17, Chs 6 and 11].

Because of the dependence of the current output on both the past and future values of the input, 'normal' equations of the form in (47) cannot be written for the non-causal case. Hence, we develop another method to handle the non-causal case.

From (16) and (44), we obtain

$$S_{3y}(\mathbf{u};\mathbf{v}) = \gamma_{3w}\frac{\mathcal{B}(\mathbf{u})\mathcal{B}(\mathbf{v})\mathcal{B}^*(\mathbf{u}+\mathbf{v})}{\mathcal{A}(\mathbf{u})\mathcal{A}(\mathbf{v})\mathcal{A}^*(\mathbf{u}+\mathbf{v})} \quad (77)$$

where $\mathcal{A}$ and $\mathcal{B}$ are the $z$-transforms of the finite-support, $d$-D sequences, $a(\mathbf{i})$ and $b(\mathbf{i})$ respectively. The time-domain equivalent of (77) is given by

$$\gamma_{3w}\beta(\mathbf{m},\mathbf{n}) = \sum_{\mathbf{i},\mathbf{j}\in\mathcal{N}_\alpha} \alpha(\mathbf{i};\mathbf{j})C_{3y}(\mathbf{m}-\mathbf{i};\mathbf{n}-\mathbf{j}) \tag{78}$$

$$= 0 \text{ if } (\mathbf{m},\mathbf{n}) \notin N_\beta \tag{79}$$

where

$$\alpha(\mathbf{m};\mathbf{n}) = \sum_{\mathbf{i}\in\mathcal{N}_a} a(\mathbf{i})a(\mathbf{i}+\mathbf{m})a(\mathbf{i}+\mathbf{n}) \tag{80}$$

$$\beta(\mathbf{m};\mathbf{n}) = \sum_{\mathbf{i}\in\mathcal{N}_b} b(\mathbf{i})b(\mathbf{i}+\mathbf{m})b(\mathbf{i}+\mathbf{n}) \tag{81}$$

and $\mathcal{N}_\alpha$ ($\mathcal{N}_\beta$) is the non zero region of support of $\alpha(\mathbf{m},\mathbf{n})$ ($\beta(\mathbf{m},\mathbf{n})$).

Using (78) and (79), we can set up a system of linear equations to solve for the unknown parameters, the $\{\alpha(\mathbf{m};\mathbf{n})\}$'s. Notice the similarity between (51), (80) and (81) . The $\alpha$'s ($\beta$'s) are the cumulants of a MA system whose parameters are the $a$'s ($b$'s). Thus, using (53), the AR parameters may be obtained in closed form as

$$a(\mathbf{n}) = \frac{\alpha(\mathbf{p},\mathbf{n})}{\alpha(\mathbf{p};\mathbf{0})} \tag{82}$$

where $\mathbf{p} = (p1, p2)$. Once the $\alpha$ parameters have been determined, (78) can be used to solve for the remaining unknowns, the $\{\beta(\mathbf{i},\mathbf{j})\}$'s, as well. Thus, analogous to (53), the MA parameters may be obtained as

$$b(\mathbf{n}) = \frac{\beta(\mathbf{q};\mathbf{n})}{\beta(\mathbf{q};\mathbf{0})} \tag{83}$$

The non-causal ARMA parameter estimation algorithm is summarized by:

1. Form the system of linear equations resulting from (79); its solution determines $\alpha(\mathbf{m};\mathbf{n})$.
2. Estimate $\gamma_{3w}\beta(\mathbf{m};\mathbf{n})$ from (78).
3. Estimate the ARMA parameters, $a(\mathbf{m})$ and $b(\mathbf{m})$, from (82) and (83).

As shown in Section 4.1, this algorithm directly extends to higher-order cumulants [see (58) and (59)].

