ELSEVIER

Signal Processing 41 (1995) 239–256

SIGNAL PROCESSING

A new identification algorithm for allpass systems by higher-order statistics

Chong-Yung Chi*, Jung-Yuan Kung

Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 30043, ROC

Received 20 January 1992; revised 28 September 1992

Foreword

The following paper was transmitted to Signal Processing Editor M. Kunt, by first author Prof. Chong-Yung Chi in January 1992. The paper was revised on 28 September 1992, and was accepted for publication on 7 January 1993. On 12 December 1993, Prof. Chi wrote to Editor M. Kunt that he had recently learned that, unbeknownst to him, his second author Jung-Yuan Kung (a former Ph.D. student who did not finish his Ph.D. work for Prof. Chi) had published an article [submitted 11 June 1993 (long after the Chi-Kung paper had been accepted for publication) and accepted for publication on 3 August 1993] in the International Journal of Electronics that duplicated the already accepted paper. This International Journal of Electronics article, whose major intellectual contributions were Prof. Chi's, was not co-authored by Chi, but, instead was co-authored by Prof. Chi's former Department Chairman, Prof. Yung-Chang Chen, for whom Mr. Kung did finish his Ph.D. work.

Prof. Chi had already received the page proofs for the Signal Processing paper. Editor M. Kunt who was unaware of all of the facts connected to this case, subsequently decided not to publish the already accepted paper. Prof. Chi's honesty and integrity cost him the publication of original research. When the international research community found out about this they alerted editor Kunt about the situation. Prof. Kunt subsequently reversed his decision. The result is that the originally accepted paper is being published here, as it would have been under normal circumstances.

Prof. J. Mendel
University of Southern California

Abstract

Based on a single cumulant of any order $M \geqslant 3$, a new allpass system identification algorithm with only non-Gaussian output measurements is proposed in this paper. The proposed algorithm, which includes both parameter estimation and order determination of linear time-invariant (LTI) allpass systems, outperforms other cumulant based methods such as least-squares estimators simply due to the more accurate model (allpass model) used by the former. It is applicable in channel equalization for the case of a phase distorted channel. Moreover, the well-known (minimum-phase) prediction error filter has been popularly used to deconvolve seismic signals where the source wavelet can be nonminimum phase

* Corresponding author.

SSDI 0165-1684(93)E0019-H

and speech signals where the vocal-tract filter can be nonminimum phase. Therefore, the proposed algorithm can be used to remove the remaining phase distortion of the nonminimum-phase source wavelet and nonminimum-phase vocal-tract filter in predictive deconvolved seismic signals and speech signals, respectively. It is also applicable in the minimum-phase – allpass decomposition based ARMA system identification method. Some simulation results and experimental results with real speech data are provided to support the claim that the proposed algorithm works well.

Zusammenfassung

In dieser Arbeit wird ein neuer Allpaß-Systemidentifikationsalgorithmus für nicht gaußsche Ausgangssignale vorgeschlagen, der auf einem einzigen Kumulanten beliebiger Ordnung $M \geqslant 3$ basiert. Der vorgeschlagene Algorithmus, der sowohl die Parameterschätzung als auch die Abschätzung der Ordnung linearer zeitinvarianter Allpaßsysteme beinhaltet, ist anderen auf Kumulanten basierenden Methoden wie Least-squares Schätzern überlegen, da hier ein besser angepaßtes Modell (Allpaß-Modell) zugrundeliegt. Er ist anwendbar zur Kanalentzerrung für den Fall phasenverzerrter Kanäle. Weiterhin ist das bekannte (minimalphasige) Prädiktionsfehlerfilter häufig zur Entfaltung seismischer Signale angewendet worden, wobei das Quellensignal nichtminimalphasig sein kann, als auch zur Entfaltung von Sprachsignalen, wobei das Vokaltrakt-Filter ebenfalls nichtminimalphasig sein kann. Deshalb kann der vorgeschlagene Algorithmus dazu benutzt werden, die verbleibenden Phasenverzerrungen von nichtminimalphasigen Quellensignalen bei seismischen Signalen oder von nichtminimalphasigen Vokaltrakt-Filtern in durch Prädiktion entfalteten Sprachsignalen zu korrigieren. Er ist ebenfalls auf die Zerlegung von ARMA-Systemen in minimalphasige und Allpaß-Anteile anwendbar. Er werden einige Simulationen und experimentelle Resultate anhand realer Sprachsignale vorgestellt, die die Funktionsfähigkeit des vorgeschlagenen Algorithmus belegen.

Résumé

Cet article propose un nouvel algorithme basé sur un seul cumulant d'ordre quelconque $M \geqslant 3$ permettant l'identification des systèmes passe-tout tout en nécessitant seulement la mesure des sorties non gaussiennes. L'algorithme proposé, qui inclut à la fois l'estimation des paramètres et la détermination de l'ordre des systèmes passe-touts invariant dans le temps, permet d'obtenir, grâce à l'utilisation d'un modèle plus adéquat (modèle passe-tout), des performances supérieures à celles obtenues par les autres méthodes cumulatives tel que la méthode d'estimation aux moindres carrées. Il est utilisable en égalisation de canaux dans le cas de canaux à distorsion de phase. De plus, le filtre d'erreur de prédiction (à phase minimum) a été largement utilisé pour la déconvolution des signaux sismiques dans lesquels l'ondelette de source peut être à phase non-minimal ainsi que pour la déconvolution des signaux de parole où le filtre de poursuite de la voix peut être à phase non-minimal. Par conséquent, l'algorithme proposé peut être utilisé pour éliminer la distorsion de phase résiduelle de l'onde de source à phase non-minimal et des filtres de poursuite de la parole à phase non-minimal respectivement pour la déconvolution prédictive des signaux sismiques et des signaux de parole. Il est également applicable à l'identification basée sur une décomposition à phase minimale passe-tout de systèmes ARMA. Des résultats de simulation et des résultats expérimentaux utilisant des données de parole réelles sont présentés pour démontrer que l'algorithme proposé fonctionne bien.

Keywords: Allpass systems; Cumulants; Non-Gaussian; Nonminimum-phase; Deconvolution

1. Introduction

Identification of linear time-invariant (LTI) systems with only output measurements is very important in many signal processing areas such as seismic deconvolution, channel equalization (in communications), radar, sonar, oceanography, speech signal processing, and image processing. Recently, cumulant (higher-order statistics) based identification [2, 3, 6–10, 17, 18, 26–28] of nonminimum-phase LTI systems with only non-Gaussian output measurements has drawn extensive attention in the previously mentioned signal processing areas because cumulants, which are blind to any kind of Gaussian process [17, 18], not only extract the amplitude information but also the phase information of nonminimum-phase LTI systems; meanwhile they are inherently immune from Gaussian measurement noise.

A general parametric model for a nonminimum-phase LTI system is known as the autoregressive moving average (ARMA) model, denoted $H(z)$. Giannakis and Mendel's well-known minimum-phase (MP) – allpass (AP) decomposition based methods [8, 9] basically consist of two steps. The first step includes the estimation of the spectrally equivalent minimum-phase $H_{\mathrm{MP}}(z)$ by a correlation based spectral estimation method. In the second step, they preprocess the output measurements of $H(z)$ by the inverse filter $1/H_{\mathrm{MP}}(z)$ to obtain an innovations process $\hat{u}(k)$. Thus, the identification of $H(z)$ is equivalent to the identification of the allpass system $H_{\mathrm{AP}}(z) = H(z)/H_{\mathrm{MP}}(z)$ [8, 9] with the preprocessed data $\hat{u}(k)$. In [8], $H_{\mathrm{AP}}(z)$ is estimated from slices of the sixth-order cumulant function of $\hat{u}(k)$ through a quite complicated procedure. In [9], fourth-order cumulants of $\hat{u}(k)$ are used to estimate the AR parameters of $H_{\mathrm{AP}}(z)$, which automatically provide the estimates of MA parameters. Chi and Kung [2, 3] proposed an allpass system classification algorithm for determining the $H_{\mathrm{AP}}(z)$ from all possible candidates of $H_{\mathrm{AP}}(z)$ based on the fact that poles of $H_{\mathrm{AP}}(z)$ are also zeros of the $H_{\mathrm{MP}}(z)$ obtained in the first step only by a single cumulant sample. However, their algorithm is not applicable in the case that $H(z)$ contains allpass factors. On the other hand, in some applications, the LTI system $H(z)$ of interest is known to be an allpass system such as phase distortion channels (allpass system) in channel equalization [4] where the channel input is often a sequence of non-Gaussian M-ary symbols. Another interesting instance is that in the well-known predictive deconvolution [22, 25], the LTI system of interest such as the source wavelet in seismic signals and the vocal-tract filter in speech signals is assumed to be minimum phase but it might be nonminimum phase in practice. Therefore, the deconvolved results can be viewed as the output of a phase distortion channel whose input is a non-Gaussian sparse spike reflectivity sequence in the seismic case or a non-Gaussian quasi-periodic positive pulse train in the voiced speech case [21]. Based on this fact and a key conclusion by an analysis that Wiggins' minimum entropy deconvolution (MED) algorithm has poor performance when the nonminimum-phase source wavelet is not an allpass system, Longbottom et al. [15] proposed a judicious deconvolution processing procedure summarized as follows.

