Fourier Series Based Nonminimum Phase Model for Statistical Signal Processing

Chong-Yung Chi, *Senior Member, IEEE*

***Abstract*— In this paper, a parametric Fourier series based model (FSBM) for or as an approximation to an arbitrary nonminimum-phase linear time-invariant (LTI) system is proposed for statistical signal processing applications where a model for LTI systems is needed. Based on the FSBM, a (minimum-phase) linear prediction error (LPE) filter for amplitude estimation of the unknown LTI system together with the Cramér–Rao (CR) bounds is presented. Then, an iterative algorithm for obtaining the optimum LPE filter with finite data is presented that is also an approximate maximum-likelihood algorithm when data are Gaussian. Then three iterative algorithms using higher order statistics (HOS) with finite non-Gaussian data are presented to estimate parameters of the FSBM followed by some simulation results as well as some experimental results with real speech data to support the efficacy of the proposed algorithms using the FSBM. Finally, we draw some conclusions.**

I. Introduction

In many statistical signal processing areas such as signal modeling, power spectral and polyspectral estimation, system identification, deconvolution and equalization, a widely known problem is the identification and estimation of an unknown linear time-invariant (LTI) system $h(n)$ (which can be nonminimum-phase) driven by an unknown random signal $u(n)$ with only a given set of output measurements $x(n)$

$$x(n) = u(n) * h(n) = \sum_{k=-\infty}^{\infty} u(k)h(n-k). \tag{1}$$

The system $h(n)$ in, for example, adaptive signal processing is often assumed to be an LTI system over a time segment for convenience even if it is time varying. A parametric model for the LTI system $h(n)$ is usually used in the design of statistical signal processing algorithms because $h(n)$ can be easily characterized by its parameters, and thus, estimation of $h(n)$ becomes a parameter estimation problem that often leads to mathematically tractable solutions with predictable performance. The system function $H(z)$ is often modeled as a parametric rational function [1], [2], i.e.,

$$H(z) = \frac{B(z)}{A(z)} \tag{2}$$

Manuscript received August 25, 1997; revised January 25, 1999. This work was supported by the National Science Council under Grants NSC 86-2213-E-007-037 and NSC 87-2213-E007-026. The associate editor coordinating the review of this paper and approving it for publication was Dr. Jonathon A. Chambers.

The author is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C. (e-mail: cychi@ee.nthu.edu.tw).

Publisher Item Identifier S 1053-587X(99)05413-6.

where $A(z)$ and $B(z)$ are a pth-order polynomial and a qth-order polynomial of z^{-1}, respectively. The model $H(z)$ given by (2) is referred to as an ARMA(p,q) model, which includes an AR(p) model (when $B(z)=1$) and an MA(q) model (when $A(z)=1$) as special cases. Therefore, finding a rational model approximation to the system $h(n)$ from measurements is a parameter estimation problem. The performance of such as AR parameter estimation algorithms [1], [2] can be predicted from the power of linear prediction error (LPE) or from the variance of AR parameter estimates as well as the associated Cramér-Rao (CR) bounds.

Except for a scale factor, the ARMA model can also be expressed as [3]

$$H(z) = z^r \cdot C(z) \cdot D(z) \tag{3}$$

where

$$C(z) = \frac{\prod_{k=1}^{M_i}(1-a_k z^{-1})}{\prod_{k=1}^{N_i}(1-c_k z^{-1})} \tag{4}$$

and

$$D(z) = \frac{\prod_{k=1}^{M_o}(1-b_k z)}{\prod_{k=1}^{N_o}(1-d_k z)} \tag{5}$$

where $|a_k|$, $|b_k|$, $|c_k|$, and $|d_k|$ are all less than unity. Note that $c(n)$ [the inverse z-transform of $C(z)$] is a causal stable minimum phase system with $c(0)=1$, and $d(n)$ [the inverse z-transform of $D(z)$] is an anticausal stable maximum phase system with $d(0)=1$. The expression given by (3) is called minimum-phase (MN) maximum-phase (MX) decomposition [3]. Note that system amplitude and phase are simultaneously characterized by all parameters [coefficients of $A(z)$ and $B(z)$ or a_k, b_k, c_k and d_k] of the rational model, whereas only the former (system amplitude) can be easily inferred from poles (c_k and $1/d_k$) and zeros (a_k and $1/b_k$) of the rational model. On the other hand, stability of the rational model, which must be considered in the design of statistical signal processing algorithms, requires determination of $C(z)$ and $D(z)$ from the unknown $H(z)$ and thus may lead to considerable complexity, especially in the design of iterative signal processing algorithms that repeat the determination of $C(z)$ and $D(z)$ at each iteration.