Results for the 1-D case are presented in [13, 14]. One of the main points of the three-step algorithm is the recognition that once $\alpha(\mathbf{m},\mathbf{n})$ and $\beta(\mathbf{m},\mathbf{n})$ have been estimated, the ARMA parameter estimation problem reduces to two decoupled MA parameter estimation problems. Taking into account the tradeoff between numerical accuracy and computational complexity, any of several methods may be used to estimate the ARMA parameters from their 3rd-order cumulants. Additionally, the key feature of the method presented in (77)-(83), which applies to non-causal, non-minimum phase ARMA models, is that appropriate slices of the cumulants of an MA model play the role of input/output cross-correlation [10, 13, 40].

Note that this algorithm presents an unified approach to the estimation of the parameters of non-causal ARMA models: we do not have to decompose the original system into four sub-systems, each with support over a quarter-plane (strictly causal and anti-causal in the 1-D case). We do not need to know *a priori* whether the model is causal or non-causal. From (80) and (81), we see that any consistent MA parameter identification algorithm may be used to estimate both the AR and MA parameters. Further, simplifications are possible in the case of symmetric impulse responses. State-Space approaches, along the lines in [13] for 1-D systems, are under investigation.

Once an estimate of the p.s.f., $h(\mathbf{i})$, or its Z-transform, $H(\mathbf{z})$, is obtained, a parametric polyspectral estimator is readily obtained via (15). In Section 6, we show that our parametric estimators are consistent, and discuss additional algorithms for obtaining minimum-variance and optimal weighted least-squares estimates; hence, the corresponding parametric polyspectral estimates also enjoy low-variance properties.

# 6 Consistency - Order Selection - Noise effects

In practice, sample estimates of the cumulants (see Section 2.6 for details on how this is done) must be used in place of the theoretical cumulants in the algorithms in Sections 4 and 5. Arguing as in the 1-D case, (e.g., [14]), strong consistency can be established for the sample cumulant estimator of (8): If in (10), $h(\mathbf{i})$ is exponentially stable, $w(\mathbf{i})$ is stationary, i.i.d., with finite moments up to order $2k$, then $\hat{C}_{ky}(\mathbf{m}_1; ...; \mathbf{m}_{k-1})$ converges with probability one to its theoretical value $C_{ky}(\mathbf{m}_1; ...; \mathbf{m}_{k-1})$. Expressions for the asymptotic variance of the estimates remain to be developed, along the lines of [32, pp 721-722].

If $\underline{\theta}$ denotes the vector of the unknown ARMA parameters, and $\underline{C}$ is the set of cumulants used in any of the identification algorithms in Sections 4 and 5, then, $\underline{C} = F(\underline{\theta})$ is single-valued, because, given the parameters of the model, and, hence, the impulse response of the model, the cumulants are uniquely determined via (14).

An identification algorithm yields strongly consistent estimates of the form $\hat{\underline{\theta}} = G(\hat{\underline{C}}) \to G(\underline{C})$, provided $G$ is the inverse function of $F$ in a neighborhood containing $\underline{\theta}_o$, the true parameter vector. Since the algorithms proposed in this report involve linear equations of the form $A\underline{\theta} = \underline{b}$; existence and uniqueness of the inverse function $G$ is guaranteed, if we can show that $\text{rank}(A) = \dim(\underline{\theta})$.

Following [39], it may be shown, in the separable case, that the system of linear equations obtained from the 'normal' equations

$$\sum_{\mathbf{i} \in \mathcal{N}_a} a(\mathbf{i}) C_{ky}(\mathbf{m} - \mathbf{i}; \mathbf{n}; \mathbf{0}; ...; \mathbf{0}) = 0, \ \ \mathbf{m} \ni \mathcal{N}_b \tag{84}$$

with $n_1 = q1 - p1, ..., q1$, $n_2 = q2 - p2, ..., q2$, $m_1 = q1 + 1, ...., q1 + p1, ...$, and $m_2 = q2 + 1, ..., q2 + p2, ...$, has rank $p1 \times p2$. However, as in the case of the autocorrelation-based $d$-D normal equations, the identifiability of the AR parameters of a non-separable ARMA model based on finite slices of cumulants is an open problem.