MP–AP-removal procedure. Remove the MP part of the unknown nonminimum-phase system through a whitening processing such as predictive deconvolution and then remove the AP part of the unknown system by allpass system deconvolution algorithms.

They assumed that predictive deconvolved seismic signals, denoted $x(k)$, are the output of a constant phase shift system $H(f) = H(z = \mathrm{e}^{\mathrm{j}2\pi f}) = \mathrm{e}^{\mathrm{j}\varphi_0}$, which is a particular allpass system, driven by reflectivity sequences, and they processed $x(k)$ to remove the constant phase shift distortion by the minimum entropy criterion [15, 29] based on a single fourth-order cumulant. Shalvi and Weinstein [24] proposed an inverse filtering criterion also based on a single fourth-order cumulant which can surely be applied to identify any unknown allpass system, if the unknown allpass system is treated as a general LTI system.

In this paper, with a given set of non-Gaussian output measurements, we propose a new higher-order cumulant based allpass system identification algorithm without any prior knowledge about pole locations and order of the system. A popular parametric allpass model [19] given by (3) below, whose phase can be arbitrary rather than constant phase, is used. The new algorithm is also based on a single higher-order cumulant whose order can be any integer greater than two, and it also includes the order determination.

In Section 2, we present the new allpass system identification algorithm. Some simulation results and experimental results with real voiced speech data are then provided to support the proposed algorithm in Section 3. Finally, we draw some conclusions.

2. The new allpass system identification algorithm

Assume that data $x(k)$, $k = 0, 1, \dots, N-1$, were generated from a real stable pth-order allpass LTI system with input $u(k)$ as follows:

$$x(k) = v(k) + w(k), \tag{1}$$

where

$$v(k) = -\sum_{i=1}^{p} a_i v(k-i) + u(k-p) + \sum_{i=1}^{p} a_i u(k-p+i), \quad (2)$$

and $w(k)$ is additive measurement noise. Equivalently, the allpass system has a rational transfer function given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{z^{-p} + a_1 z^{-p+1} + \cdots + a_p}{1 + a_1 z^{-1} + \cdots + a_p z^{-p}}, \quad (3)$$

where $A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p}$ and $B(z) = A(z^{-1}) z^{-p}$. The new allpass system identification algorithm to be presented below is based on the following modeling assumptions:

(A1) The allpass system $H(z)$ is causal and exponentially stable, i.e., $A(z) \neq 0$ for $|z| \geqslant 1$.

(A2) The input $u(k)$ is real, zero-mean, stationary, independent identically distributed (i.i.d.), non-Gaussian with nth-order cumulant γ_n. Moreover, $|\gamma_n| < \infty$ for $2 \leqslant n \leqslant 2M$ where $M \geqslant 3$ is a positive integer.

(A3) The measurement noise $w(k)$ is Gaussian which can be white or colored with unknown statistics.

(A4) The input $u(k)$ is statistically independent of $w(k)$.

For ease of later use, let $S_{AP}(L)$ denote the set of all Lth-order anticausal stable allpass filters $\hat{H}_L(z)$ where

$$\hat{H}_L(z) = \frac{1 + \hat{a}_1 z^{-1} + \cdots + \hat{a}_L z^{-L}}{z^{-L} + \hat{a}_1 z^{-L+1} + \cdots + \hat{a}_L}, \quad (4)$$

with L poles outside the unit circle. Note that $|\hat{H}_L(f)| = |\hat{H}_L(z = \exp\{j2\pi f\})| = 1$ for all f, and that $|H(f)| = 1$ for all f and $1/H(z) \in S_{AP}(p)$. Next, we present the following theorem on which the new allpass system identification algorithm is based.

Theorem 1. *Assume that $x(k)$ was generated from (1) under the previous assumptions (A1)–(A4). Let $y(k)$ be the output of a pth-order allpass filter $\hat{H}_p(z) \in S_{AP}(p)$ with the input $x(k)$. Then the absolute Mth-order ($M \geqslant 3$) cumulant $|C_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)|$ is maximum if and only if $\hat{H}_p(z) = 1/H(z)$ where $C_{M,y}(k_1, k_2, \ldots, k_{M-1})$ is the Mth-order cumulant function of $y(k)$. Furthermore,* $\max\{|C_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)|\} = |\gamma_M|$.

The proof of this theorem is given in Appendix A.

Based on Theorem 1, when the order of the allpass system $H(z)$ is known a priori, one can identify $H(z)$ by maximizing the following objective function:

$$J(L) = \hat{C}^2_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0), \quad (5)$$

with L set to p, where $\hat{C}_{M,y}(k_1, k_2, \ldots, k_{M-1})$ is the Mth-order sample cumulant function [17, 26] of the output, $y(k)$, of $\hat{H}_L(z) \in S_{AP}(L = p)$ with the input $x(k)$. For example, the biased third-order sample cumulant $\hat{C}_{3,y}(k_1, k_2)$ is given by

$$\hat{C}_{3,y}(k_1, k_2) = \frac{1}{N} \sum_{k=0}^{N-1} y(k) y(k+k_1) y(k+k_2). \quad (6)$$

Remark that the output $y(k)$ of $\hat{H}_L(z)$ must be computed backwards as follows:

$$y(k) = -\sum_{j=1}^{L} a_j y(k+j) + x(k+L) + \sum_{j=1}^{L} a_j x(k+L-j), \quad (7)$$

$k = N-1, N-2, \ldots, 0$, because $\hat{H}_L(z)$ is anticausal stable. Moreover, since $J(L)$ is a highly nonlinear function of the coefficients of $\hat{H}_L(z)$, it is almost impossible to find a closed-form solution for the coefficients of the optimum $\hat{H}_L(z)$. Instead, we resort to an iterative numerical optimization method to search for the desired $\hat{H}_L(z)$. Next we illuminate the identification procedure.

When the order of the allpass $H(z)$ is known a priori, the proposed algorithm estimates $\hat{H}(z)$ through the following procedure:

Parameter estimation

(s0) Set $L = p$.

(s1) Search for $J_{max}(L)$ (maximum of $J(L)$) and the associated Lth-order $\hat{H}_L(z) \in S_{AP}(L)$ by a Newton–Raphson type iterative algorithm. The optimum estimates of $H(z)$ and γ_M are given by

$$\hat{H}(z) = 1/\hat{H}_L(z) \quad (8)$$

and

$$\hat{\gamma}_M = \hat{C}_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0), \quad (9)$$

respectively.

The Newton–Raphson type iterative algorithm used in (s1) for the case of $M = 3$ is summarized in Appendix B.

It is advisable here to turn to the differences between the proposed algorithm and Shalvi–Weinstein's inverse filtering algorithm [24] and those between the proposed algorithm and a minimum entropy criterion based algorithm used by Longbottom et al. [15].

Let us express noisy measurements $x(k)$ modeled by (1) in the following convolutional form:

$$x(k) = u(k)*h(k) + w(k), \quad (10)$$

where $h(k)$ is the impulse response of the allpass system $H(z)$. Let $g(k)$ denote the inverse filter of $h(k)$ and

$$z(k) = x(k)*g(k) \quad (11)$$

be the output of $g(k)$ in response to $x(k)$. Shalvi and Weinstein [24] proposed a blind equalizer for estimating $u(k)$ by maximizing the absolute fourth-order cumulant of $z(k)$, denoted J_{SW}, i.e.,

$$J_{\mathrm{SW}}(z(k)) = |C_{4,z}(0,0,0)| \quad (12)$$

with respect to the coefficients of the assumed FIR filter $g(k)$ under the constraint $E\{|z(k)|^2\} = E\{|u(k)|^2\}$. The obtained optimum equalizer corresponds to the inverse filter $1/H(z)$ except for a constant delay factor. Their algorithm requires M (cumulant order) to be equal to four. Let us emphasize that they treat $h(k)$ as a general LTI system no matter whether $h(k)$ is an allpass system or not. On the other hand, the proposed algorithm maximizes $C^2_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)$ with the constraint of allpass filter with unity gain which also implies $E\{|y(k)|^2\} = E\{|u(k)|^2\}$. Note that the proposed algorithm does not require $M = 4$ as long as $\gamma_M \neq 0$ and $M \geqslant 3$ although maximizing $C^2_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)$ is equivalent to maximizing $J_{\mathrm{SW}}(y(k))$ for $M = 4$; meanwhile, it takes into account the particular structure of the allpass model given by (3). Moreover, it also includes order determination, to be presented below.

Next, let us present the differences between the proposed algorithm and the minimum entropy criterion based allpass system identification algorithm reported in [15].

