Linear Prediction, Maximum Flatness, Maximum Entropy, and AR Polyspectral Estimation

Chong-Yung Chi, *Senior Member, IEEE*

***Abstract*—It is well known in second-order statistics based autoregressive (AR) spectral estimation that the linear prediction spectral estimator is equivalent to the maximum entropy spectral estimator and they are also equivalent to the maximum spectral flatness spectral estimator for AR processes of known order. In this paper, we present a new theoretical background for the polyspectral estimation and modeling of non-Gaussian AR processes which includes a new higher order statistics (HOS) based linear prediction error filter and associated linear prediction polyspectral estimator, a maximum polyspectral flatness polyspectral estimator, a maximum higher order entropy polyspectral estimator, as well as the equivalencies among these polyspectral estimators.**

I. Introduction

PARAMETRIC spectral estimation of a wide-sense stationary process [6]-[10] can be found in various science and engineering areas such a speech processing, seismology, radar, sonar, radio astronomy, biomedicine, image processing, vibration analysis, and oceanography. The autoregressive (AR) spectral estimator is the most popular because estimates of the AR parameters can be found by solving a set of linear equations, the so-called Yule-Walker equations. When the AR modeling assumption is valid, spectral estimators are obtained which are less biased and have a lower variability than conventional Fourier based spectral estimators [9]-[13]. Furthermore, the AR spectral estimator is supported by the well-known fact that the AR spectral estimator is equivalent to the linear prediction spectral estimator as well as the maximum entropy spectral estimator [1]-[4] and they are also equivalent to the maximum spectral flatness spectral estimator [5], [6] as the order of AR processes is known *a priori*. However, all power spectral estimators are based on second-order statistics (power spectra or correlation functions) and thus are phase blind whether the process is Gaussian or not. Obviously, if the signal phase is needed, the performance degradation to any correlation based signal processing methods is inevitable.

Recently, higher order statistics (HOS) based identification of nonminimum-phase linear time-invariant systems with only non-Gaussian output measurements has drawn extensive attention in the aforementioned areas of science and engineering. These statistics, known as cumulants, and their associated Fourier transforms, known as polyspectra, not only extract the amplitude information but also phase information, whereas cumulants are totally zero if the process of interest is Gaussian. In the real world, measurements are very often non-Gaussian random signals, such as voiced speech signals, binary sequences in digital communications, and seismograms, while noise is very often Gaussian. Hence, cumulant based signal processing methods are naturally immune from Gaussian noise whether the noise is white or colored with either known or unknown statistics. On the other hand, correlation based signal processing methods are neither phase-sensitive nor immune from Gaussian noise. These facts recently prompted rapid research in signal processing with HOS.

As in parametric spectral estimation, parametric polyspectral estimation is also based on the assumption that the given non-Gaussian linear process is an AR, a moving average (MA), or an autoregressive moving average (ARMA) process. Thus parametric polyspectral estimation and system identification with only output measurements are basically the same problem for estimating the unknown parameters from available finite data. Giannakis and Mendel [20], Giannakis [21], and Swami and Mendel [22], [25] estimate the unknown parameters basically by fitting a set of linear equations with sample cumulants in least squares sense. There are many other methods, such as closed-form solution based recursive methods [19], [20], bicepstral methods reported in Pan and Nikias [27], Lii and Rosenblatt's exhaustive search method [29], Tugnait's cumulant matching methods [31], [32] and Friedlander and Porat's optimization method [33]. References [17] and [18] provide the reader with a nice tutorial review of HOS in signal processing and system theory. Among the existing cumulant based AR parameter estimators, many of them, such as [20]-[25], [30], are developed based on fitting a set of linear equations associated with (7) (to be discussed later) by the least squares method without resort to the minimization of a cost function of the prediction error, which is the backbone of AR power spectral estimators. To the author's knowledge, this is because a well-defined prediction error filter based on HOS has never been reported in the open literature. Naturally, it is not yet known whether all the previously men-

Manuscript received September 13, 1991; revised July 11, 1992. The associate editor coordinating the review of this paper and approving it for publication was Dr. Steven Elgar. This work was supported by the National Science Council under Grant NSC 81-0404-E-007-010 and the Telecommunications Laboratories under Grant TL-NSC-81-5201.

The author is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China.

IEEE Log Number 9208201

1053-587X/93$03.00 © 1993 IEEE

tioned equivalencies among AR power spectral estimators imply the corresponding equivalencies among AR polyspectral estimators. This motivated the study presented in this paper.

In Section II, we briefly review the background of non-Gaussian AR processes for ease of later use. Then we present a new prediction error filter based on the HOS of stationary non-Gaussian linear processes in Section III. Sections IV and V discuss polyspectral estimation based on the measure of polyspectral flatness and that based on the maximum higher order entropy criterion, repectively. Finally, we provide a discussion and draw some conclusions.

II. Non-Gaussian AR Processes

Assume that $x(k)$ is the output of a casual autoregressive (AR) model of order p as follows:

$$x(k) = -\sum_{i=1}^{p} a(i)x(k-i) + u(k) \tag{1}$$

where $u(k)$ is a real, zero-mean, independent identically distributed (i.i.d.) non-Gaussian process. The transfer function $1/A(z)$, where

$$A(z) = \sum_{i=0}^{p} a(i)z^{-i} = 1 + \sum_{i=1}^{p} a(i)z^{-i} \tag{2}$$

and $a(0) = 1$, is assumed to be minimum phase (with all zeros inside the unit circle) for $x(k)$ to be stationary. The Mth-order cumulant $\text{cum}(x_1, x_2, \cdots, x_M)$ of random variables $x_1, x_2, \cdots, x_M$ is defined in [17], [18] as the coefficient of $(v_1 \cdot v_1 \cdots v_M)$ in the Taylor series expansion of the cumulant generating function

$$K(\boldsymbol{v}) = \ln E\{\exp(\boldsymbol{v}'\boldsymbol{x})\} \tag{3}$$

where $\boldsymbol{v} = (v_1, v_2, \cdots, v_M)'$ and $\boldsymbol{x} = (x_1, x_2, \cdots, x_M)'$. It is well known that the Mth-order cumulant function, denoted $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$, of the stationary $x(k)$ can be expressed by [17], [18]

$$\begin{aligned} C_{M,x}&(k_1, k_2, \cdots, k_{M-1}) \\ &= \text{cum}\,(x(k), x(k+k_1), x(k+k_{M-1})) \\ &= \gamma_M \sum_{k=0}^{\infty} h(k)h(k+k_1)\cdots h(k+k_{M-1}) \end{aligned} \tag{4}$$

where γ_M is the Mth-order cumulant of $u(k)$ and $h(k)$ is the impulse response of the AR system. Note that $h(0) = 1$ and $h(k) = 0$ for $k < 0$ since the AR system is casual. The Mth-order polyspectrum, denoted $S_{M,x}(f_1, f_2, \cdots, f_{M-1})$, of $x(k)$, which is the $(M-1)$-dimensional Fourier transform of $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$, is given by [17], [18].

$$\begin{aligned} S_{M,x}(f_1, f_2, \cdots, f_{M-1}) &= \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{M-1}=-\infty}^{\infty} C_{M,x}(k_1, k_2, \cdots, k_{M-1}) \exp\{-j2\pi[f_1k_1 + \cdots + f_{M-1}k_{M-1}]\} \\ &= \gamma_M \frac{1}{A(f_1)} \cdots \frac{1}{A(f_{M-1})} \cdot \frac{1}{A^*(f_1 + \cdots + f_{M-1})} \end{aligned} \tag{5}$$

where $A(f) = A(z = e^{j2\pi f})$. A well-known linear equation relating Mth-order cumulant function $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$ to AR coefficients can be obtained from (5) as follows.