Order determination, and hence, consistency of parameter estimates, for non-causal models is even more complicated; results developed for 1-D non-causal models in [13], may be extended to 2-D non-causal separable models. From a purely *modeling* point of view (i.e., approximate cumulant matching), this is a rather moot point; however, this is important from a *realization* point of view (i.e., exact cumulant matching). Empirical and AIC based order-determination algorithms are discussed in [19]; although, these are appropriate only for Gaussian processes, they may be useful indicators for non-Gaussian processes as well.

Finally, we note that a minimum-variance simultaneous ARMA parameter estimation algorithm (appropriate for the realization problem) can be developed along the lines of [31] where causal 1-D models are considered. In this algorithm, which involves non-linear minimization, the parameter vector is found as,

$$\hat{\underline{\theta}}_{\text{min var}} = \arg\min \ \ (\underline{C} - \hat{\underline{C}})^T \Sigma^{-1} (\underline{C} - \hat{\underline{C}}) \tag{85}$$

where $\Sigma = E\{\hat{\underline{C}}\hat{\underline{C}}^T\}$, and $\underline{C}$ and $\Sigma$ are both functions of the unknown parameter vector $\underline{\theta}_o$. The theoretical cumulants $\underline{C}(\underline{\theta})$, corresponding to a proposed model $\underline{\theta}$, can be computed, for example, by truncating the infinite summation in (14). An estimate of $\Sigma(\underline{\theta}_o)$ may be obtained following the procedure in Section 2.6. This procedure implicitly assumes that there is a one-to-one mapping between $\underline{C}$ and $\underline{\theta}$. The algorithm in (85) is asymptotically optimal in the sense that the resulting parameter estimates have the smallest variance in the class of algorithms based on the set of cumulant statistics $\underline{C}$; this algorithm *does not* achieve the Cramer-Rao lower bound (the input p.d.f. is not assumed to be known in our model), and is computationally demanding; good initial estimates for the minimum-variance algorithm may be found via the linear algorithms of Sections 4 and 5. Further, for the least-squares algorithms in Sections 4 and 5, optimal weighted least squares estimates can be obtained, along the lines of [31], where the 1-D case is discussed; this algorithm yields the minimum variance parameter estimate in the class of weighted least-squares algorithms.

In theory, third-order cumulants are not affected by additive, possibly colored, noise with symmetric distributions (e.g., Gaussian), because the third-order cumulants of such a process vanish. In practice, when sample averaging is used to compute the cumulants, using finite records, the estimates may show high variance, depending on the distributions of the input and output noises. Similar comments hold when $k$-th order cumulants are used, and the additive noise is colored Gaussian, or has zero $k$-th order cumulant.

In the deterministic case, the observed $d$-D sequence is modeled as the impulse response of a $d$-D model; hence, with $\gamma_{kw} = 1$, the algorithms developed for the stochastic case extend to the deterministic case, which is potentially useful for reconstructing multidimensional signals in noise, analyzing transients, and designing multidimensional filters. The only exception is that the observed $d$-D sequence cannot be segmented for computing sample cumulants. In order to suppress the effects of additive noise, the mean must be removed from each observed record; hence, algorithms that assume $H(\mathbf{0}) \neq 0$ cannot be used in this case, and the alternatives discussed earlier must be used instead. Additionally, fourth-order cumulant based deterministic modeling differs considerably from the third-order case; this is discussed in detail in [39]), where it is shown that if cumulants of deterministic signals (in noise) are defined analogous to those for stochastic signals, with sample averages replacing the expectation operator, then, the $k$-th order moment of the deterministic sequence is identical to $k$-th order cumulant of a stochastic sequence obtained by exciting the linear model with an i.i.d. sequence.