Dianat and Raghuveer [4] proposed a parametric Fourier series based model (FSBM) for both magnitude and phase of non-Gaussian signals with the model parameters estimated from bispectra of data. Recently, Chien *et al.* [5] proposed a parametric cumulant-based method for estimating the phase of the unknown system $h(n)$ through allpass filtering of measurements $x(n)$ when $x(n)$ is non-Gaussian. Their method is applicable for both one- and two-dimensional (1-D and 2-D) systems. They used FSBM for an optimal allpass filter, which leads to a consistent estimate for the system phase $(\arg\{H(z=e^{j\omega})\})$ by maximizing a single absolute higher order cumulant of the allpass filter output. In this paper, an FSBM, which is always stable with a finite Fourier series model for log amplitude and a finite Fourier series model for phase, as an approximation to an arbitrary nonminimum-phase LTI system, is proposed for applications in the aforementioned statistical signal processing areas.

Section II presents the nonminimum-phase FSBM. Then, an LPE filter based on the proposed FSBM for amplitude estimation of the system is presented in Section III. Section IV presents estimation of FSBM (amplitude and phase) parameters, and then, some simulation results as well as some experimental results with real speech data are presented in Section V. Finally, we draw some conclusions.

II. Nonminimum-Phase FSBM

Assume that $h(n)$ is a real nonminimum-phase LTI system with the frequency response $H(\omega)=H(z=e^{j\omega})=H^*(-\omega)$ defined as the magnitude (MG)-phase (PS) decomposition

$$H(\omega)=H_{\rm MG}(\omega)\cdot H_{\rm PS}(\omega) \tag{6}$$

where

$$H_{\rm MG}(\omega)=|H(\omega)|=\exp\left\{\sum_{i=1}^{p}\alpha_i\cos(i\omega)\right\} \tag{7}$$

and

$$H_{\rm PS}(\omega)=\exp\left\{j\sum_{i=1}^{q}\beta_i\sin(i\omega)\right\} \tag{8}$$

include the amplitude and phase information of $H(\omega)$, respectively, where $p\geq 1$ and $q\geq 1$ are integers, and α_i and β_i are real. Four worthy remarks regarding the distinctions of the proposed nonminimum-phase FSBM(p,q) given by (6) and the rational model (i.e., AR, MA and ARMA models) and the advantages of the former over the latter are given as follows.

R1) System magnitude and phase are simultaneously characterized by poles (for peaks in magnitude response) and zeros (for notches in magnitude response) for the ARMA model, whereas they are characterized by amplitude parameters (α_i) and phase parameters (β_i), respectively, for the FSBM. Moreover, the magnitude response $H_{\rm MG}(\omega)$ given by (7) is also a cascade of p comb (peaking or notching) filters with amplitude responses $\exp\{\alpha_i\cos(i\omega)\}$, $i=1,2,\ldots,p$, where each of the peak bandwidth and the notch bandwidth are smaller (larger) as $|\alpha_i|$ is larger (smaller). Therefore, when the FSBM is used for approximation to a comb filter with l peaks and l notches, a single nonzero parameter α_l may be sufficient. If the comb filter is further required to be minimum phase, the FSBM must take the form of (14) (see below) $(\beta_l=-\alpha_l)$.

R2) The FSBM(p,q) given by (6) is always a stable IIR system no matter whether it is causal or noncausal because it is a periodic continuous function of ω [3] with period equal to 2π. Therefore, when the system to be designed is a noncausal stable system such as the noncausal inverse filter $1/H(z)$ [when $h(n)$ is not minimum phase] in blind deconvolution and channel equalization, the proposed FSBM is more suitable than the ARMA model for more efficient algorithm design and simpler signal processing procedure because the stability issue is never existent for the former.