Observe from (5) that

$$S^*_{M,x}(f_1, \cdots, f_{M-1})\, A(f_1 + \cdots + f_{M-1}) = \gamma_M \frac{1}{A^*(f_1)} \cdots \frac{1}{A^*(f_{M-1})}. \tag{6}$$

It can be easily shown, by taking inverse $(M-1)$-dimensional Fourier transform of both sides of (6), that

$$\begin{aligned} \sum_{i=0}^{p} a(i) C_{M,x}&(i-k_1, \cdots, i-k_{M-1}) \\ &= \gamma_M h(-k_1)h(-k_2)\cdots h(-k_{M-1}) \\ &= \begin{cases} \gamma_M, & k_1 = k_2 = \cdots = k_{M-1} = 0 \\ 0, & \text{for any } k_i > 0 \end{cases}. \end{aligned} \tag{7}$$

Let $R_{M,p}$ denote the finite domain of support associated with the Mth-order cumulant function of any non-Gaussian MA process of order p. For instance, for $M = 3$, the finite region $R_{3,p}$ is well known to be the point set

$$R_{3,p} = \{(k_1, k_2) \mid |k_1| \le p, |k_2| \le p, |k_1 - k_2| \le p\}$$

which is symmetric with respect to the origin. If $(k_1, k_2) \in R_{3,p}$ where $0 \le k_2 \le k_1 \le p$, then the associated points (k_2, k_1), $(-k_1, k_2 - k_1)$, $(k_2 - k_1, -k_1)$, $(k_1 - k_2, -k_2)$, and $(-k_2, k_1 - k_2)$ belong to $R_{3,p}$ and all the values of $C_{3,x}(k_1, k_2)$ at these points are the same due to the inherent symmetry properties of $C_{3,x}(k_1, k_2)$. Given $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$ for $(k_1, k_2, \cdots, k_{M-1}) \in R_{M,p}$, one can obtain $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$ for any other $(k_1, k_2, \cdots, k_{M-1}) \notin R_{M,p}$, from these known samples either by (7) or by the inherent symmetry properties of $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$. Remark that for $M = 2$, (7) reduces to the well-known Yule–Walker equations as follows:

$$\sum_{i=0}^{p} a(i)\, r_{xx}(i-k) = \begin{cases} \sigma^2, & k = 0 \\ 0, & k > 0 \end{cases} \tag{8}$$

where σ^2 is the variance of $u(k)$ and $r_{xx}(k)$ is the autocorrelation function of $x(k)$. Moreover, $R_{2,p}$ reduces to $\{-p, -p+1, \cdots, p\}$. It is well known [6], [10] that given $\{r_{xx}(k), k \in R_{2,p}\}$, $r_{xx}(k)$ for any other $k \notin R_{2,p}$ can be computed from these known samples either by (8) or by the symmetry property $r_{xx}(k) = r_{xx}(-k)$.

Surely, one can uniquely solve for $a(i)$ using a set of linearly independent equations associated with (7). How to select those "linearly independent" equations has been addressed by such as Giannakis [21], Swami and Mendel [22], and Tugnait [30]. Swami and Mendel, as well as Giannakis, have shown that $a(i)$ can be determined when $p + 1$ slices of Mth-order cumulants are used. To guarantee the identifiability of $a(i)$, many more than p linear equations associated with (7) using cumulant samples taken from $p + 1$ cumulant slices are suggested by them. Unfortunately, the cumulant samples needed by them are not fully taken from the finite region $R_{M,p}$ in spite of the fact that for $M = 2$, one can uniquely solve for $a(i)$ using the previous Yule–Walker equations which only need correlation samples within $R_{2,p}$. In other words, whether it is possible to uniquely determine $a(i)$ with only cumulant samples within $R_{M,p}$ is still an open question.

The AR polyspectral estimator provides the polyspectral estimate by substituting the estimates $\hat{A}(z)$ and $\hat{\gamma}_M$ obtained somehow from available finite data into (5). How to accurately and efficiently estimate $\hat{A}(z)$ and $\hat{\gamma}_M$ from finite data is the common goal of all AR polyspectral estimators. Next, let us present the new HOS based prediction error filter.

III. Prediction Error Filter Based on HOS

To reduce confusion, $x(k)$ will now denote the given real stationary non-Gaussian linear process, otherwise, we will clearly mention that $x(k)$ refers to (1). Let $\hat{A}(z)$ be the transfer function of the linear prediction error filter of order p with $\hat{a}(0) = 1$ and the input $x(k)$. The output, denoted $e(k)$, of the prediction error filter is then

$$e(k) = \sum_{i=0}^{p} \hat{a}(i)x(k-i) = x(k) + \sum_{i=0}^{p} \hat{a}(i)x(k-i) \tag{9}$$

and its Mth-order polyspectrum can be shown to be

$$S_{M,e}(f_1, f_2, \cdots, f_{M-1}) = \hat{A}(f_1) \cdots \hat{A}(f_{M-1}) \cdot \hat{A}^*(f_1 + \cdots + f_{M-1}) \cdot S_{M,x}(f_1, f_2, \cdots, f_{M-1}). \tag{10}$$

The optimum $\hat{A}(z)$ is the one which minimizes the sum of absolute squares of the Mth-order cumulant function of $e(k)$ as follows:

$$J_M(\hat{A}(z)) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{M-1}=-\infty}^{\infty} |C_{M,e}(k_1, k_2, \cdots, k_{M-1})|^2 \tag{11}$$

which, even for $M = 2$, seems totally different from the mean square error $E[e^2(k)]$ on which the conventional prediction error filter [6], [14]–[16] is based. Next, let us show that for $M = 2$ minimizing $J_M(\hat{A}(z))$ is indeed equivalent to minimizing $E[e^2(k)]$ when $x(k)$ is given by (1).

For $M = 2$, $J_M(\hat{A}(z))$ defined by (11) reduces to

$$J_2(\hat{A}(z)) = \sum_{k=-\infty}^{\infty} |r_{ee}(k)|^2 \tag{12}$$

where $r_{ee}(k)$ is the autocorrelation function of $e(k)$. It is sufficient that $J_2(\hat{A}(z))$ given by (12) is minimum if $|r_{ee}(k)|$ is minimum for all k. A set of sufficient conditions for $|r_{ee}(k)|$ to be minimum for all k is that $r_{ee}(0) = E[e^2(k)] > 0$ is minimum and $r_{ee}(k) = 0$ for $k \neq 0$ in the meantime. It is well known that the conventional pth-order prediction error filter, which minimizes the mean square error of $e(k)$, is a whitening filter ($r_{ee}(k) = 0$ for $k \neq 0$) with transfer function equal to $A(z)$ [6], [16]. In other words, minimizing $E[e^2(k)]$ is equivalent to minimizing $J_2(\hat{A}(z))$ given by (12) because the former requires that the previous sufficient conditions associated with the latter be true.