# 7 Conclusions

Cumulants, polyspectra and polycepstra of multidimensional non-Gaussian linear processes were introduced and studied in this report. Unique (up to a linear term) phase estimates were obtained using single-slice, linear equation-based, recursive, and closed-form phase reconstruction algorithms which employ the bispectral phase. Several consistent cumulant-based algorithms were also developed for parameter estimation of multidimensional causal ARMA $(\mathbf{p},\mathbf{q})$ processes which may be corrupted by additive colored Gaussian noise of unknown covariance. These algorithms involve the solution of Yule-Walker type equations for AR estimation. For MA estimation, one method employs a single cumulant slice of the residual series, whereas the other method uses $\mathbf{q}$ output cumulant slices for estimating the MA part through the impulse response coefficients. Several other algorithms using one or more slices of cumulants of different orders were also proposed. A unified cumulant based parameter estimation approach, which does not require a priori knowledge of whether the multidimensional ARMA model is (non-)causal, (non-)minimum phase or (non-)separable, was also developed. Although the presentation focussed on ARMA estimates obtained with linear equations, asymptotic minimum-variance cumulant matching approaches involving non-linear optimization were also discussed. It appears that cumulants and polyspectra are useful tools for (non-)parametric modeling of multidimensional signals, because processing is carried out in a high SNR domain, which preserves all the relevant information about the signal.

# References


[1] H. Bartelt and B. Wirnitzer, "Shift-invariant imaging of photon-limited data using bispectral analysis", *Optics Comm.*, v53, 13-16, Feb. 1985.

[2] H. Bartelt, A.W. Lohmann and B. Wirnitzer, "Phase and amplitude recovery from bispectra", *Applied Optics*, v23, 3121-29, 1984.

[3] D.R. Brillinger, "An Introduction to Polyspectra", *Ann. Math. Statist.*, v.36, 1351-74, 1965.

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

[5] J. Capon, "High-resolution frequency-wavenumber spectral analysis", *Proc. IEEE*, v. 57, 1408-18, 1969.

[6] R. Chellappa and R.L. Kashyap, "Digital image restoration using spatial interaction models", *IEEE Trans ASSP*, 30, 461-472, June 1982.

[7] P. Comon, "Separation of sources using higher-order moments", to be presented at *SPIE Conf. Adv. Alg. and Arch. for Signal Processing*, San Diego, CA, Aug 1989.

[8] S.A. Dianat and M.R. Raghuveer, "Two-dimensional non-minimum phase signal reconstruction", *Proc. Workshop on Higher-Order Spectral Analysis*, 112-117, Vail, CO, June 1989.

[9] D.E. Dudgeon and R.M. Mersereau, *Multidimensional digital signal processing*, Prentice-Hall, New jersey, 1984.

[10] G.B. Giannakis, "Cumulants: A powerful tool in signal processing", *Proc. IEEE*, v74, 1333-4, Sept 1987.

[11] G.B. Giannakis, "Signal reconstruction from multiple correlations: Frequency- and time-domain approaches", to appear in *J. Opt. Soc. Amer.*, 682-697, May 1989.

[12] G.B. Giannakis and A. Delopoulos, "Non-parametric estimation of autocorrelation and spectra using cumulants and polyspectra", University of Virginia Report No UVA/192/444b/EE89/113, June 1989; also submitted to *IEEE Trans. ASSP*.

[13] G. B. Giannakis and A. Swami, "New Results on State-Space and Input-Output Identification of Non-Gaussian Processes using Cumulants", *Proc. SPIE-87*, Vol. 826, *Advanced Algorithms and Architectures for Signal Processing II*, 199-204, San Diego, CA, Aug. 1987.

[14] G.B. Giannakis and A. Swami, "On Estimating Non-Causal Non-Minimum Phase ARMA Models of Non-Gaussian Processes", *Proc. IV IEEE ASSP Workshop on Spectrum Estimation and Modeling*, 187-192, Minneapolis, MN, 187-192, Aug. 1988; to appear in *IEEE Trans. on ASSP*, March 1990.

[15] T.E. Hall and S.G. Wilson, "Stochastic image modeling using cumulants with application to predictive image coding", *Proc. Workshop on Higher-Order Spectral Analysis*, 239-244, Vail, CO, June 1989.