R3) The complex cepstrum $\tilde{h}(n)$ (inverse Fourier transform of $\ln[H(\omega)]$), which has been used in speech processing [6]–[8], biomedical signal processing [9], and seismic signal processing [10], [11], associated with the ARMA model given by (3) requires the compensation for the time delay term z^r (i.e., $r=0$) in advance and can be shown to be [3]

$$\tilde{h}(n)=\begin{cases}0, & n=0\\ -\sum_{k=1}^{M_i}a_k^n/n+\sum_{k=1}^{N_i}c_k^n/n, & n>0\\ \sum_{k=1}^{M_o}b_k^{-n}/n-\sum_{k=1}^{N_o}d_k^{-n}/n, & n<0\end{cases}. \tag{9}$$

Although we can use numerical polynomial rooting algorithms to find all the poles c_k and $1/d_k$ and zeros a_k and $1/b_k$ of $H(z)$, the obtained a_k, $1/b_k$, c_k, and $1/d_k$ may not be very reliable when the system order is large. On the other hand, the complex cepstrum associated with the FSBM(p,q) given by (6) can be easily shown to be

$$\tilde{h}(n)=\begin{cases}\frac{1}{2}(\alpha_n-\beta_n), & 1\leq n\leq\max\{p,q\}\\ \frac{1}{2}(\alpha_{-n}+\beta_{-n}), & -\max\{p,q\}\leq n\leq -1\\ 0, & \text{otherwise}\end{cases} \tag{10}$$

where $\alpha_i=0$ for $i>p$, and $\beta_i=0$ for $i>q$. The finite complex cepstrum $\tilde{h}(n)$ of FSBM(p,q) given by (10) corresponds to a windowing approximation to that of an arbitrary LTI system (i.e., the true system complex cepstrum multiplied by a rectangular window).

R4) For any arbitrary stable LTI system $H(\omega)$ whose amplitude and phase are continuous (without zeros on the unit circle), the $\log$ function of the FSBM(p,q) given by (6) with

$$\alpha_i=\frac{1}{\pi}\int_{-\pi}^{\pi}[\ln|H(\omega)|]\cos(i\omega)\,d\omega \tag{11}$$

$$\beta_i=\frac{1}{\pi}\int_{-\pi}^{\pi}[\arg\{H(\omega)\}]\sin(i\omega)\,d\omega \tag{12}$$

converges uniformly to $\ln[H(\omega)]$ as $p\to\infty$ and $q\to\infty$ (by the properties of Fourier series). However, when the LTI system has a pair of zeros at $e^{\pm j\omega_0}$ (on the unit circle), $\ln|H(\pm\omega_0)|=-\infty$ and $j\arg\{H(\omega)\}$ have a pair of discontinuities of magnitude equal to π at $\omega=\pm\omega_0$. In this case, the larger p, the better $\ln[H_{\rm MG}(\omega)]$ approximates to $\ln|H(\omega)|$, and the larger q, the better $\ln[H_{\rm PS}(\omega)]$ approximates to $j\arg\{H(\omega)\}$.

Minimum-phase (MP) allpass (AP) decomposition and MN-MX decomposition given by (3) for the ARMA model have been used in deconvolution [12]–[14], channel equalization [15]–[19], and system identification [19]–[24]. These two decompositions for the FSBM are as follows.

A) *MP-AP decomposition*:

$$H(\omega) = H_{\rm MP}(\omega) \cdot H_{\rm AP}(\omega) \tag{13}$$

where

$$\begin{aligned} H_{\rm MP}(\omega) &= \exp\left\{\sum_{i=1}^{p} \alpha_i \cos(i\omega) - j\sum_{i=1}^{p} \alpha_i \sin(i\omega)\right\} \\ &= \exp\left\{\sum_{i=1}^{p} \alpha_i e^{-j\omega i}\right\} \end{aligned} \tag{14}$$

$$H_{\rm AP}(\omega) = \exp\left\{j \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i)\sin(i\omega)\right\} \tag{15}$$

where $\alpha_i = 0$ for $i > p$, and $\beta_i = 0$ for $i > q$.