Notice, from (11), that $J_M(\hat{A}(z))$ is not a quadratic function but a highly nonlinear function of the coefficients of $\hat{A}(z)$ if $M > 2$. Therefore, the optimum $\hat{A}(z)$ cannot be obtained by solving a set of linear equations, as was done with (8) for the conventional prediction error filter. Some discussion about this is in Section VI.

It is also well known that the conventional prediction error filter based on the minimization of $E[e^2(k)]$ is minimum phase [14], [15]. The corresponding fact associated with the proposed prediction error filter based on the minimization of $J_M(\hat{A}(z))$ defined by (11) is described in the following.

Fact 1: The optimum $\hat{A}(z)$ associated with $J_M(\hat{A}(z))$ (see (11)) cannot have zeros outside the unit circle (minimum phase).

Proof: Assume that $\hat{A}(z)$ is not minimum phase with a zero z_i outside the unit circle (i.e., $|z_i| > 1$). We can then express $\hat{A}(z)$ as

$$\hat{A}(z) = (1 - z_i z^{-1})\tilde{A}(z) \tag{13}$$

where $\tilde{A}(z)$ is a $(p-1)$th-order polynomial of z^{-1}. Let

$$\hat{B}(z) = (1 - (1/z_i^*)z^{-1})\tilde{A}(z). \tag{14}$$

We need the following equality in the proof:

$$\begin{aligned} |1 - z_i \exp\{-j2\pi f\}| &= |z_i| \cdot \left|\frac{1}{z_i} - \exp\{-j2\pi f\}\right| \\ &= |z_i| \cdot |\exp\{j2\pi f\} - 1/z_i^*| \\ &= |z_i| \cdot |1 - (1/z_i^*)\exp\{-j2\pi f\}|. \end{aligned} \tag{15}$$

One can easily see from (13)–(15) that

$$|\hat{A}(f)| = |z_i| \cdot |\hat{B}(f)| > |\hat{B}(f)|. \tag{16}$$

Next, we infer, from (10), (11), and (16), that

$$\begin{aligned} J_M(\hat{A}(z)) &= \int_0^1 \cdots \int_0^1 |S_{M,e}(f_1, \cdots, f_{M-1})|^2 \, df_1 \, df_2 \cdots df_{M-1} \quad \text{(by Parseval's theorem)} \\ &= \int_0^1 \cdots \int_0^1 |S_{M,x}(f_1, \cdots, f_{M-1})|^2 \cdot |\hat{A}(f_1) \hat{A}(f_2) \cdots \hat{A}(f_{M-1}) \\ &\quad \cdot \hat{A}^*(f_1 + \cdots + f_{M-1})|^2 \, df_1 \, df_2 \cdots df_{M-1} \quad \text{(since (10))} \\ &> \int_0^1 \cdots \int_0^1 |S_{Mx}(f_1, \cdots, f_{M-1})|^2 \cdot |\hat{B}(f_1) \hat{B}(f_2) \cdots \hat{B}(f_{M-1}) \\ &\quad \cdot \hat{B}^*(f_1 + \cdots + f_{M-1})|^2 \, df_1 \, df_2 \cdots df_{M-1} \quad \text{(since (16))} \\ &= J_M(\hat{B}(z)) \end{aligned} \tag{17}$$

which implies that $J_M(\hat{A}(z))$ can never be minimum unless $\hat{A}(z)$ is minimum phase without zeros outside the unit circle. Q.E.D.

Next, let us discuss when the optimum $\hat{A}(z)$ cannot have zeros on the unit circle. Assume that $z_i = \rho e^{j2\pi\eta}$ in $\hat{A}(z)$ given by (13). Then

$$|\hat{A}(f)|^2 = g(f) |\tilde{A}(f)|^2 \tag{18}$$

where

$$g(f) = 1 - 2\rho \cos(2\pi(f - \eta)) + \rho^2. \tag{19}$$

By substituting (18) into the second line of (17), $J_M(\hat{A}(z))$ can be expressed as

$$J_m(\hat{A}(z)) = \int_0^1 \cdots \int_0^1 G_1(f_1, \cdots, f_{M-1}) \cdot G_2(f_1, \cdots, f_{M-1}) \, df_1 \, df_2 \cdots df_{M-1} \tag{20}$$

where

$$\begin{aligned} G_1(f_1, \cdots, f_{M-1}) &= |S_{M,x}(f_1, \cdots, f_{M-1})|^2 \cdot |\tilde{A}(f_1) \tilde{A}(f_2) \\ &\quad \cdots \tilde{A}(f_{M-1}) \cdot \tilde{A}^*(f_1 + \cdots + f_{M-1})|^2 \end{aligned} \tag{21}$$

and

$$\begin{aligned} G_2(f_1, \cdots, f_{M-1}) &= g(f_1) g(f_2) \cdots g(f_{M-1}) \\ &\quad \cdot g(f_1 + \cdots + f_{M-1}). \end{aligned} \tag{22}$$

It is easy to show that $dG_2/d\rho > 0$ for $\rho > 1$, $dG_2/d\rho \geq 0$ for $\rho = 1$ and the case of $dG_2/d\rho = 0$ for $\rho = 1$ happens only when $f_i = \eta$ or $f_1 + f_2 + \cdots + f_{M-1} = \eta$. This implies that $dJ_M/d\rho > 0$ for $\rho > 1$ which is consistent with Fact 1, and that $dJ_M/d\rho > 0$ for $\rho = 1$ if $G_1(f_1, \cdots, f_{M-1}) > 0$ is a continuous function of $f_1, \cdots, f_{M-1}$ which happens when $S_{M,x}(f_1, \cdots, f_{M-1})$ is a continuous function of $f_1, \cdots, f_{M-1}$. Therefore, the J_M for $\hat{A}(z)$ having a zero on the unit circle is never a local minimum when $S_{M,x}(f_1, \cdots, f_{M-1})$ is a continuous function of $f_1, \cdots, f_{M-1}$. In other words, we have the following conclusion

(C1) The optimum $\hat{A}(z)$ associated with $J_M(\hat{A}(z))$ cannot have zeros on the unit circle when $x(n)$ has a continuous Mth-order polyspectrum.

IV. Maximum Flatness of Polyspectra

In this section, we concentrate on the measure of polyspectral flatness. Let $\xi_e(L)$ denote the flatness of Mth-order polyspectrum of $e(k)$ defined as

$$\xi_e(L) = \frac{\exp\left\{ \int_0^1 \cdots \int_0^1 \ln \left[|S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^L \right] df_1 \cdots df_{M-1} \right\}}{\int_0^1 \cdots \int_0^1 |S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^L \, df_1 \cdots df_{M-1}} \tag{23}$$

where L is a positive integer. Note that $\xi_e(L)$ is the geometric mean of $|S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^L$ divided by the arithmetic mean and therefore can be shown to satisfy $0 \leq \xi_e(L) \leq 1$. Note that $\xi_e(L) = 1$ if $|S_{M,e}(f_1, f_2, \cdots, f_{M-1})|$ equals a constant for all $(f_1, f_2, \cdots, f_{M-1})$. On the other hand, $\xi_e(L) \approx 0$ if $|S_{M,e}(f_1, f_2, \cdots, f_{M-1})|$ is very peaky. Note that for $L = 1$ and $M = 2$, $\xi_e(L)$ is exactly the same as that associated with the well-known Gray and Markel's spectral flatness measure [5], [6]. Now let us show the following fact associated with the maximization of $\xi_e(L)$.