Wiggins proposed an MED algorithm [29] which deconvolves $x(k)$ by maximizing the varimax norm of $z(k)$,

$$V_z = E\{z^4(k)\}/\{\mathrm{var}[z(k)]\}^2. \quad (13)$$

Longbottom et al. [15] reported that the MED algorithm has poor performance when $h(k)$ is not an allpass system. The reader can refer to [15] for a detailed analysis of the performance of the MED algorithm. Instead, they [15] applied the MED criterion to the identification of a constant phase shift system, which is a particular allpass system with $H(f) = \mathrm{e}^{\mathrm{j}\varphi_0}$. With the inverse filter $G(f)$ set to $G(f) = \mathrm{e}^{\mathrm{j}\varphi}$, they try to remove the constant phase distortion $H(f) = \mathrm{e}^{\mathrm{j}\varphi_0}$ by maximizing the standardized fourth-order cumulant of $z(k)$ defined as

$$Q_z = \frac{C_{4,z}(0,0,0)}{\{\mathrm{var}[z(k)]\}^2} = V_z - 3 \quad (14)$$

with respect to the single parameter φ. The optimum φ turns out to be equal to $-\varphi_0$. On the other hand, we assumed that the unknown allpass system $h(k)$ is a parametric allpass model given by (3) whose phase is arbitrary, and the constant phase shift system can be approximated by (3) with a sufficient order. However, one can easily see, from (14) and the fact that $\mathrm{var}[z(k)] = \mathrm{var}[u(k)]$, that maximizing $C^2_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)$ for $M = 4$ is indeed equivalent to maximizing Q_y when $C_{4,y}(0,0,0) > 0$, but $C_{4,y}(0,0,0)$ can be negative for some non-Gaussian processes. The proposed allpass system identification algorithm is theoretically complete for all $M \geqslant 3$. Moreover, it includes order determination to be presented next.

When the order of the allpass system $H(z)$ is not known a priori, one must determine the order of $H(z)$ prior to the estimation of the coefficients of $H(z)$. The order determination algorithm to be presented below is based on the following fact.

Fact 1. *Assume that $x(k)$ was generated from (1) under the previous assumptions (A1)–(A4). Let $x(k)$ be the input of an arbitrary allpass system with amplitude response equal to unity and $y(k)$ be the corresponding output. Then the following equation holds:*

$$\sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{M-1}}^{\infty} C_{M,y}^2(k_1,\ldots,k_{M-1}) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{M-1}=-\infty}^{\infty} C_{M,x}^2(k_1,\ldots,k_{M-1}) = \gamma_M^2, \tag{15}$$

where $M \geqslant 3$.

Fact 1 can be easily shown via Parseval's theorem since $|S_{M,y}(f_1,\ldots,f_{M-1})| = |S_{M,x}(f_1,\ldots,f_{M-1})| = |\gamma_M|$ where $S_{M,x}(f_1,\ldots,f_{M-1})$, known as the Mth-order polyspectrum of $x(k)$, is the $(M-1)$-dimensional Fourier transform of $C_{M,x}(k_1,\ldots,k_{M-1})$ and $S_{M,y}(f_1,\ldots,f_{M-1})$ is the Mth-order polyspectrum of $y(k)$.

The order determination algorithm is also a threshold decision rule given by (21) below based on the first- and second-order statistics of $\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)$ (see (20) below). Next, let us present how we estimate the mean and the variance of $\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)$.

A large-sample property for the third-order sample cumulant $\hat{C}_{3,y}(m,n)$ was reported in [11] as follows:

(p1) For N sufficiently large, $\hat{C}_{3,y}(m,n)$ is approximately Gaussian distributed with mean $C_{3,y}(m,n)$ and variance

$$\sigma^2\{\hat{C}_{3,y}(m,n)\} = \frac{1}{N}\sum_{i=-q-n}^{q+m}\left(1-\frac{|i|}{N}\right) \times E\{[y(0)y(m)y(n) - C_{3,y}(m,n)] \times [y(i)y(i+m)y(i+n) - C_{3,y}(m,n)]\}, \tag{16}$$

where the integer q is associated with an approximate MA(q) process to the ARMA process $y(k)$.

As mentioned by Giannakis and Mendel [11], the estimate of the variance of $\hat{C}_{3,y}(0,0)$ can be obtained by (16), in which the ensemble average is replaced by time average, as follows:

$$\hat{\sigma}^2\{\hat{C}_{3,y}(0,0)\} = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=-q}^{q}\left(1-\frac{|i|}{N}\right) \times [y(j)^3 - \hat{C}_{3,y}(0,0)] \times [y(i+j)^3 - \hat{C}_{3,y}(0,0)]. \tag{17}$$

However, for any other $M>3$, $\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)$ is also asymptotically Gaussian with mean $C_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)$, but the variance $\sigma^2[\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)]$ does not seem to have a neat closed-form expression as the one given by (16) for $M=3$. The reader can refer to [5, 11, 14, 23] for the derivation of $\sigma^2[\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)]$ for $M>3$.

Note, by Theorem 1 and Fact 1, that when $\hat{H}_p(z) = 1/H(z)$,

$$C_{M,y}^2(k_1=0, k_2=0,\ldots,k_{M-1}=0) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{M-1}=-\infty}^{\infty} C_{M,y}^2(k_1,\ldots,k_{M-1}) = \gamma_M^2. \tag{18}$$

Therefore, we estimate the mean of $\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)$ by

$$\hat{E}[\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)] = \operatorname{sign}(\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)) \times \left\{\sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{M-1}=-\infty}^{\infty} \hat{C}_{M,y}^2(k_1,\ldots,k_{M-1})\right\}^{1/2}, \tag{19}$$

where $\operatorname{sign}(\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0))$ denotes the sign of $\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)$.

Our order determination is based on the following statistic $T(y(k))$:

$$T(y(k)) = \{\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)\} \times [\hat{\sigma}\{\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)\}]^{-1} - \{\hat{E}[\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)]\} \times [\hat{\sigma}\{\hat{C}_{M,y}(k_1=0, k_2=0,\ldots,k_{M-1}=0)\}]^{-1} \tag{20}$$

Note that $T(y(k))$ is approximately a Gaussian random variable with zero mean and unit variance when $\hat{H}_p(z) = 1/H(z)$ and N is sufficiently large. Our identification algorithm proceeds with the order determination of $H(z)$ as follows:

Order determination.

(s2) Set $L = 0$ and compute $J_{\max}(0) = \hat{C}^2_{M,x}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)$.

(s3) Set $L = L + 1$.

(s4) Execute (s1).

(s5) If $|J_{\max}(L) - J_{\max}(L-1)|/J_{\max}(L-1) > \xi$ where ξ is a preassigned positive constant, then go to (s3).

(s6) If

$$|T(y(k))| \leqslant \eta, \tag{21}$$

where η is the threshold associated with a given confidence $(1 - \Delta) = P_r[|T(y(k))| \leqslant \eta]$, then the estimated order $\hat{p} = L - 1$, otherwise, go to (s3).

When the identification procedure is complete, $\hat{H}(z) = 1/\hat{H}_{L-1}(z)$ and the estimate $\hat{\gamma}_M$ is given by (9) associated with $\hat{H}_{L-1}(z)$. The following two remarks are necessary now:

(R1) For the case of $M = 3$, $T(y(k))$ can be easily computed by substituting (6), (17) and (19) into (20) and the threshold η for a given confidence $(1 - \Delta)$ can be easily found from tables of Gaussian distribution [20] since $T(y(k))$ is approximately Gaussian. However, for the case of $M > 3$, one must be able to estimate $\sigma^2[\hat{C}_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)]$ whose derivation is quite tedious. For some signal processing applications, by our experience, removing procedure (s6) for the case of $M > 3$ seems acceptable. Example 2 in Section 3 will provide some simulation results for the case of $M = 4$ in which the system order was determined through the procedure (s2)–(s5) without (s6).

(R2) The second term of $\hat{E}[\hat{C}_{M,y}(k_1 = 0, k_2 = 0, \ldots, k_{M-1} = 0)]$ given by (19) can only be calculated over a finite $(M-1)$-dimensional region. We calculate it over the finite domain of support $F(q)$ of the Mth-order cumulant function of the previously mentioned approximate MA(q) process to the non-Gaussian ARMA process $y(k)$ (see (p1)). For instance, for $M = 3$, the finite region $F(q)$ is the following hexagonal region:

$$F(q) = \{(k_1, k_2) \,|\, |k_1| \leqslant q, |k_2| \leqslant q, \\ |k_1 - k_2| \leqslant q\}, \tag{22}$$

which is centered at the origin and symmetric with respect to the origin.

The proposed algorithm works well due to the following characteristics:

(C1) The proposed algorithm begins with the zeroth-order $\hat{H}_L(z) = 1$ $(L = 0)$. In the Lth iteration, the AR parameter estimates $\hat{a}_1, \ldots, \hat{a}_{L-1}$ associated with the optimum $\hat{H}_{L-1}(z)$ together with $\hat{a}_L = 0$ can be used as the initial guess for the parameters of $\hat{H}_L(z)$. On the other hand, in each iteration of the Newton–Raphson type iterative algorithm in (s1), the objective function $J(L)$ is guaranteed to increase. Therefore, $J_{\max}(L) \geqslant J_{\max}(L-1)$ or $J_{\max}(L)$ increases monotonically with L. Furthermore, one can easily show that $J_{\max}(L)$ is bounded as long as $H(z)$ is an arbitrary exponentially stable LTI system. Hence, the proposed algorithm is guaranteed to converge.

(C2) The proposed identification algorithm performs the order determination and parameter estimation simultaneously.