[16] *IEEE Trans. Biomed. Engg.*, Special Issue on Computerized Medical Imaging, BME-28, Feb 1981.

[17] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989.

[18] R.L. Kashyap, "Characterization and Estimation of Two-Dimensional ARMA Models", *IEEE Trans. Info Theory*, IT-30, 736-45, 1984.

[19] R.L. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial interaction models of images", *IEEE Trans. Info Theory*, 60-72, 1983.

[20] S.Y. Kung, K.S. Arun and D.V. Bhaskar Rao, "State Space and SVD Based Approximation Methods for the Harmonic Retrieval Problem", *J. Opti. Soc. Amer.*, 73, 1799-1811, 1983.

[21] K.S. Lii and K.N. Helland, "Cross-bispectrum computation and variance estimation", *ACM Trans. Math Software*, v7, 284-94, 1981.

[22] K.S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes", *Ann. Stat.*, v10, 1195-1208, 1982.

[23] K.S. Lii, K.N. Helland and M. Rosenblatt, "Estimating three-dimensional energy transfer in isotropic turbulence", *J. Time Series Analysis*, 1, 1982.

[24] A.W. Lohmann and B. Wirnitzer, "Triple Correlations", *Proc. IEEE*, 72, 889-901, 1984.

[25] A.W. Lohmann, G. Weigelt and B. Wirnitzer, "Speckle Masking in Astronomy: Triple Correlation Theory and Applications", *Applied Optics*, 22, 4028-4037, 1983.

[26] T. Matsuoka and T.J. Ulrych, "Phase estimation using the bispectrum", *Proc. IEEE*, v72, 1403-1411, 1984.

[27] C.L. Nikias and M. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework", *Proc. IEEE*, 75, 869-891, 1987.

[28] M.J. Northcott, G.R. Ayers, and J.C. Dainty, "Algorithms for image reconstruction from phase-limited data using triple correlation", *J. Opt. Soc. Amer.*, v. 5, 986-992, 1988.

[29] A.V. Oppenheim and J.S. Lim, "The importance of phase in signals", *Proc. IEEE*, v69, 529-41, 1981.

[30] A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, New Jersey: Prentice-Hall, 1975.

[31] B. Porat and B. Friedlander, "Optimal Estimation of MA and ARMA Parameters of Non-Gaussian Processes From Higher-Order Cumulants", *Proc. IV IEEE ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, MN, 208-212, Aug. 1988.

[32] M.B. Priestley, "Spectral Analysis and Time Series", *Academic Press*, London, 1981.

[33] *Proc. IEEE*, Special Issue on Computerized Tomography, 71, March 1983.

[34] M.R. Raghuveer and S.A. Dianat, "Detection of non-linear phase coupling in multidimensional stochastic processes", *Intl. Conf. on Adv. in Contr. and Comm.*, pp 729-732, Baton Rouge, LA, October 1988.

[35] M.R. Raghuveer, S.A. Dianat and G. Sundaramoorthy, "Reconstruction of non-minimum phase multidimensional signals using the bispectrum", *Proc. III SPIE Conf. on Visual Communication and Image Processing*, Cambridge, MA, Nov 1988.

[36] E.A. Robinson, "Spectral Approach to Geophysical Inversion by Lorentz, Fourier, and Radon Transforms", *Proc. IEEE*, Special Issue on Spectral Estimation, 1039-1054, Sept 1982.

[37] M. Rosenblatt, *Stationary Sequences and Random Fields*, Boston: Birkhauser, 1985.

[38] B.M. Sadler, "Shift and rotation invariant object reconstruction using the bispectrum", *Proc. Workshop on Higher-Order Spectral Analysis*, 106-111, Vail, CO, June 1989.

[39] A. Swami, *System Identification Using Cumulants*, Ph.D. Dissertation, University of Southern California, Los Angeles, Oct 1988.