Fact 2: Let $Q(L)$ denote the set of all optimum $\hat{A}(z)$'s associated with maximum $\xi_e(L)$. a) If an $\hat{A}(z) \in Q(L)$ has p zeros at $z = z_1, \cdots, z = z_p$, then all $\hat{D}(z)$'s $\in Q(L)$ where $\hat{D}(z) = (1 - \alpha_1 z^{-1}) \cdots (1 - \alpha_p z^{-1})$ in which $\alpha_k = z_k$ or $1/z_k^*$, $k = 1, 2, \cdots, p$. b) Maximizing $\xi_e(L)$ is equivalent to minimizing the following objective function

$$\tilde{J}_L(\hat{A}(z)) = \int_0^1 \cdots \int_0^1 |S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^L \, df_1 \cdots df_{M-1}. \tag{24}$$

Proof: Let us prove statement a) and then statement b). Assume that an $\hat{A}(z) \in Q(L)$ is expressed by (13). Moreover, assume that $x(k)$ is also the input applied to the $\hat{B}(z)$ as given by (14) and $\mu(k)$ is the corresponding output $\hat{B}(z)$. It can be easily shown by (10) and the equality in (16) that

$$|S_{M,e}(f_1, f_2, \cdots, f_{M-1})| = F \cdot |S_{M,\mu}(f_1, f_2, \cdots, f_{M-1})| \tag{25}$$

where $F = |z_i|^M$ is a constant. Therefore, $\xi_e(L) = \xi_\mu(L)$ by (23) and (25) which also indicates polyspectral flatness is the same for any two scale-factor related polyspectra. Therefore, $\hat{B}(z)$ is also an optimum solution or $\hat{B}(z) \in Q(L)$. Thus statement a) is true.

Next, let us prove statement b). The following result is needed in the proof. If $\hat{A}(z)$ is minimum phase [6], [16] with all zeros inside the unit circle, then

$$\int_0^1 \ln |\hat{A}(f)|^2 \, df = 0. \tag{26}$$

The equality (26) also holds when $\hat{A}(z)$ has zeros on the unit circle. The proof is as follows. Again, assume that $\hat{A}(z)$ is expressed by (13) where $z_i = \rho e^{j2\pi\eta}$ and all the zeros of $\hat{A}(z)$ are inside the unit circle. Then substituting (18) into the integral given by (26) for $\rho = 1$ gives rise to

$$\begin{aligned}
\int_0^1 \ln |\hat{A}(f)|^2 \, df &= \int_0^1 \ln g(f) df + \int_0^1 \ln |\tilde{A}(f)|^2 \, df \\
&= \int_0^1 \ln g(f) \, df \quad \text{(since all zeros of } \hat{A}(z) \text{ are inside the unit circle)} \\
&= \int_{-1/2+\eta}^{1/2+\eta} \ln 2[1 - \cos(2\pi(f - \eta))] \, df \\
&\quad \text{(since } g(f) \text{ has a period of 1)} \\
&= \ln 2 + \frac{1}{\pi} \int_0^\pi \ln(1 - \cos y) \, dy \\
&= \ln 2 + \frac{1}{\pi}(-\pi \ln 2) = 0
\end{aligned}$$

where we have used a definite integral formula in the last line of the proof. Next, we have

$$\begin{aligned}
&\int_0^1 \cdots \int_0^1 \ln |S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^L \, df_1 \cdots df_{M-1} \\
&= \frac{L}{2} \int_0^1 \cdots \int_0^1 \ln |S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^2 \, df_1 \cdots df_{M-1} \\
&= \frac{L}{2} \int_0^1 \cdots \int_0^1 \ln |S_{M,x}(f_1, f_2, \cdots, f_{M-1})|^2 \, df_1 \cdots df_{M-1} + \frac{L}{2} \int_0^1 \cdots \int_0^1 \{\ln |\hat{A}(f_1)|^2 \\
&\quad + \cdots + \ln |\hat{A}(f_{M-1})|^2 + \ln |\hat{A}^*(f_1 + \cdots + f_{M-1})|^2\} \, df_1 \cdots df_{M-1} \quad \text{(since (10))} \\
&= \int_0^1 \cdots \int_0^1 \ln |S_{M,x}(f_1, f_2, \cdots, f_{M-1})|^L \, df_1 \cdots df_{M-1} \quad \text{(since (26))}
\end{aligned} \tag{27}$$

which is not a function of $\hat{A}(z)$. Substituting (27) into (23) provides

$$\xi_e(L) = \frac{\exp\left\{\int_0^1 \cdots \int_0^1 \ln \{|S_{M,x}(f_1, f_2, \cdots, f_{M-1})|^L\} \, df_1 \cdots df_{M-1}\right\}}{\int_0^1 \cdots \int_0^1 |S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^L \, df_1 \cdots df_{M-1}}. \tag{28}$$

Thus, maximizing $\xi_e(L)$ is equivalent to minimizing the denominator of $\xi_e(L)$ which is identical with the objective function $\tilde{J}_L(\hat{A}(z))$. The assumption of minimum-phase $\hat{A}(z)$ in the proof is not restrictive by statement a).

Q.E.D.

Remark that $\tilde{J}_L(\hat{A}(z))$ given by (24) for $L = 1$ and $M = 2$ reduces to $\tilde{J}_L(\hat{A}(z)) = r_{ee}(0) = E[e^2(k)]$ (by Parseval's theorem) with which the conventional prediction error filter is associated. Furthermore, one may ask if the optimum $\hat{A}(z)$ associated with the objective function $\tilde{J}_L(\hat{A}(z))$ is identical with the proposed HOS based prediction error filter associated with (11). Letting $L = 2$ in $\tilde{J}_L(\hat{A}(z))$ yields

$$\begin{aligned}
\tilde{J}_2(\hat{A}(z)) &= \int_0^1 \cdots \int_0^1 |S_{M,e}(f_1, f_2, \cdots, f_{M-1})|^2 \cdot df_1 \cdots df_{M-1} = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{M-1}=-\infty}^{\infty} \cdot |C_{M,e}(k_1, k_2, \cdots, k_{M-1})|^2 \\
&\quad \text{(by Parseval's theorem)} \\
&= J_M(\hat{A}(z)).
\end{aligned} \tag{29}$$

Hence, we have shown Fact 3 described as follows:

Fact 3: The $\hat{A}(z) \in Q(L)$ (defined in Fact 2) with minimum phase for $L = 2$ is identical with the HOS based prediction error filter (also minimum phase) associated with $J_M(\hat{A}(z))$. Hence, the maximum polyspectral flatness polyspectrum associated with the minimum-phase $\hat{A}(z)$ for $L = 2$ is the same as the linear prediction polyspectrum.

By Fact 3 and (C1) we also have the following conclusion.

(C2) The $\hat{A}(z) \in Q(L)$ (defined in Fact 2) with minimum phase for $L = 2$ cannot have zeros on the unit circle when $x(k)$ has a continuous polyspectrum.

Next, let us show the following fact.

Fact 4: The $\hat{A}(z) \in Q(L)$ (defined in Fact 2) with minimum phase is identical with $A(z)$ when $x(k)$ is given by (1).