(C3) Without doubt, various cumulant based least squares (LS) methods [6, 7, 10, 26, 28] can be applied to estimating the AR parameters of the allpass system $H(z)$ by treating $H(z)$ as a general ARMA model, while the MA parameters of the allpass system $H(z)$ are automatically determined (see (3)). The proposed algorithm outperforms cumulant based LS methods simply because of the more accurate model (allpass model) used by the former.

(C4) The proposed allpass system identification algorithm is a consistent estimator. This fact can be justified as follows. The cumulant estimate $\hat{C}_{M,y}(k_1, k_2, \ldots, k_{M-1})$ (see (6) for $M = 3$) is a consistent estimate of $C_{M,y}(k_1, k_2, \ldots, k_{M-1})$ due to assumptions

(A1) and (A2) [17, 10]. Thus, $J(L)$ is a consistent estimate of $C^2_{M,y}(k_1 = 0,\ k_2 = 0, \dots, k_{M-1} = 0)$. Therefore, we can conclude, by Theorem 1 and Fact 1, that $J_{\max}(L)$ is a consistent estimate of $C^2_{M,y}(k_1 = 0,\ k_2 = 0, \dots, k_{M-1} = 0)$ associated with $\hat{H}_L(z) = 1/H(z)$ when $L \geqslant p$. In other words, the proposed allpass system identification algorithm is a consistent estimator.

(C5) The proposed identification algorithm can be used for any $M \geqslant 3$ (order of cumulants) as long as the Mth-order cumulant γ_M of the driving input $u(k)$ is not equal to zero.

3. Simulation results and experimental results with voiced speech data

In this section, we show two simulation examples and one set of experimental results with real voiced speech data to demonstrate that the proposed all-pass system identification algorithm works well.

3.1. Simulation results

We calculated the quantity $\hat{E}[\hat{C}_{M,y}(k_1 = 0,\ k_2 = 0, \dots, k_{M-1} = 0)]$ given by (19) over the finite region $F(15)$ (see (22)) during the identification procedure (s6) of the proposed algorithm in our simulation. The first example includes various performance tests of the proposed algorithm, while the second example is to employ the proposed algorithm to remove the remaining phase distortion of source wavelets in predictive deconvolved data. Then, we employed the proposed algorithm to remove the remaining phase distortion of a vocal-tract filter in predictive deconvolved speech data. Now, let us turn to Example 1.

Example 1. The driving noise $u(k)$ used was a zero-mean, exponentially distributed i.i.d. random sequence with variance $\sigma_u^2 = 1$ and skewness $\gamma_3 = 2$. We let this sequence pass through a selected allpass model $H(z)$ to obtain the noise-free output signal $v(k)$ and then added a zero-mean white or colored Gaussian noise sequence $w(k)$ to $v(k)$ to form the noisy data $x(k)$. The order of cumulants used was $M = 3$ and the length of data was $N = 1024$. Two cases in our simulations are presented below. In the first case the order of $H(z)$ is known a priori. The second case concerns order determination.

Case I: Allpass system with known order. In this case, we first present some simulation results associated with a second-order allpass system and then some simulation results associated with a sixth-order allpass system. Four sets of synthetic data were generated for four different signal-to-noise ratios (SNR) of ∞, 100, 10 and 1, respectively. Each set of data includes 30 independent realizations of $x(k)$.

(A) *Second-order allpass system.* The second-order allpass system with the AR parameters $a_1 = -0.3$ and $a_2 = -0.4$ (taken from [10]) was used. First of all, let us show some simulation results (see Table 1) using the synthetic data for $w(k)$ as white Gaussian noise. The simulation results shown in Table 1 were obtained through the previous procedure, (s0) and (s1). One can see, from Table 1, that both bias and standard deviation become smaller as SNR gets larger and that the values of bias and standard deviation for SNR = 1 (a quite low SNR) are also small. These simulation results support the good performance of the proposed allpass system identification algorithm. On the other hand, with the same synthetic data, we also obtained the corresponding results using a cumulant based least squares (LS) estimator reported in [6, 7, 10, 26, 28], which treats $H(z)$ as a general ARMA (p, p) model, as follows:

$$C_{3,x}(-m, -m) + \sum_{k=1}^{p} a_k C_{3,x}(-m+k,\ -m+k) = 0, \quad \text{for } m \geqslant p+1, \tag{23}$$

where $p = 2$. In this case, we concatenated (23) for $m = 3, 4, \dots, 12$, with $C_{3,x}(-m,\ -m)$ replaced by $\hat{C}_{3,x}(-m,\ -m)$ and then obtained the LS estimates of the AR parameters a_1 and a_2. We then estimated γ_3 as

$$\hat{\gamma}_3 = \frac{1}{N} \sum_{k=0}^{N-1} y^3(k), \tag{24}$$

Table 1
Case I(A) (second-order allpass system) of Example 1. Simulation results obtained by the proposed algorithm for $w(k)$ being white Gaussian ($a_1 = -0.3$, $a_2 = -0.4$, $\gamma_3 = 2$, $N = 1024$, 30 independent runs), estimated values (mean ± one standard deviation)

SNR	$\hat{a}_1$	$\hat{a}_2$	$\hat{\gamma}_3$
∞	−0.3021 ± 0.0147	−0.4001 ± 0.0162	1.9892 ± 0.0638
100	−0.3028 ± 0.0193	−0.4025 ± 0.0173	1.9807 ± 0.0791
10	−0.3037 ± 0.0235	−0.3961 ± 0.0351	2.0204 ± 0.1290
1	−0.3191 ± 0.0378	−0.4127 ± 0.0483	2.0359 ± 0.2443

Table 2
Case I(A) (second-order allpass system) of Example 1. Simulation results obtained by the LS estimator for $w(k)$ being white Gaussian ($a_1 = -0.3$, $a_2 = -0.4$, $\gamma_3 = 2$, $N = 1024$, 30 independent runs), estimated values (mean ± one standard deviation)

SNR	$\hat{a}_1$	$\hat{a}_2$	$\hat{\gamma}_3$
∞	−0.2984 ± 0.1165	−0.3687 ± 0.1065	1.8600 ± 0.1524
100	−0.2944 ± 0.1158	−0.3718 ± 0.1076	1.8544 ± 0.1521
10	−0.2858 ± 0.1240	−0.3740 ± 0.1194	1.8208 ± 0.1679
1	−0.2486 ± 0.2067	−0.3039 ± 0.1824	1.5577 ± 0.4814

where $y(k)$ is the output of the inverse filter $\hat{H}_2(z)$ (see (4)) with $\hat{a}_1$ and $\hat{a}_2$ replaced by the LS estimates. These simulation results are shown in Table 2. Comparing Table 2 with Table 1, one can see that both bias and standard deviation of the estimates $\hat{a}_1, \hat{a}_2$ and $\hat{\gamma}_3$ shown in Table 1 are smaller than the corresponding bias and standard deviation shown in Table 2 for each SNR except that the bias of $\hat{a}_1$ shown in Table 2 is slightly smaller than the corresponding value shown in Table 1 for SNR = ∞. The lower the SNR, the more the proposed identification algorithm outperforms the LS method. These results also support (C3).

Next, we show some simulation results for the case of $w(k)$ being colored Gaussian noise which was generated from an MA(2) system, $B(z) = 1 - 1.2z^{-1} + 0.32z^{-2}$, with input being a zero-mean white Gaussian sequence. The amplitude spectrum of $B(z)$ is shown in Fig. 1 which indicates that $B(z)$ is a highpass filter. The obtained simulation results using the proposed algorithm and those using the previous LS method are shown in Tables 3 and 4, respectively. Again, the same conclusion as in the previous case of white Gaussian noise, that our algorithm works well and outperforms the cumulant based LS method, can be drawn from these two tables.

(B) *Sixth-order allpass system.* A sixth-order allpass system with poles located at 0.98, 0.7, − 0.5, 0.4 ± 0.4j, − 0.2, was used. We performed the simulation with the synthetic data for $w(k)$ as white Gaussian noise. The simulation results obtained by the proposed algorithm are shown in Table 5. One can see, from Table 5, that the proposed algorithm can also accurately estimate the coefficients of this sixth-order allpass system which has a strong pole ($z = 0.98$) (close to the unit circle) and a weak pole ($z = -0.2$) (close to the origin) among the six poles. These simulation results also imply that the performance of the proposed algorithm is not sensitive to the pole locations of allpass systems.