[40] A. Swami and G.B. Giannakis, "ARMA Modeling and Phase Reconstruction of Multidimensional Non-Gaussian Processes Using Cumulants", *Proc. ICASSP-88*, New York, NY, 729-732, April 1988.

[41] A. Swami and J.M. Mendel, "Adaptive System Identification Using Cumulants", *Proc. ICASSP-88*, 2248-51, New York, NY, April 1988; also *USC-SIPI Report-117*, University of Southern California, Los Angeles, Dec. 1987.

[42] A. Swami and J.M. Mendel, "Cumulant Based Approach to the Harmonic Retrieval Problem", *Proc. ICASSP-88*, New York, NY, 2264-67, April 1988.

[43] A. Swami and J.M. Mendel, "ARMA Parameter Estimation Using Only Output Cumulants", *Proc. IV IEEE ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, MN, 193-198, Aug. 1988.

[44] A.M. Tekalp and A.T. Erdem, "Two-dimensional higher-order spectrum factorization with applications in non-Gaussian image modeling", *Proc. Workshop on Higher-Order Spectral Analysis*, 186-190, Vail, CO, June 1989.

[45] A.M. Tekalp and A.T. Erdem, "Higher-order spectrum factorization in one and two dimensions with applications in signal modeling and non-minimum phase system identification", *IEEE Trans on ASSP*, October 1989, (to appear).

[46] A.M. Tekalp, H. Kaufman and J.W. Woods, "Identification of image and blur parameters for the restoration of non-causal filters", *IEEE Trans. ASSP*, v. 34, 963-972, 1986.

[47] P.V. Tryon, *The Bispectrum and Higher-Order Spectra: A Bibliography*, Nat. Bur. Stand., Washington, DC, NBS Tech. Note 1036, April 1981.

[48] M. Tsatsanis and G.B. Giannakis, "Object detection and classification using matched filtering and higher-order statistics", *Proc. VI Workshop on Multi-dimensional Signal Processing*, Monterey, CA, September 1989.

[49] J.K. Tugnait, "Approaches to F.I.R. system identification with noisy data using higher-order statistics", *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, CO, 13-18, June 1989.

[50] J. K. Tugnait, "Identification of Linear Stochastic Systems via Second- and Fourth-Order Cumulant Matching", *IEEE Trans. Info Thy.*, IT-33, 393-407, 1987.

# Appendix A: The q-slice and 2-slice algorithms

Detailed derivation of the q-slice and 2-slice algorithms for 2-D processes are given here.

## A.1 The Model

The system is described by the 2-D ARMA model:

$$\sum_{i=0}^{p1}\sum_{j=0}^{p2} a(i,j)y(k-i,t-j) = \sum_{i=0}^{q1}\sum_{j=0}^{q2} b(i,j)u(k-i,t-j) \tag{A.1}$$

where the order $\{(p1,p2),(q1,q2)\}$, is assumed known. The noisy output is given by,

$$z(k,t) = y(k,t) + w(k,t) \tag{A.2}$$

The impulse response $h(m,n)$ satisfies

$$\begin{aligned}\sum_{i=0}^{p1}\sum_{j=0}^{p2} a(i,j)h(m-i,n-j) &= \sum_{i=0}^{q1}\sum_{j=0}^{q2} b(i,j)\delta(m-i,n-j) \\ &= b(m,n)\end{aligned} \tag{A.3}$$

The $m$-th order cumulants are related to the impulse response, via [40]

$$\begin{aligned}&C_{m,y}\{(t_{i,1},t_{j,1}),(t_{i,2},t_{j,2}),...,(t_{i,m-1},t_{j,m-1})\} \\ &= \gamma_{m,u}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} h(i,j)h(i+t_{i,1},j+t_{j,1})\cdots h(i+t_{i,m-1},j+t_{j,m-1})\end{aligned} \tag{A.4}$$