Proof: Let us solve for the optimum $\hat{A}(z)$ with minimum phase by letting $\xi_e(L) = 1$ (maximum value of $\xi_e(L)$) which requires

$$\begin{aligned}
|S_{M,e}(f_1, f_2, \cdots, f_{M-1})| &= |S_{M,x}(f_1, f_2, \cdots, f_{M-1})| \cdot |\hat{A}(f_1)| \cdot |\hat{A}(f_2)| \\
&\quad \cdots |\hat{A}(f_{M-1})| \cdot \hat{A}(f_1 + f_2 + \cdots + f_{M-1})| \\
&\quad \text{(since (10))} \\
&= |\gamma_M| \cdot \left|\frac{\hat{A}(f_1)}{A(f_1)}\right| \cdot \left|\frac{\hat{A}(f_2)}{A(f_2)}\right| \cdots \left|\frac{\hat{A}(f_{M-1})}{A(f_{M-1})}\right| \\
&\quad \cdot \left|\frac{\hat{A}(f_1 + f_1 + \cdots + f_{M-1})}{A(f_1 + f_2 + \cdots + f_{M-1})}\right| \quad \text{(since (5))} \\
&= G, \quad \text{for all } (f_1, f_2, \cdots, f_{M-1}) \qquad (30)
\end{aligned}$$

where G is a constant. Note that both $A(z)$ and $\hat{A}(z)$ are minimum phase by assumption. Therefore, the rational function $\hat{A}(z)/A(z)$, which is also minimum phase, never forms an all-pass filter which is nonminimum phase [34]. Thus, we conclude, from (30) that

$$\hat{A}(z)/A(z) = G' \qquad (31)$$

where G' is also a constant. Since $\hat{a}(0) = a(0) = 1$, the constant G' can only take the value of unity. In other words, $\hat{A}(z) = A(z)$. The proof is thus completed.

Q.E.D.

Furthermore, notice from (30) and (31) that

$$S_{M,e}(f_1, \cdots, f_{M-1}) = \gamma_M$$

or

$$C_{M,e}(k_1, \cdots, k_{M-1}) = \gamma_M \, \delta(k_1) \cdots \delta(k_{M-1})$$

where $\delta(k)$ is the discrete delta function ($\delta(k) = 1$ for $k = 0$, $\delta(k) = 0$ for $k \neq 0$). Therefore, both the linear prediction polyspectral estimator and the maximum polyspectral flatness polyspectral estimator perform as an "Mth-order whitening filter," which suggests that $\hat{\gamma}_M$ be estimated as the sample cumulant $\hat{C}_{M,e}(k_1 = 0, \cdots, k_{M-1} = 0)$ when the number of data is finite. Remark that the conventional prediction error filter is known as a second-order whitening filter whose output is an uncorrelated process but not an independent process except for the Gaussian case. Similarly, the non-Gaussian output of an Mth-order ($M \geq 3$) whitening filter can be viewed as an "Mth-order uncorrelated" process which is not an independent process either except that $x(k)$ is given by (1) where $u(k)$ is i.i.d. Next, let us present the criterion of maximum higher order entropy.

V. Maximum Higher Order Entropy

We define the Mth-order entropy, denoted $\Gamma(\hat{S}_{M,x})$, of the polyspectrum $\hat{S}_{M,x}(f_1, \cdots, f_{M-1})$ of $x(k)$, as

$$\Gamma(\hat{S}_{M,x}) = \frac{1}{2} \int_0^1 \cdots \int_0^1 \ln |\hat{S}_{M,x}(f_1, \cdots, f_{M-1})|^2 \, df_1 \cdots df_{M-1}. \qquad (32)$$

The maximum Mth-order entropy polyspectrum of $x(k)$ is the one which maximizes $\Gamma(\hat{S}_{M,x})$ given (32) subject to the constraint

$$\hat{C}_{M,x}(k_1, \cdots, k_{M-1}) = C_{M,x}(k_1, \cdots, k_{M-1}) \quad \forall\, (k_1, \cdots, k_{M-1}) \in R_{M,p} \qquad (33)$$

where $R_{M,p}$, which was defined in Section II, is the domain of support associated with the Mth-order cumulant function of any non-Gaussian MA process of order p, and $C_{M,x}(k_1, \cdots, k_{M-1})$ for $(k_1, \cdots, k_{M-1}) \in R_{M,p}$ are the known samples of cumulant function $\hat{C}_{M,x}(k_1, \cdots, k_{M-1})$ associated with $\hat{S}_{M,x}(f_1, \cdots, f_{M-1})$. In other words, the maximum Mth-order entropy polyspectral estimator assumes that the Mth-order cumulant function for $(k_1, \cdots, k_{M-1}) \in R_{M,p}$ is known exactly and attempts to extrapolate the Mth-order cumulant function for $(k_1, \cdots, k_{M-1}) \notin R_{M,p}$. Thus, the maximum Mth-order entropy polyspectral estimator bypasses the problems that arise from the use of window functions, a feature that is common to all Fourier-based methods of polyspectrum analysis.

It is worthwhile to remark that unlike the "entropy" [3] which indicates the randomness of a time series, we do not have any definite physical interpretation for the "Mth-order entropy" defined by (32). Moreover, for $M = 2$, $\Gamma(\hat{S}_{M,x})$ reduces to

$$\Gamma(\hat{S}_{xx}(f)) = \int_0^1 \ln \hat{S}_{xx}(f)\, df \qquad (34)$$

which is proportional to the entropy rate associated with the well-known Burg's maximum entropy spectral estimator [1]–[4], where $\hat{S}_{xx}(f)$ denotes a power spectrum estimate of $x(k)$. We next show the following fact associated with $\Gamma(\hat{S}_{M,x})$.

Fact 5: The maximum Mth-order entropy polyspectrum $\hat{S}_{M,x}(f_1, \cdots, f_{M-1})$ can be expressed as (40) below.

Proof: Let

$$\mathcal{L} = \Gamma(\hat{S}_{M,x}) - \sum_{(k_1, \cdots, k_{M-1}) \in R_{M,p}} \cdots \sum \lambda(k_1, \cdots, k_{M-1}) \cdot \{\hat{C}_{M,x}(k_1, \cdots, k_{M-1}) - C_{M,x}(k_1, \cdots, k_{M-1})\} \quad (35)$$

where $\lambda(k_1, \cdots, k_{M-1})$ is the Lagrange multiplier. Note that $\lambda(k_1, \cdots, k_{M-1})$ is real since $\hat{C}_{M,x}(k_1, \cdots, k_{M-1})$ is real. Taking partial derivative of $\mathcal{L}$ given by (35) with respect to $\hat{C}_{M,x}(k_1, \cdots, k_{M-1})$ and then setting the result to zero yields

$$\frac{1}{2}\int_{\frac{1}{2}}^{1} \cdots \int_{\frac{1}{2}}^{1} \left\{ \frac{e^{-j2\pi(f_1k_1 + \cdots + f_{M-1}k_{M-1})}}{\hat{S}_{M,x}(f_1, \cdots, f_{M-1})} + \frac{e^{j2\pi(f_1k_1 + \cdots + f_{M-1}k_{M-1})}}{\hat{S}^*_{M,x}(f_1, \cdots, f_{M-1})} \right\} df_1 \cdots df_{M-1} = \begin{cases} \lambda(k_1, \cdots, k_{M-1}), & (k_1, \cdots, k_{M-1}) \in R_{M,p} \\ 0 & \text{otherwise.} \end{cases} \quad (36)$$