B) *MN-MX decomposition*:

$$H(\omega) = H_{\rm MN}(\omega) \cdot H_{\rm MX}(\omega) \tag{16}$$

where

$$\begin{aligned} H_{\rm MN}(\omega) &= \exp\left\{\sum_{i=1}^{\max\{p,q\}} \frac{1}{2}(\alpha_i - \beta_i)\cos(i\omega) - j\sum_{i=1}^{\max\{p,q\}} \frac{1}{2}(\alpha_i - \beta_i)\sin(i\omega)\right\} \\ &= \exp\left\{\sum_{i=1}^{\max\{p,q\}} \frac{1}{2}(\alpha_i - \beta_i) e^{-j\omega i}\right\} \end{aligned} \tag{17}$$

$$\begin{aligned} H_{\rm MX}(\omega) &= \exp\left\{\sum_{i=1}^{\max\{p,q\}} \frac{1}{2}(\alpha_i + \beta_i)\cos(i\omega) + j\sum_{i=1}^{\max\{p,q\}} \frac{1}{2}(\alpha_i + \beta_i)\sin(i\omega)\right\} \\ &= \exp\left\{\sum_{i=1}^{\max\{p,q\}} \frac{1}{2}(\alpha_i + \beta_i) e^{j\omega i}\right\} \end{aligned} \tag{18}$$

where $\alpha_i = 0$ for $i > p$, and $\beta_i = 0$ for $i > q$.

It has been shown in [3] that when the complex cepstrum $\tilde{g}(n)$ of a signal $g(n)$ is causal (anticausal), the signal $g(n)$ is causal minimum phase (anticausal maximum phase) with $g(0) = \exp\{\tilde{g}(0)\}$. Because $\tilde{h}_{\rm MP}(n)$ [$= \alpha_n$ for $n > 0$ by (14)] and $\tilde{h}_{\rm MN}(n)$ [$= \tilde{h}(n)$ for $n > 0$ by (17) and (10)] are causal with $\tilde{h}_{\rm MP}(0) = \tilde{h}_{\rm MN}(0) = 0$, $H_{\rm MP}(\omega)$ and $H_{\rm MN}(\omega)$ [counterpart of $C(z)$] are causal minimum phase with $h_{\rm MP}(0) = h_{\rm MN}(0) = 1$. Because $\tilde{h}_{\rm MX}(n)$ [$= \tilde{h}(n)$ for $n < 0$ by (18)] is anticausal with $\tilde{h}_{\rm MX}(0) = 0$, $H_{\rm MX}(\omega)$ [counterpart of $D(z)$] is anticausal maximum phase with $h_{\rm MX}(0) = 1$. Moreover, based on the three decompositions above, FSBM(p,q) can be simplified for some special LTI systems as summarized in the following remark:

R5) When $\alpha_i = 0$ for all $1 \le i \le p$, the FSBM is an allpass system [see (7) and (8)]; when $\beta_i = 0$ for all $1 \le i \le q$, the FSBM is a zero phase system [see (7) and (8)]; when $p = q$ and $\beta_i = -\alpha_i$ for all $1 \le i \le p$, the FSBM is a causal minimum-phase system [see (14) and (15)]; when $p = q$ and $\beta_i = \alpha_i$ for all $1 \le i \le p$, the FSBM is an anticausal maximum-phase system [see (17) and (18)].

Next, let us present how to estimate the amplitude parameters of FSBM(p,q) by linear prediction and phase parameters using higher order statistics with only measurements $x(n)$.

III. FSBM For LPE Filters

Let us briefly review the conventional LPE filter for ease of later need for the presentation of LPE filters using the FSBM.