Case II. Order determination. The same second-order allpass system as that in Case I(A) was used in this case. The order of the system was determined through the procedure (s2)–(s6). We set the confidence $(1 - \Delta)$ to 95% and consequently the

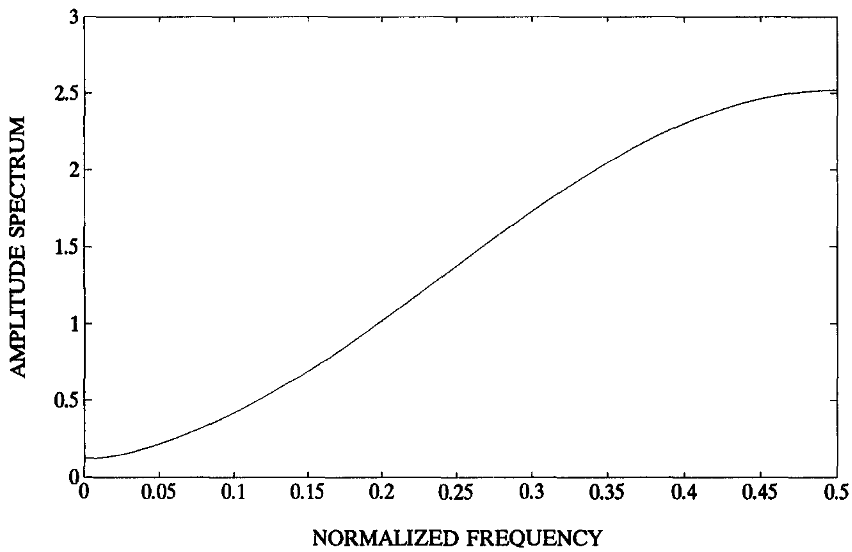


Fig. 1. The amplitude spectrum of the MA(2) used for generating colored Gaussian noise $w(k)$.

Table 3
Case I(A) (second-order allpass system) of Example 1. Simulation results obtained by the proposed algorithm for $w(k)$ being coloured Gaussian ($a_1 = -0.3$, $a_2 = -0.4$, $\gamma_3 = 2$, $N = 1024$, 30 independent runs), estimated values (mean ± one standard deviation)

SNR	$\hat{a}_1$	$\hat{a}_2$	$\hat{\gamma}_3$
∞	−0.3021 ± 0.0147	−0.4001 ± 0.0162	1.9892 ± 0.0638
100	−0.3032 ± 0.0182	−0.3974 ± 0.0187	1.9765 ± 0.0823
10	−0.3048 ± 0.0246	−0.3952 ± 0.0393	2.0287 ± 0.1303
1	−0.3276 ± 0.0715	−0.3723 ± 0.0658	2.0474 ± 0.2918

Table 4
Case I(A) (second-order allpass system) of Example 1. Simulation results obtained by the LS estimator for $w(k)$ being colored Gaussian ($a_1 = -0.3$, $a_2 = -0.4$, $\gamma_3 = 2$, $N = 1024$, 30 independent runs), estimated values (mean ± one standard deviation)

SNR	$\hat{a}_1$	$\hat{a}_2$	$\hat{\gamma}_3$
∞	−0.2984 ± 0.1165	−0.3687 ± 0.1065	1.8600 ± 0.1524
100	−0.2961 ± 0.1196	−0.3670 ± 0.1099	1.8581 ± 0.1609
10	−0.2934 ± 0.1373	−0.3608 ± 0.1268	1.8315 ± 0.2029
1	−0.2627 ± 0.2275	−0.3077 ± 0.2200	1.5991 ± 0.4287

threshold η was 1.9601 in (s6). We performed the simulation with the synthetic data for $w(k)$ as white Gaussian noise for the case of SNR ∈ − 5–20 dB. For each SNR, 100 independent runs were conducted. The successful number of order determination among 100 independent runs for each SNR is shown in Fig. 2. From Fig. 2, one can see that the number of successes is greater than 93 for each SNR for the given confidence of 95% except that the number of successes is 90 for

Table 5
Case I(B) (sixth-order allpass system) of Example 1. Simulation results obtained by the proposed algorithm for $w(k)$ being white Gaussian ($a_1 = -1.78$, $a_2 = 0.714$, $a_3 = 0.3106$, $a_4 = -0.306$, $a_5 = 0.0451$, $a_6 = 0.0219$, $\gamma_3 = 2$, $N = 1024$, 30 independent runs), estimated values (mean $\pm$ one standard deviation)

SNR	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	$\hat{\gamma}_3$
∞	-1.7257 ± 0.0156	0.6843 ± 0.0086	0.2952 ± 0.0019	-0.2955 ± 0.0046	0.0218 ± 0.0093	0.0259 ± 0.0108	1.8004 ± 0.1633
100	-1.7258 ± 0.0137	0.6841 ± 0.0097	0.2949 ± 0.0025	-0.2952 ± 0.0051	0.0216 ± 0.0084	0.0260 ± 0.0119	1.7914 ± 0.1686
10	-1.7246 ± 0.0141	0.6820 ± 0.0128	0.2947 ± 0.0024	-0.2946 ± 0.0078	0.0211 ± 0.0089	0.0257 ± 0.0129	1.7859 ± 0.2352
1	-1.7132 ± 0.0134	0.6772 ± 0.0201	0.2940 ± 0.0033	-0.2934 ± 0.0094	0.0139 ± 0.0086	0.0167 ± 0.0124	1.7551 ± 0.3834

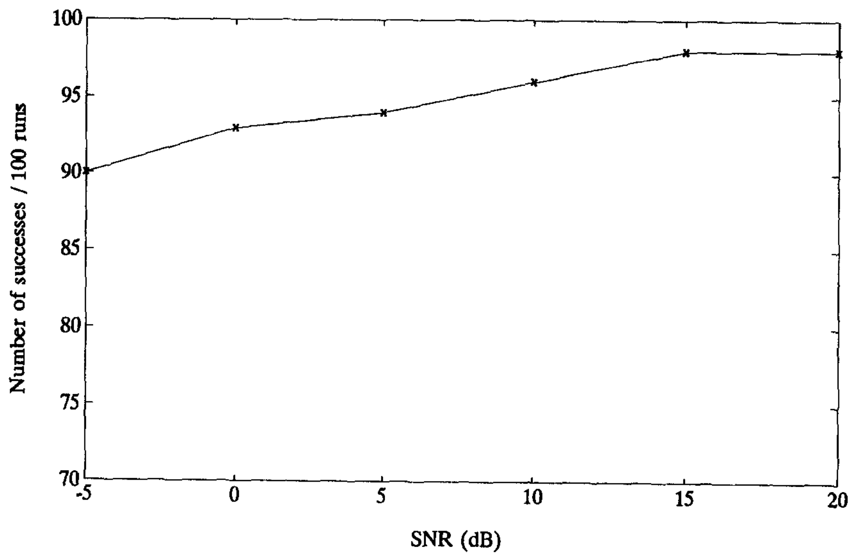


Fig. 2. Percentage of correct order determination using the proposed allpass identification algorithm with 95% confidence (Case II of Example 1).

SNR = -5 dB (a very low SNR). These simulation results support the claim that the proposed allpass system identification algorithm can determine the order of all-pass systems well.

Example 2. In seismic deconvolution, a source wavelet $h(k)$ is input to the Earth and the received noisy data $x(k)$ can be modeled as (10) where $u(k)$ is the reflectivity sequence of the Earth and $w(k)$ is the measurement noise. Deconvolution is a signal processing procedure of removing the effects of $h(k)$ and suppressing the noise $w(k)$ from data $x(k)$ such that only the desired signal $u(k)$ is left. Conventionally, $u(k)$, except for a scale factor, is estimated by the (minimum-phase) prediction error filter (PEF) [22, 25] which assumes that $u(k)$ is a white noise sequence and $h(k)$ is minimum phase. However, $u(k)$ is usually a non-Gaussian sparse spike train and $h(k)$ can be nonminimum phase in practice. Nevertheless, the PEF, whose coefficients are obtained by solving the correlation based Yule–Walker equations, has been popularly used in seismic deconvolution in the past three decades. Now, let us present some simulation results to show the removal of the remaining phase distortion of source wavelet in predictive deconvolved results by the proposed algorithm. That is to say, in this example, the signal processing procedure is exactly the MP–AP-removal procedure mentioned in Section 1.