Let us define $\Lambda(k_1, k_2, \cdots, k_{M-1})$ to be

$$\Lambda(k_1, k_2, \cdots, k_{M-1}) = \begin{cases} \lambda(-k_1, \cdots, -k_{M-1}), & (k_1, \cdots, k_{M-1}) \in R_{M,p} \\ 0 & \text{otherwise.} \end{cases} \quad (37)$$

Note that the domain of support associated with $\Lambda(k_1, k_2, \cdots, k_{M-1})$ is the same as that (i.e. $R_{M,p}$) associated with $\lambda(k_1, k_2, \cdots, k_{M-1})$ since $R_{M,p}$ is symmetric with respect to the origin. It can be easily shown from (36) and (37) that $1/\hat{S}_{M,x}(f_1, \cdots, f_{M-1})$ and $\Lambda(k_1, k_2, \cdots, k_{M-1})$ form an $(M-1)$-dimensional Fourier transform pair, i.e.,

$$\frac{1}{\hat{S}_{M,x}(f_1, \cdots, f_{M-1})} = \sum_{(k_1, \cdots, k_{M-1}) \in R_{M,p}} \cdots \sum \Lambda(k_1, \cdots, k_{M-1}) \cdot \exp\{-j2\pi(f_1k_1 + \cdots + f_{M-1}k_{M-1})\} \quad (38)$$

where $\Lambda(k_1, k_2, \cdots, k_{M-1})$ also satisfies all the symmetry properties of the Mth-order cumulant function since $\hat{S}_{M,x}(f_1, \cdots, f_{M-1})$ is a polyspectrum of a non-Gaussian linear process by assumption. Moreover, one can see from (38), that $1/\hat{S}_{M,x}(f_1, \cdots, f_{M-1})$ can be viewed as the Mth-order polyspectrum of a non-Gaussian MA process of order p and therefore can be factored as the product

$$\frac{1}{\hat{S}_{M,x}(f_1, \cdots, f_{M-1})} = \sum_{(k_1, \cdots, k_{M-1}) \in R_{M,p}} \cdots \sum \Lambda(k_1, \cdots, k_{M-1}) \cdot \exp\{-j2\pi(f_1k_1 + \cdots + f_{M-1}k_{M-1})\} \quad (39)$$

where $\hat{A}(f) = \hat{A}(z = \exp\{j2\pi f\})$ and $\hat{A}(z)$ is a pth-order polynomial of z^{-1}, or

$$S_{M,x}(f_1, \cdots, f_{M-1}) = \hat{\gamma}_M \frac{1}{\hat{A}(f_1)} \cdots \frac{1}{\hat{A}(f_{M-1})} \cdot \frac{1}{\hat{A}^*(f_1 + \cdots + f_{M-1})}. \quad (40)$$

Q.E.D.

By substituting (40) back into (32), the Mth-order entropy $\Gamma(\hat{S}_{M,x})$ can be no further expressed as

$$\Gamma(\hat{S}_{M,x}) = \ln|\hat{\gamma}_M| - \tfrac{1}{2}\int_0^1 \cdots \int_0^1 \ln|\hat{A}(f_1)\hat{A}(f_2) \cdots \hat{A}(f_{M-1}) \cdot \hat{A}^*(f_1 + \cdots + f_{M-1})|^2 \, df_1 \cdots df_{M-1}. \quad (41)$$

Let $P(R_{M,p})$ be the set of all possible $(\hat{\gamma}_M, \hat{A}(z))$'s that satisfy (33). By Fact 5, solving for the maximum Mth-order entropy polyspectrum is equivalent to finding $P(R_{M,p})$ and then substituting $(\hat{\gamma}_M, \hat{A}(z))$ of $P(R_{M,p})$ into (41) to find the optimum $(\hat{\gamma}_M, \hat{A}(z))$ as well as the maximum $\Gamma(\hat{S}_{M,x})$. Therefore, if $P(R_{M,p})$ contains only a single $(\hat{\gamma}_M, \hat{A}(z))$, it must be the optimum solution no matter whether the optimum $\hat{A}(z)$ is minimum phase or not, although when $M = 2$, the optimum $\hat{A}(z)$ associated with Burg's maximum entropy spectrum is minimum phase. Obviously, the constraint given by (33) is satisfied for $\hat{A}(z) = A(z)$ when $x(k)$ is given by (1). Hence the following fact is true.

Fact 6: $\hat{S}_{M,x}(f_1, \cdots, f_{M-1}) = S_{M,x}(f_1, \cdots, f_{M-1})$ (see (5)) is a maximum Mth-order entropy polyspectrum when $x(k)$ is given by (1) where $A(z)$ is minimum phase. Therefore, if the solution $\hat{A}(z) = A(z)$ is unique, the optimum

$$\hat{S}_{M,x}(f_1, \cdots, f_{M-1}) = S_{M,x}(f_1, \cdots, f_{M-1})$$

is unique. However, as mentioned in Section II, it is still unknown whether the solution $\hat{A}(z) = A(z)$ under the constraint given by (33) is unique although this is true for the case of $M = 2$. Furthermore, Facts 3 through 6 imply the following fact.

Fact 7: The linear prediction polyspectral estimator and the maximum polyspectral flatness polyspectral estimator associated with the minimum phase $\hat{A}(z) \in Q(L)$ for $L = 2$ are also a maximum Mth-order entropy polyspectral estimator when $x(k)$ is given by (1) where $A(z)$ is minimum phase.

As a final remark, the sampling points associated with known cumulant samples are not necessarily limited to the finite region $R_{M,p}$ as in (33). The resultant maximum Mth-order entropy polyspectral estimator can also extrapolate or interpolate the cumulant function for the other points. However, solving for the maximum Mth-order entropy polyspectrum becomes a quite difficult nonlinear optimization problem.

VI. Discussion and Conclusions

In the paper, we first presented a new prediction error filter based on HOS (see (11)) of a stationary non-Gaussian linear process $x(k)$. Then we defined the measure of maximum flatness of polyspectra (see (23)) and the criterion of maximum Mth-order entropy (see (32)). We have shown that when $x(k)$ is given by (1) and the order p is known *a priori*, the proposed HOS based prediction error criterion, maximum flatness measure of polyspectrum, and maximum Mth-order entropy criterion are associated with the same AR parameters which are identical with the true AR parameters. These results are summarized in Facts 1 through 7 in the previous sections. Besides, Fact 8 is to be given later in this section. We also showed that these results for which the order M of cumulant function is set to two reduce to the corresponding results associated with second-order statistics (power spectra) but it is still an open question whether or not the solution of maximum Mth-order entropy polyspectrum is unique. However, when the order p of the AR process of interest is not known in advance, the proposed HOS based prediction error filter $\hat{A}(z)$ is then no longer identical with the true $A(z)$ when the order p of $\hat{A}(z)$ is less than p, and thus the prediction error is not Mth-order white any more for this case. On the other hand, if $\hat{p} \geq p$, then $\hat{a}(i) = a(i)$ for $1 \leq i \leq p$ and $\hat{a}(i) = 0$ for $i > p$.