## A.2 AR Parameter Estimation

Let

$$\mathrm{C}_m((t_1,t_2);(\tau_1,\tau_2)) \equiv C_{m,y}((0,0),...,(0,0),(t_1,t_2),(\tau_1,\tau_2)) \tag{A.5}$$

$$= \gamma_{m,u}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} h^{m-2}(i,j)h(i+t_1,j+t_2)h(i+\tau_1,j+\tau_2) \tag{A.6}$$

Define the function $f_m$ by

$$f_m((t_1,t_2);(\tau_1,\tau_2)) \equiv \sum_{i=0}^{p1}\sum_{j=0}^{p2} a(i,j)\mathrm{C}_m((t_1,t_2);(\tau_1-i,\tau_2-j)) \tag{A.7}$$

Substituting for $\mathrm{C}_m$ from (A.6) into (A.7) yields,

$$\begin{aligned}&f_m((t_1,t_2);(\tau_1,\tau_2)) \\ &= \gamma_{m,u}\sum_{i=0}^{p1}\sum_{j=0}^{p2} a(i,j)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} h^{m-2}(k_1,k_2)h(k_1+t_1,k_2+t_2)h(k_1+\tau_1-i,k_2+\tau_2-j) \\ &= \gamma_{m,u}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} h^{m-2}(k_1,k_2)h(k_1+t_1,k_2+t_2)b(k_1+\tau_1,k_2+\tau_2)\end{aligned} \tag{A.8}$$

where we have interchanged summations and used (A.3) to obtain the last equality. The right-hand side of (A.8) is zero, when $\tau_1 > q1$ or $\tau_2 > q2$; thus, the AR parameters can be estimated as the solution to a system of linear equations.

## A.3 The q-slice Algorithm

Since $b(i,j)$ has finite support over $[0,q1]\times[0,q2]$, letting $\tau_1 = q1$ and $\tau_2 = q2$ in (A.8) yields,

$$f_m((t_1,t_2);(q1,q2)) = \gamma_{m,u}h^{m-2}(0,0)b(q1,q2)h(t_1,t_2) \quad \text{(A.9)}$$

Assuming $h(0,0) = 1$, we obtain from (A.9), $f_m((0,0);(q1,q2)) = \gamma_{m,u}b(q1,q2)$; hence,

$$h(t_1,t_2) = f_m((t_1,t_2);(q1,q2))/f_m((0,0);(q1,q2)) \quad \text{(A.10)}$$

which is the 2-D counterpart to (30). Letting $t_1 = 0,1,...,q1$ and $t_2 = 0,1,...,q2$, yields $h(n,m)$; the $b(n,m)$'s are then obtained via (A.3).

## A.4 The 2-slice Algorithm

Equation (A.8) may be written as

$$f_m((t_1,t_2);(\tau_1,\tau_2)) = \gamma_{m,u}\sum_{k_1=0}^{q1}\sum_{k_2=0}^{q2} b(k_1,k_2)h^{m-2}(k_1-\tau_1,k_2-\tau_2)h(k_1+t_1-\tau_1,k_2+t_2-\tau_2) \quad \text{(A.11)}$$

where we have used the causality of $h(n,m)$.

The $(t_1,t_2)$ slice of the cumulant is obtained by fixing $t_1$ and $t_2$ in (A.5); this $(t_1,t_2)$ 'slice' is a function of two parameters, because we are dealing with 2-D processes. Suppose $\tau_2 = q2$. Then, Eq. (A.11) is

$$f_m((t_1,t_2);(\tau_1,q2)) = \gamma_{m,u}\sum_{k=0}^{q1} b(k,q2)h^{m-2}(k-\tau_1,0)h(k+t_1-\tau_1,t_2) \quad \text{(A.12)}$$

When $t_2 = 0$ in (A.12), we essentially have a 1-D problem involving the $b(k,q2)$'s and $h(k,0)$'s. In the 1-D case, we have [39]

$$f_m(t;\tau) \;:=\; \sum_{k=0}^{p} a(k)C_{m,y}(\tau-k,t,0,\cdots,0) \quad \text{(A.13)}$$

$$= \gamma_{m,u}\sum_{k=0}^{q} b(k)h(k-\tau+t)h^{m-2}(k-\tau) \quad \text{(A.14)}$$