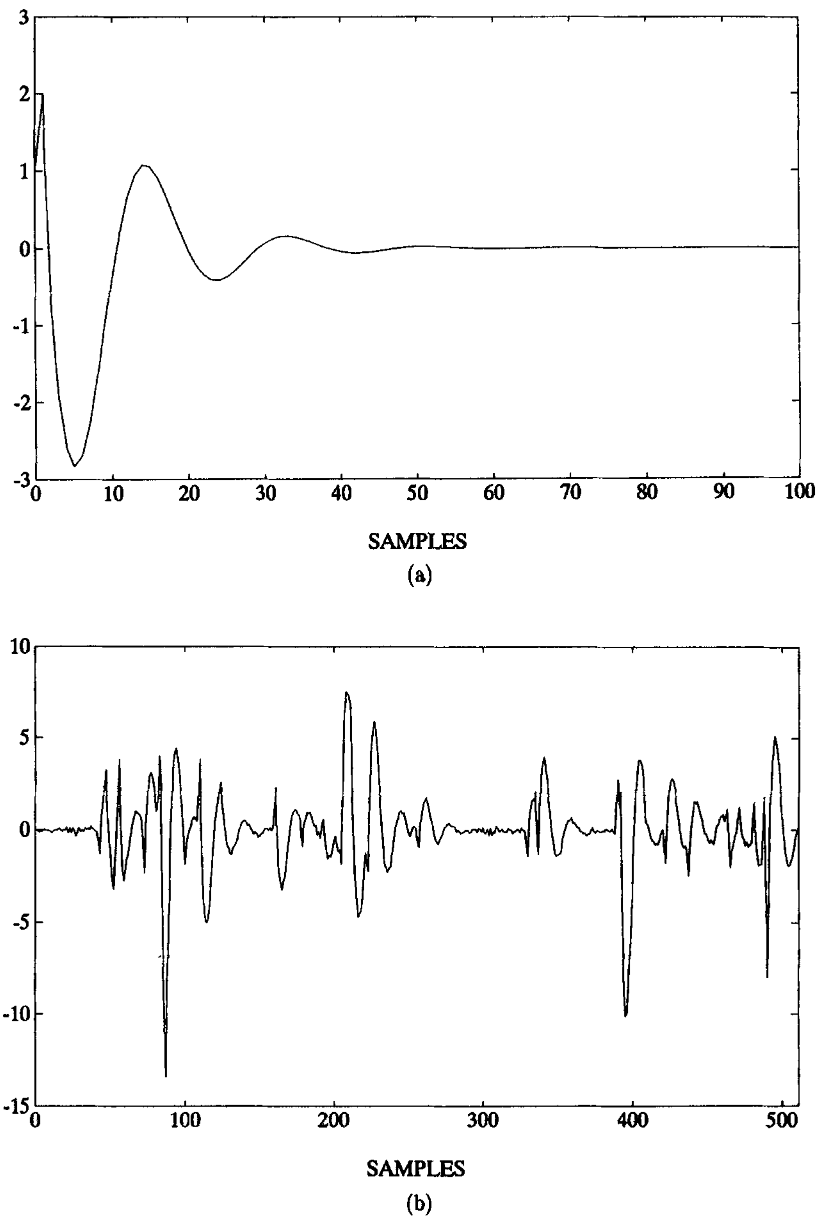


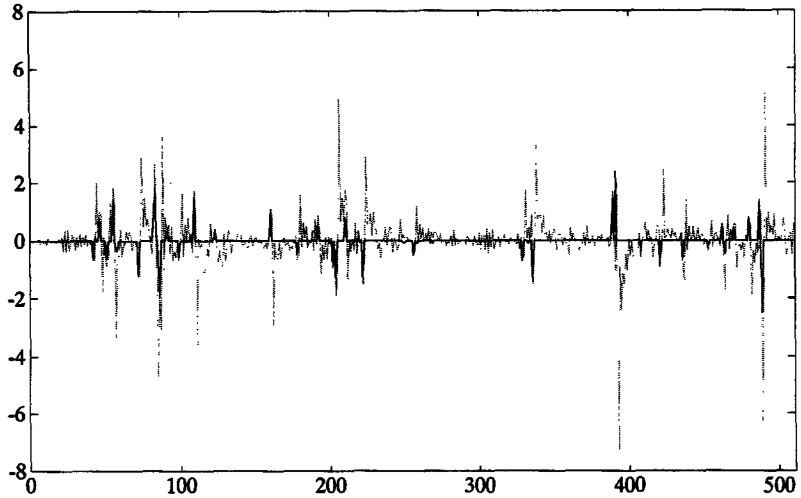
Fig. 3. (a) Source wavelet $h(k)$, (b) synthetic noisy data for SNR = 27 dB.

The nonminimum-phase source wavelet $h(k)$ shown in Fig. 3(a) was used whose transfer function is given by

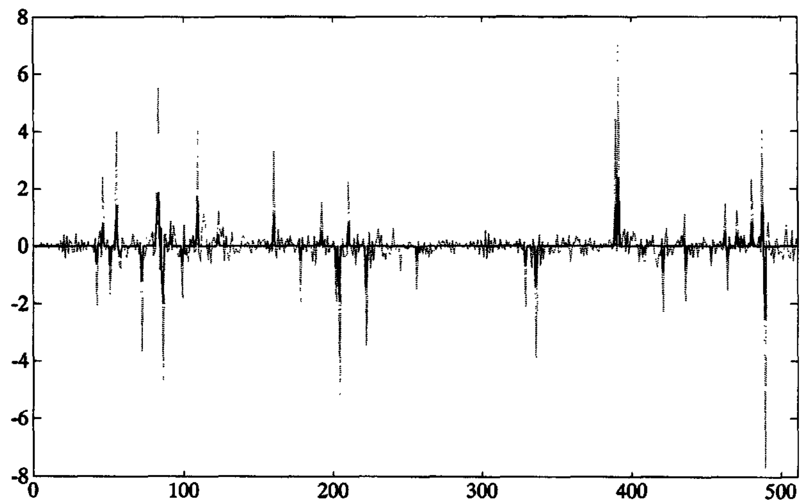
$$H(z) = \frac{1 + 0.1z^{-1} - 3.2725z^{-2} + 1.41125z^{-3}}{1 - 1.9z^{-1} + 1.1525z^{-2} - 0.1625z^{-3}}, \tag{25}$$

which has zeros located at $-2.0415, 1.4719, 0.4696$ and poles located at $0.2, 0.85 \pm 0.3j$. The driving input was an i.i.d. Bernoulli–Gaussian (B–G) sequence [13, 16], which has been used to model the sparse reflectivity sequence in seismic deconvolution, defined by

$$u(k) = r(k)q(k) \tag{26}$$



(c)



(d)

Fig. 3. (c) True input signal $u(k)$ (solid line) and the predictive deconvolved data $e(k)$ (dotted line); (d) true input signal $u(k)$ (solid line) and the output $y(k)$ (dotted line) of the optimum allpass filter $\hat{H}_L(z)$ where $L = 2$.

where $r(k)$ is a zero-mean white Gaussian random sequence with variance σ_r^2 and $q(k)$ is a Bernoulli sequence for which

$$P_r[q(k)] = \begin{cases} \lambda, & q(k) = 1, \\ 1 - \lambda, & q(k) = 0. \end{cases} \tag{27}$$

A B–G sequence $u(k)$ was generated with parameters $\lambda = 0.1$ and $\sigma_r^2 = 1$, and then $N = 512$ synthetic data of $x(k)$ shown in Fig. 3(b) were generated based on (10) for SNR = 27 dB and $w(k)$ being white Gaussian noise. The skewness of $u(k)$ is $\gamma_3 = 0$ and the kurtosis of $u(k)$ is $\gamma_4 = 0.27$ for this

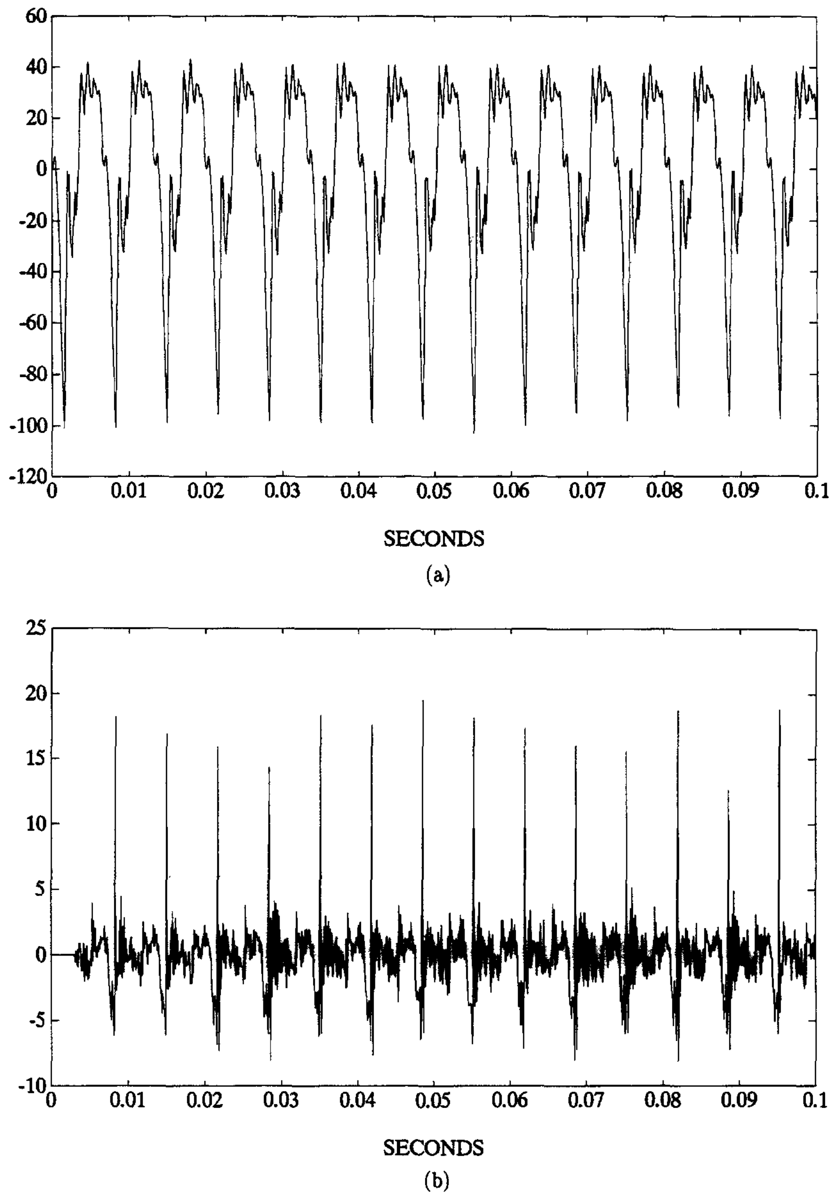
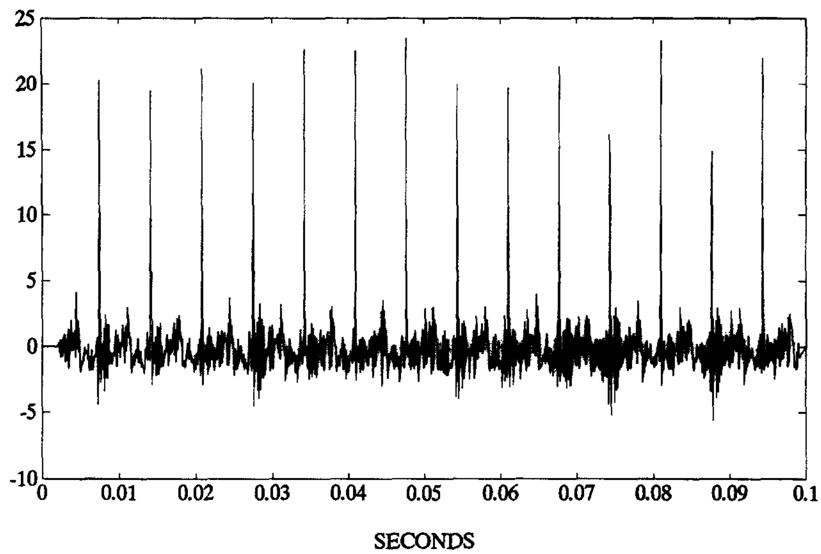