We would like to mention that Facts 1–3 and 5 are valid even when $x(k)$ is a noncausal AR process where $A(z)$ is nonminimum phase since the previous proofs for these four facts only require $x(k)$ to be a real stationary non-Gaussian linear process, and that Fact 5 can also be proven for the case of complex $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$ although the previous maximum Mth-order entropy polyspectrum was defined for real $C_{M,x}(k_1, k_2, \cdots, k_{M-1})$. A remark associated with Fact 7 is as follows. It is not yet known if the linear prediction polyspectral estimator and the maximum polyspectral flatness polyspectral estimator are still a maximum higher-order entropy polyspectral estimator or not when $x(k)$ is not an AR process.

Next, let us discuss the phase sensitivity of the previous linear prediction polyspectral flatness polyspectral estimator, and maximum Mth-order entropy polyspectral estimator. Assume that $y(k)$ is the output of an all-pass filter with the input being $x(k)$, then an identical maximum polyspectral flatness polyspectrum will be obtained using either $x(k)$ or $y(k)$ (because the all-pass factor in $|S_{M,e}(f_1, \cdots, f_{M-1})|$ disappears, thereby resulting in the same $\xi_e(L)$ (see (23)). By Fact 3, the linear prediction polyspectrum associated with $y(k)$ is also the same as that associated with $x(k)$. In other words, these two polyspectral estimators try to flatten polyspectra of $x(k)$ only in amplitude. Hence the following fact is true.

Fact 8: Both the linear prediction polyspectral estimator and the maximum polyspectral flatness polyspectral estimator are all-pass factor blind.

Corresponding to Fact 8 for $M = 2$ is the well-known fact that all power spectral estimators are all-pass factor blind. On the other hand, the maximum Mth-order entropy polyspectral estimator can be phase-sensitive due to the reasons as follows. Let $P_x(R_{M,p})$ associated with $\Gamma(\hat{S}_{M,x})$ and $P_y(R_{M,p})$ associated with $\Gamma(\hat{S}_{M,y})$ denote the solution set $P(R_{M,p})$ of $(\hat{\gamma}_M, \hat{A}(z))$'s that satisfy (33) (defined in Section V). Because cumulant function is phase sensitive, one can easily infer that

$$C_{M,x}(k_1, k_2, \cdots, k_{M-1}) \neq C_{M,y}(k_1, k_2, \cdots, k_{M-1})$$

for $(k_1, k_2, \cdots, k_{M-1}) \in R_{M,p}$. Therefore, $P_x(R_{M,p})$ is totally different from $P_y(R_{M,p})$ which implies that the maximum Mth-order entropy polyspectral estimator is phase sensitive.

Notice from (11) that solving for the proposed prediction error coefficients by minimizing (11) becomes a nonlinear optimization problem because the closed form solution does not seem to exist. Let us check an internal consistency that $\hat{A}(z) = A(z)$ should be a local minimum of $J_M(\hat{A}(z))$ when $x(k)$ is given by (1) since the optimum $\hat{A}(z)$ has been shown to be identical with $A(z)$ in the previous sections.

For simplicity, let $M = 3$. Note that $e(k) = u(k)$ as $\hat{A}(z) = A(z)$. Thus

$$C_{3,e}(k_1, k_2) = C_{3,u}(k_1, k_2) = \gamma_3 \cdot \delta(k_1) \cdot \delta(k_2). \quad (42)$$

Moreover, it can be easily seen from (1) that as $\hat{A}(z) = A(z)$,

$$\begin{aligned} E[e(k)\,x(k-k_1)\,x(k-k_2)] &= E[u(k)\,x(k-k_1)\,x(k-k_2)] \\ &= 0, \quad \text{if } k_1 > 0 \text{ or } k_2 > 0 \end{aligned} \quad (43)$$

and

$$E[e^2(k)\,x(k-k_1)] = E[u^2(k)\,x(k-k_1)] = \gamma_3 h(-k_1) = \begin{cases} \gamma_3, & k_1 = 0 \\ 0, & \text{if } k_1 > 0. \end{cases} \quad (44)$$

Let $\boldsymbol{x}(k) = (x(k), x(k-1), \cdots, x(k-p+1))'$. Taking partial derivative of $J_3(\hat{A}(z))$ (see (11)) with respect to $\hat{\boldsymbol{a}} = (\hat{a}(1), \hat{a}(2), \cdots, \hat{a}(p))'$ yields

$$\begin{aligned} \left.\frac{\partial J_3}{\partial \hat{\boldsymbol{a}}}\right|_{\hat{A}(z)=A(z)} &= 2 \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} C_{3,e}(k_1, k_2) \cdot \left.\frac{\partial C_{3,e}(k_1, k_2)}{\partial \hat{\boldsymbol{a}}}\right|_{\hat{A}(z)=A(z)} \\ &= 2\gamma_3 \left.\frac{\partial C_{3,e}(0, 0)}{\partial \hat{\boldsymbol{a}}}\right|_{\hat{A}(z)=A(z)} \quad \text{(since (42))} \\ &= 6\gamma_3 E[e^2(k)\boldsymbol{x}(k-1)]\Big|_{\hat{A}(z)=A(z)} \\ &= 0 \quad \text{(since (44))} \end{aligned} \quad (45)$$

which implies $\hat{A}(z) = A(z)$ is either a local minimum or a local maximum of $J_3(\hat{A}(z))$. Furthermore, we obtain from (45)

$$\left.\frac{\partial^2 J_3}{\partial \hat{a}^2}\right|_{\hat{A}(z)=A(z)} = 2\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\left\{\frac{\partial C_{3,e}(k_1,k_2)}{\partial \hat{a}}\right\}\left\{\frac{\partial C_{3,e}(k_1,k_2)}{\partial \hat{a}}\right\}'\Bigg|_{\hat{A}(z)=A(z)}$$

$$+ 2\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} C_{3,e}(k_1,k_2)\left\{\frac{\partial^2 C_{3,e}(k_1,k_2)}{\partial \hat{a}^2}\right\}\Bigg|_{\hat{A}(z)=A(z)}$$

$$= 2\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\left\{\frac{\partial C_{3,e}(k_1,k_2)}{\partial \hat{a}}\right\}\left\{\frac{\partial C_{3,e}(k_1,k_2)}{\partial \hat{a}}\right\}'\Bigg|_{\hat{A}(z)=A(z)}$$

$$+ 2\gamma_3\{E[e(k)\,\boldsymbol{x}(k-1)\,\boldsymbol{x}'(k-1)]\}\Big|_{\hat{A}(z)=A(z)} \qquad \text{(since (42))}$$

$$= 2\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\left\{\frac{\partial C_{3,e}(k_1,k_2)}{\partial \hat{a}}\right\}\left\{\frac{\partial C_{3,e}(k_1,k_2)}{\partial \hat{a}}\right\}'\Bigg|_{\hat{A}(z)=A(z)} \qquad \text{(since (43))} \tag{46}$$

which is obviously nonnegative definite. Furthermore, (46) implies that

$$\left.\frac{\partial^2 J_3}{\partial \hat{a}^2}\right|_{\hat{A}(z)=A(z)} \geq 2\sum_{k_2=1}^{p}\left\{\frac{\partial C_{3,e}(0,k_2)}{\partial \hat{a}}\right\}\left\{\frac{\partial C_{3,e}(0,k_2)}{\partial \hat{a}}\right\}'\Bigg|_{\hat{A}(z)=A(z)}. \tag{47}$$