The $b(k)$'s and $h(k)$'s can be recursively estimated as follows:
For $\tau = q, q-1, ..., q/2$, with $h(0) = 1$,

$$\gamma_{m,u}b(\tau) = f_m(0;\tau) - \sum_{k=\tau+1}^{q}\gamma_{m,u}b(k)h^{m-1}(k-\tau) \quad \text{(A.15)}$$

$$h(q-\tau+1) = \frac{f_m(1;\tau) - \sum_{k=\tau}^{q-1}\gamma_{m,u}b(k)h^{m-2}(k-\tau)h(k+1-\tau)}{\gamma_{m,u}b(q)h^{m-2}(q-\tau)} \quad \text{(A.16)}$$

and, finally,

$$b(k) = \sum_{i=0}^{k} a(i)h(k-i), \qquad k = 0, 1, ..., q/2 \tag{A.17}$$

Equation (A.17) evaluated at $k = q/2$ yields $b(q/2)$, whereas (A.15) evaluated at $\tau = q/2$ yields $\gamma_{m,u}b(q/2)$; hence, $\gamma_{m,u}$ can be estimated, and the estimates $\gamma_{m,u}b(\tau)$, obtained from (A.15), can then be corrected for the scale factor $\gamma_{mu}$.

Comparing (A.12) (with $t_2 = 0$) and (A.14) (for the 1-D case), it is clear that the recursive solution given by (A.15)-(A.17) can be used to estimate $\gamma_{m,u}$, $b(k, q2)$ and $h(k, 0)$, for $k = 0, 1, ..., q1$; this involves the two 'slices' specified by $(t_1, t_2) = (0, 0)$ and $(t_1, t_2) = (1, 0)$. Since, $\gamma_{m,u}$ has been estimated, we will assume that the sample cumulants have been scaled by it to obtain an effective $\gamma_{m,u}$ of unity in the following. Then, when $t_1 = 0$ and $t_2 = 1$ [i.e., using the $(0, 1)$ slice], (A.12) yields $h(k, 1), k = 0, ..., q1$, via

$$\begin{aligned} f_m((0,1);(\tau_1, q2)) &= \sum_{k=\tau_1}^{q1-1} b(k, q2)h^{m-2}(k-\tau_1, 0)h(k-\tau_1, 1) \\ &\quad + b(q1, q2)h^{m-2}(q1-\tau_1, 0)h(q1-\tau_1, 1) \end{aligned}$$

with $\tau_1 = q1, q1-1, ..., 0$.

Now let $\tau_2 = q2 - 1$. Then, the right-hand side of (A.11) is

$$\begin{aligned} &\sum_{k=0}^{q1} b(k, q2-1)h^{m-2}(k-\tau_1, 0)h(k+t_1-\tau_1, t_2) \\ + &\sum_{k=0}^{q1} b(k, q2)h^{m-2}(k-\tau_1, 1)h(k+t_1-\tau_1, t_2+1) \end{aligned}$$

When $t_1 = t_2 = 0$, the only unknowns in the above expression are the $b(k, q2-1)$ terms, since $h(k, 0)$, $h(k, 1)$ and $b(k, q2)$ have already been estimated. Thus, the $(t_1, t_2) = (0, 0)$ slice then directly yields $b(k, q2-1)$, by letting $\tau_1 = q1, ..., 0$. The $(0, 1)$ slice then yields the $h(k, 2)$'s in a similar manner. Continuing in this fashion, all the $h(i, j)$'s and $b(i, j)$'s can be obtained.

It is clear that the method can be extended to $m$-D processes.

If the MA coefficients are symmetric, i.e., $b(k, l) = b(l, k)$ (usually, a reasonable assumption in image processing), then, only two slices are required [i.e., (j1,j2) = (0,0) and (0,1)]; if the $b(n, m)$'s are separable, then the 1-D algorithm applies directly.