Fig. 4. (a) Speech data $x(k)$ of phoneme /ɑ:/ uttered by a man, (b) the predictive deconvolved data $e(k)$.

case. Hence, the cumulant order $M = 4$ was used in this example.

First of all, a minimum-phase PEF of order equal to 40 was obtained from $x(k)$ by Burg's algorithm [1, 12]. Then we processed $x(k)$ by the obtained PEF to get the deconvolved data $e(k)$ which is shown in Fig. 3(c) where the solid line depicts the true input $u(k)$ and the dotted line depicts $e(k)$. One can see, from Fig. 3(c), that each spike in $u(k)$ is associated with a wavelet in $e(k)$ which begins with two opposite peaks and gradually decays due to the remaining phase distortion of the source wavelet.



(c)

Fig. 4. (c) The improved deconvolved data $y(k)$ which is the output of the optimum allpass filter $\hat{H}_L(z)$ where $L = 9$.

Next, we processed $e(k)$ using the proposed allpass identification algorithm. Through the procedure (s2)–(s5) (without (s6)), the order L of the optimum $\hat{H}_L(z)$ turned out to be 2, and the output $y(k)$ (dotted line) of the optimum $\hat{H}_2(z)$ along with the true input $u(k)$ (solid line) are shown in Fig. 3(d). From Fig. 3(d), one can see that $y(k)$ approximates $u(k)$ very well except for a scale factor. In other-words, the phase distortion in $e(k)$ has been considerably removed by the proposed algorithm. Comparing Figs. 3(c) and 3(d), one can see that $y(k)$ is indeed a better estimate of $u(k)$ than $e(k)$.

3.2. *Experimental results with voiced speech data*

It is well known that the voiced speech signal can be modeled as the output of the vocal-tract filter driven by a non-Gaussian positive quasi-periodic pulse train [21]. In the experiment, the speech sound /ɑ:/ uttered by a man was filtered by a low-pass filter with cutoff frequency set to 3 kHz and then sampled by a 12 bit A/D converter with sampling frequency 10 kHz. The speech data of $x(k)$, which can be viewed as output measurements based on the convolutional model given by (10) where $h(k)$ is the impulse response of the vocal-tract filter, are shown in Fig. 4(a). Again, we followed the MP–AP-removal procedure mentioned in Section 1 to process $x(k)$. We preprocessed the speech data $x(k)$ by a minimum-phase PEF of order equal to 30 obtained by Burg's algorithm to get the 'second-order white' signal $e(k)$ shown in Fig. 4(b). Then the deconvolved signal $e(k)$ shown in Fig. 4(b) can be viewed as the output of an allpass system (a phase distortion system), and we processed $e(k)$ by the proposed allpass identification algorithm to remove the phase distortion in $e(k)$. With the cumulant order M set to 3 and the threshold $\eta = 1.9601$ (i.e., the confidence $(1 - \Delta) = 95\%$) in (s6), through the procedure (s2)–(s6), the order L of the optimum $\hat{H}_L(z)$ turned out to be 9. The output $y(k)$ of the optimum $\hat{H}_9(z)$ is shown in Fig. 4(c). One can see, from Figs. 4(b) and 4(c), that $y(k)$ approximates a positive quasi-periodic pulse train much better than $e(k)$ since the vocal-tract filter is nonminimum phase.

4. Conclusions

In this paper, we have presented a new higher-order cumulant based allpass system identification algorithm with only non-Gaussian output measurements. The proposed algorithm possesses five nice characteristics described in (C1)–(C5) in Section 2. It is applicable for any LTI phase distortion systems such as phase-distorted communication channels, phase distortion of source wavelet in predictive deconvolved seismic signals and phase distortion of vocal-tract filter in predictive deconvolved speech signals. We also showed two simulation examples and one set of experimental results with real voiced speech data to support the proposed algorithm. As mentioned in Section 1, the identification of an ARMA system $H(z)$ can be converted into the identification of an allpass system [8, 9] when MP–AP decomposition based methods are used, and the proposed allpass system identification algorithm can be applied no matter whether $H(z)$ includes allpass factors, whereas Chi and Kung's [2, 3] algorithm is not applicable when $H(z)$ includes allpass factors.

Appendix A. Proof of Theorem 1

By the assumption (A1), the Mth-order cumulant function of $u(k)$ is

$$C_{M,u}(k_1,k_2,\dots,k_{M-1}) = \gamma_M \delta(k_1)\delta(k_2)\delta(k_3)\cdots\delta(k_{M-1}), \tag{A.1}$$

where $\delta(k)$ is the discrete delta function. Let $H_{\mathrm{AP}}(z) = H(z)\hat{H}_p(z)$. Assume that

$$H_{\mathrm{AP}}(f) = H_{\mathrm{AP}}(z = \exp\{j2\pi f\}) = \exp\{j2\pi\phi(f)\} \tag{A.2}$$

with $\phi(0) = 0$ without loss of generality, where the phase $\phi(f)$ is a continuous odd function of f. Because the Mth-order ($M \geqslant 3$) cumulant function of Gaussian noise $w(k)$ is zero, the polyspectrum, $S_{M,y}(f_1,\dots,f_{M-1})$, of $y(k)$ is then given by [17, 18]

$$S_{M,y}(f_1,f_2,\dots,f_{M-1}) = S_{M,u}(f_1,f_2,\dots,f_{M-1}) \times \left\{\prod_{i=1}^{M-1} H_{\mathrm{AP}}(f_i)\right\} H^*_{\mathrm{AP}}(f_1+\cdots+f_{M-1}) = \gamma_M \exp\{j2\pi P(f_1,\dots,f_{M-1})\}, \tag{A.3}$$

where

$$P(f_1,\dots,f_{M-1}) = \phi(f_1)+\cdots+\phi(f_{M-1}) - \phi(f_1+\cdots+f_{M-1}) \tag{A.4}$$

is also a real continuous function of $f_1,\dots,f_{M-1}$. Then, we obtain from (A.3),

$$\left|C_{M,y}(0,0,\dots,0)\right| = \left|\gamma_M \int_{-1/2}^{1/2}\cdots\int_{-1/2}^{1/2} \exp\{j2\pi P(f_1,\dots,f_{M-1})\}\,\mathrm{d}f_1\,\mathrm{d}f_2\cdots\mathrm{d}f_{M-1}\right| \leqslant |\gamma_M| \int_{-1/2}^{1/2}\cdots\int_{-1/2}^{1/2} \left|\exp\{j2\pi P(f_1,\dots,f_{M-1})\}\right| \times \mathrm{d}f_1\,\mathrm{d}f_2\cdots\mathrm{d}f_{M-1} = |\gamma_M|. \tag{A.5}$$

It is trivial to see that if $H_{\mathrm{AP}}(z) = 1$, i.e., $\phi(f) = 0$, the equality in (A.5) holds. Next, we show that when the equality in (A.5) holds, $H_{\mathrm{AP}}(z) = 1$.