Moreover,

$$\left.\frac{\partial C_{3,e}(0,k_2)}{\partial \hat{a}}\right|_{\hat{A}(z)=A(z)} = 2E[e(k)\,e(k+k_2)\,\boldsymbol{x}(k-1)] + E[e^2(k)\,\boldsymbol{x}(k+k_2-1)]\Big|_{\hat{A}(z)=A(z)}$$

$$= \gamma_3 \boldsymbol{v}_{k_2}, \quad \text{for } 1 \leq k_2 \leq p \quad \text{(since (43) and (44))} \tag{48}$$

where

$$\boldsymbol{v}_{k_2} = (h(k_2-1), \cdots, h(0)=1, \cdots, 0)' \tag{49}$$

whose k_2th component is equal to one and the last $(p - k_2)$ components are equal to zero. We conclude from (47) and (48) that

$$\left.\frac{\partial^2 J_3}{\partial \hat{a}^2}\right|_{\hat{A}(z)=A(z)} \geq 2\gamma_3^2 \sum_{k_2=1}^{p} \boldsymbol{v}_{k_2}\boldsymbol{v}'_{k_2} \tag{50}$$

whose right-hand side is obviously positive definite since $[\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_p]$ forms a basis of the p-dimensional Euclidean space. In other words, $\hat{A}(z) = A(z)$ is indeed a local minimum of $J_3(\hat{A}(z))$.

Although many existing cumulant based AR parameter estimators are based on fitting (7) with sample cumulants computed from finite data, currently, we are investigating the solution of AR coefficients based on (11) instead of (7). The results will be reported in a separate paper. On the other hand, many well-known order determination methods are based on the conventional minimum mean-square prediction error such as the final prediction error (FPE) [35], Akaike information criterion (AIC) [36], and the criterion autoregressive transfer function (CAT) [37]. We are also investigating some new order determination methods based on the proposed HOS based prediction error. We believe that the results presented in this paper provide a theoretical background on the polyspectral estimation and modeling of non-Gaussian AR processes.

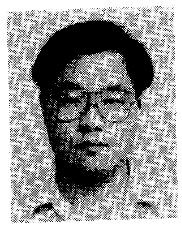
Acknowledgment

The research described in this paper was performed at the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China.

References

[1] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford University, May 1975.
[2] S. S. Haykin and S. Kesler, "Prediction error filtering and maximum-entropy spectral estimation," in *Nonlinear Methods of Spectral Analysis*, 2nd ed., S. Haykin, Ed. New York: Springer-Verlag, 1983, ch. 2.
[3] E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proc. IEEE*, vol. 70, pp. 939–952, 1982.
[4] A. Van De Bos, "Alternative interpretation of maximum entropy spectral analysis," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 493–494, July 1971.
[5] A. H. Gray, Jr., and J. D. Markel, "A spectral flatness measure for studying the autocorrelation method of linear prediction of speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 207–217, June 1974.
[6] Steven M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[7] R. H. Shumway, *Applied Statistical Time Series Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[8] W. W. S. Wei, *Time Series Analysis*. Addison-Wesley, 1990.
[9] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day, 1970.
[10] S. Lawrence Marple, Jr., *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[11] P. Bloomfield, *Fourier Analysis of the Time Series: An Introduction*. New York: Wiley, 1976.
[12] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco: Holden-Day, 1968.

[13] L. H. Koopmans, *The Spectral Analysis of Time Series*. New York: Academic, 1974.
[14] S. W. Lang and J. H. McClellan, "A simple proof of stability for all-pole linear prediction models," *Proc. IEEE*, vol. 67, pp. 860-861, May 1979.
[15] L. Pakula and S. Kay, "Simple proofs of the minimum phase property of the prediction error filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 501-502, Apr. 1983.
[16] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. New York: Springer-Verlag, 1976.
[17] J. M. Mendel, "Tutorial on higher order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278-305, Mar. 1991.
[18] C. L. Nikias and M. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, pp. 869-891, July 1987.
[19] G. B. Giannakis, "Cumulants: A powerful tool in signal processing," *Proc. IEEE*, vol. 75, pp. 1333-1334, 1987.
[20] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 360-377, Mar. 1989.
[21] G. B. Giannakis, "On the identifiability of non-Gaussian ARMA models using cumulants," *IEEE Trans. Automat. Contr.*, vol. 35, no. 1, pp. 18-26, Jan. 1990.
[22] A. Swami and J. M. Mendel, "Identifiability of the AR parameters of an ARMA process using cumulants," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 268-273, Feb. 1991.
[23] M. R. Raghuveer and C. L. Nikias, "Bispectrum estimation: A parametric approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 5, pp. 1213-1230, Oct. 1985.
[24] M. R. Raghuveer and C. L. Nikias, "Bispectrum estimation via AR modeling," *Signal Processing* (Special Issue on Modern Trends in Spectral Analysis), vol. 9, no. 1, pp. 35-48, Jan. 1986.
[25] A. Swami and J. M. Mendel, "AR identifiability using cumulants," in *Proc. Workshop Higher-Order Spectral Analysis*, Vail, CO, 1989, pp. 1-6.
[26] M. Huzii, "Estimation of an AR process by using a higher-order moment," *J. Time Series Anal.*, vol. 2, pp. 87-93, 1981.
[27] R. Pan and C. L. Nikias, "The complex cepstrum of higher order cumulants and nonminimum phase system identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 186-205, 1988.
[28] C. L. Nikias, "Higher-order spectral analysis," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed., vol. 1. Englewood Cliffs, NJ: Prentice-Hall, 1991, ch. 7.
[29] K.-S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Ann. Stat.*, vol. 10, pp. 1195-1208, 1982.
[30] J. K. Tugnait, "Recovering the poles from third-order cumulants of system output," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 1085-1089, 1989.
[31] J. K. Tugnait, "Identification of linear stochastic systems via second- and fourth-order cumulant matching," *IEEE Trans. Inform. Theory*, vol. 33, pp. 393-407, 1987.
[32] J. K. Tugnait, "Identification of nonminimum phase linear stochastic systems," *Automatica*, vol. 22, pp. 454-464, 1986.
[33] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from high-order moments," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 27-35, 1990.
[34] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[35] H. Akaike, "Statistical predictor identification," *Ann. Inst. Statis. Math.*, vol. 22, pp. 203-217, 1970.
[36] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716-723, 1974.
[37] E. Parzen, "An approach to time series modeling and forecasting illustrated by hourly electricity demands," Tech. Rep. 37, Stat. Sci. Div., State Univ. of New York, Jan. 1976.



Chong-Yung Chi (S'83-M'83-SM'89) was born in Taiwan, Republic of China, on August 7, 1952. He received the B.S. degree from the Tatung Institute of Technology, Taipei, Taiwan, in 1975, the M.S. degree from the National Taiwan University, Taipei, in 1977, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1983, all in electrical engineering.

From July 1983 to September 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA, where he worked on the design of various spaceborne radar remote sensing systems including radar scatterometers, SAR's, altimeters, and rain mapping radars. From October 1988 to July 1989, he was a visiting specialist at the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, Republic of China. Since August 1989 he has been a Professor with the Department of Electrical Engineering, National Tsing Hua University, Taiwan, Republic of China. He has published more than sixty technical papers in radar remote sensing, deconvolution and channel equalization, digital filter design, spectral estimation, and higher order statistics (HOS) based signal processing. His research interests include digital signal processing, statistical signal processing, system identification and estimation theory, signal processing with HOS, and adaptive filtering.