From (A.4) and (A.5), one can infer that when the equality holds,

$$\phi(f_1)+\cdots+\phi(f_{M-1}) - \phi(f_1+\cdots+f_{M-1}) = \theta + l \quad \forall (f_1,\dots,f_{M-1}) \tag{A.6}$$

where $-1/2 < \theta < 1/2$ is a constant and l is an integer. Letting $f_1 = f_2 = \cdots = f_{M-1} = 0$ in (A.6), we obtain $\theta + l = (M-2)\phi(0) = 0$ since $\phi(0) = 0$. This leads to $\theta = -l$, which implies $\theta = l = 0$ since $-1/2 < \theta < 1/2$. Therefore, (A.6) reduces to

$$\phi(f_1)+\cdots+\phi(f_{M-1}) = \phi(f_1+\cdots+f_{M-1}), \tag{A.7}$$

which implies $\phi(\cdot)$ is a linear operator or $\phi(f) = \alpha f$, or

$$H_{\mathrm{AP}}(z) = H(z)\hat{H}_p(z) = z^{\alpha}, \tag{A.8}$$

where α is a constant. Since $1/H(z) \in S_{\mathrm{AP}}(p)$ and $\hat{H}_p(z) \in S_{\mathrm{AP}}(p)$, $H_{\mathrm{AP}}(z)$ never takes the form z^{α_1} with $\alpha_1 \neq 0$. Thus, Eq. (A.8) holds only when $\alpha = 0$. Therefore, $H_{\mathrm{AP}}(z) = 1$ when the equality in (A.5) holds, or equivalently $H(z) = 1/\hat{H}_p(z)$. We thus have completed the proof.

Appendix B. Newton–Raphson type algorithm for the case $M = 3$ in (s1)

Let $\boldsymbol{\theta} = (a_1, \ldots, a_L)^T$. The Newton–Raphson type algorithm for finding a local maximum of the non-linear objective function $J(L)$ updates $\hat{\boldsymbol{\theta}}$ at the ith iteration by

$$\hat{\boldsymbol{\theta}}(i) = \hat{\boldsymbol{\theta}}(i-1) + \rho H_{i-1}^{-1} \boldsymbol{g}_{i-1}, \tag{B.1}$$

where $0 < \rho \leqslant 1$ is a constant, $\boldsymbol{g}_{i-1}$ and H_{i-1} denote the gradient and the Hessian matrix for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)$, respectively, as follows:

$$\boldsymbol{g}_{i-1} = \left. \frac{\partial J(L)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)}, \tag{B.2}$$

$$H_{i-1} = \left. \frac{\partial^2 J(L)}{\partial \boldsymbol{\theta}^2} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)}. \tag{B.3}$$

Next, we show how to compute $\boldsymbol{g}_{i-1}$ and H_{i-1}. For simplicity, assume that $M = 3$. We see, from (5), (6), (B.2) and (B.3), that $\boldsymbol{g}_{i-1}$ and H_{i-1} can be further expressed as

$$\begin{aligned} \boldsymbol{g}_{i-1} &= 2\hat{C}_{3,y}(0,0) \left. \frac{\partial \hat{C}_{3,y}(0,0)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)} \\ &= 2\left(\frac{1}{N} \sum_{k=0}^{N-1} [y(k)]^3 \right) \\ &\quad \times \left. \left(\frac{1}{N} \sum_{k=0}^{N-1} 3[y(k)]^2 \left(\frac{\partial y(k)}{\partial \boldsymbol{\theta}} \right) \right) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)} \end{aligned} \tag{B.4}$$

and

$$\begin{aligned} H_{i-1} &= 2 \left. \left(\frac{\partial \hat{C}_{3,y}(0,0)}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \hat{C}_{3,y}(0,0)}{\partial \boldsymbol{\theta}} \right)^T \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)} \\ &\quad + 2\hat{C}_{3,y}(0,0) \left. \frac{\partial \hat{C}^2_{3,y}(0,0)}{\partial \boldsymbol{\theta}^2} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)}. \\ &\cong 2\left(\frac{1}{N} \sum_{k=0}^{N-1} 3[y(k)]^2 \left(\frac{\partial y(k)}{\partial \boldsymbol{\theta}} \right) \right) \\ &\quad \times \left. \left(\frac{1}{N} \sum_{k=0}^{N-1} 3[y(k)]^2 \left(\frac{\partial y(k)}{\partial \boldsymbol{\theta}} \right)^T \right) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)} \\ &\quad + 2\left(\frac{1}{N} \sum_{k=0}^{N-1} [y(k)]^3 \right) \left(\frac{1}{N} \sum_{k=0}^{N-1} 6y(k) \right. \\ &\quad \times \left. \left. \left(\frac{\partial y(k)}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial y(k)}{\partial \boldsymbol{\theta}} \right)^T \right) \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(i-1)}, \end{aligned} \tag{B.5}$$

respectively, where the term including the second derivative of $y(k)$ with respect to $\boldsymbol{\theta}$ in (B.5) is neglected. In order to compute $\boldsymbol{g}_{i-1}$ and H_{i-1}, we need $y(k)$ and $\partial y(k)/\partial a_m$ for $m = 1, 2, \ldots, L$, and how to compute them is described in the following.

Taking the partial derivative of (7) with respect to a_m, $m = 1, 2, \ldots, L$, we find

$$\begin{aligned} \frac{\partial y(k)}{\partial a_m} &= -y(k+m) - \sum_{j=1}^{L} a_j \frac{\partial y(k+j)}{\partial a_m} \\ &\quad + x(k+L-m), \quad m = 1, 2, \ldots, L. \end{aligned} \tag{B.6}$$

At each iteration, updating $\hat{\boldsymbol{\theta}}$ by (B.1) with $\rho = 1$ normally leads to increase of $J(L)$ along with an anticausal stable $\hat{H}_L(z)$ which is needed for computing $y(k)$, the gradient and the approximate Hessian matrix; otherwise, a smaller ρ must be considered.

Acknowledgments

The research described in this paper was performed at the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China, and was supported by the National Science Council under Grant NSC81-0404-E-007-001 and the Telecommunication Laboratories under Grant TL-NSC-81-5201. The authors would also like to thank the anonymous referees for their valuable comments and suggestions.

References

[1] J.P. Burg, Maximum entropy spectral analysis, Ph.D. dissertation, Stanford University, May 1975.

[2] C.-Y. Chi and J.-Y. Kung, "A fast phase determination method by a single cumulant sample", *Proc. Internat. Signal Process. Workshop on Higher Order Statistics*, 10–12 July 1991, Chamrousse, France, pp. 83–86.

[3] C.-Y. Chi and J.-Y. Kung, "A phase determination method for nonminimum phase ARMA systems by a single cumulant sample", *IEEE Trans. Signal Processing*, Vol. 41, No. 2, February 1993, pp. 981–985.

[4] A.P. Clark, *Equalizers for Digital Modems*, Wiley, New York, 1985.

[5] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from higher-order moments", *IEEE Trans. Automat. Control*, Vol. 35, No. 1, January 1990, pp. 27–35.

[6] G.B. Giannakis, "Wavelet parameter and phase estimation using cumulant slices", *IEEE Trans. Geosci. Remote Sensing*, Vol. 27, No. 4, July 1989, pp. 452–455.

[7] G.B. Giannakis, "On the identifiability of non-Gaussian ARMA models using cumulants", *IEEE Trans. Automat. Control*, Vol. 35, No. 1, January 1990, pp. 18–26.

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

[9] G.B. Giannakis and J.M. Mendel, "Approximate realization and model reduction of nonminimum phase stochastic systems", *Proc. 25th IEEE Conf. on Decision and Control*, Athens, Greece, 1986, pp. 1079–1084.

[10] G.B. Giannakis and J.M. Mendel, "Identification of nonminimum phase systems using higher order statistics", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 37, March 1989, pp. 360–377.

[11] G.B. Giannakis and J.M. Mendel, "Cumulant-based order determination of non-Gaussian ARMA models", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 38, No. 8, August 1990, pp. 1141–1423.

[12] S.M. Kay, *Modern Spectral Estimation*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[13] J. Kormylo and J.M. Mendel, "Maximum-likelihood deconvolution", *IEEE Trans. Geosci. Remote Sensing*, Vol. GE-21, 1983, pp. 72–82.

[14] K.S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes", *Ann. Statist.*, Vol. 10, No. 4, 1982, pp. 1195–1208.

[15] J. Longbottom, A.T. Walden and R.E. White, "Principles and application of maximum kurtosis phase estimation", *Geophysical Prospecting*, Vol. 36, February 1988, pp. 115–138.

[16] J.M. Mendel, *Maximum-likelihood Deconvolution: A Journey into Model-based Signal Processing*, Springer, New York, 1990.

[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, March 1991, pp. 278-305.

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

[19] A.V. Oppenheim and R.W. Schafer, *Discrete-time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.

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

[21] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, Englewood Cliffs, NJ, 1978.

[22] E.A. Robinson, "Predictive decomposition of seismic traces", *Geophysics*, Vol. 22, 1957, pp. 767–778.

[23] M. Rosenblatt, *Stationary Sequences and Random Fields*, Birkhäuser, Boston, 1985.

[24] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)", *IEEE Trans. Inform. Theory*, Vol. 36, March 1990, pp. 312–321.

[25] M.T. Silvia and E.A. Robinson, *Geophysical Signal Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1980.

[26] A. Swami and J.M. Mendel, "AR identifiability using cumulants", *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, Co, June 1989, pp. 1–6.

[27] A. Swami and J.M. Mendel, "ARMA parameter estimation using only output cumulants", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 38, July 1990, pp. 1257–1265.

[28] J.K. Tugnait, "Recovering the poles from third-order cumulants of system output", *IEEE Trans. Automat. Control*, Vol. 34, October 1989, pp. 1085–1089.

[29] R.A. Wiggins, "Minimum entropy deconvolution", *Geoexploration*, Vol. 16, 1978, pp. 21–35.