A. Conventional LPE Filters

Assume that $x(n)$ is a real stationary random process modeled by (1), where $h(n)$ is a stable LTI system driven by a white noise $u(n)$ with zero mean and variance σ^2. The conventional pth-order LPE filter [1], [2]

$$A_p(z) = 1 + \sum_{i=1}^{p} a_i z^{-i} \tag{19}$$

(a causal FIR filter) processes $x(n)$ such that the prediction error

$$e(n) = x(n) * a_n = x(n) + \sum_{k=1}^{p} a_k x(n-k) \tag{20}$$

has minimum variance or average power $E[e^2(n)]$. Note that we have used the same notation a_k in the LPE filter $A_p(z)$ and in $C(z)$ [see (3) and (4)] of the ARMA model without confusion. The optimum LPE filter $\hat{A}_p(z)$ is minimum-phase and can be solved from the orthogonality principle [1], [2]

$$E[e(n)x(n-k)] = 0, \quad k = 1, 2, \ldots, p \tag{21}$$

which also forms a set of symmetric Toeplitz linear equations of $\mathbf{a}_p = (a_1, \ldots, a_p)^T$ (which are also called normal equations).

A well-known fact in estimation theory is that for any unbiased estimates $\hat{\mathbf{a}}_p$ and $\widehat{\sigma^2}$ with given finite data $\mathbf{x} = (x(0), x(1), \ldots, x(N-1))^T$, their covariance matrix is lower bounded by the CR bounds. When $x(n)$ is an AR(p) Gaussian process, the approximate probability density function

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\left(\frac{N}{2\sigma^2}\right)\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I(\omega)}{|H(\omega)|^2}\, d\omega\right\} \tag{22}$$

where $I(\omega)$ is the periodogram

$$I(\omega) = \frac{1}{N}|X(\omega)|^2 = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)\exp\{-j\omega n\}\right|^2 \tag{23}$$

was used in [1] to derive the CR bounds

$$\mathbf{C}_{\hat{\mathbf{a}}_p\widehat{\sigma^2}} \geq \frac{\sigma^2}{N}\begin{bmatrix}\mathbf{R}_{xx}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 2\sigma^2\end{bmatrix} \tag{24}$$

where $\mathbf{R}_{xx} = E[\mathbf{x}_p(n)\mathbf{x}_p^T(n)]$, in which $\mathbf{x}_p(n) = (x(n), \ldots, x(n-p+1))^T$, and $\mathbf{0}$ is a $p \times 1$ zero vector. Next, let us present the LPE filter using FSBM.

B. LPE Filters Using FSBM

Let the pth-order LPE filter $v_p(n)$ be a causal minimum-phase IIR filter with $v_p(0) = 1$ and

$$\begin{aligned} V_p(\omega) &= \exp\left\{\sum_{i=1}^{p} a_i \cos(i\omega) - j\sum_{i=1}^{p} a_i \sin(i\omega)\right\} \\ &= \exp\left\{\sum_{i=1}^{p} a_i e^{-j\omega i}\right\} \end{aligned} \tag{25}$$

and thus, the prediction error is given by

$$e(n) = x(n) * v_p(n) = x(n) + \sum_{k=1}^{\infty} v_p(k)x(n-k). \tag{26}$$

Note that for notational simplicity, we have used the same notations a_i for parameters of both the proposed LPE filter $V_p(\omega)$ and the conventional LPE filter $A_p(z)$ without confusion. The optimum LPE filter $\hat{V}_p(\omega)$ is described in the following theorem.

Theorem 1: Assume that $x(n)$ is modeled as (1) where $u(n)$ is white with zero mean and variance σ^2, and $H(\omega)$ is an FSBM($p^\star, q$) as given by (6). Let $e(n)$ be the prediction error given by (26). Then, for any $p \geq p^\star$, the optimum LPE filter $\hat{V}_p(\omega) = 1/H_{\rm MP}(\omega)$ with $\min\{E[e^2(n)]\} = E[u^2(n)] = \sigma^2$. See Appendix A for the proof of Theorem 1.

Next, based on Theorem 1, let us present the orthogonality principle for solving the optimum LPE filter $\hat{V}_p(\omega)$ that needs the expression

$$\begin{aligned} \frac{\partial e(n)}{\partial a_k} &= \frac{\partial}{\partial a_k}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} V_p(\omega)\cdot X(\omega)\cdot e^{j\omega n}\, d\omega\right\} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial V_p(\omega)}{\partial a_k}\cdot X(\omega)\cdot e^{j\omega n}\, d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} V_p(\omega)\cdot X(\omega)\cdot e^{j\omega(n-k)}\, d\omega \\ &= e(n-k) \quad \text{[since (25)]} \end{aligned} \tag{27}$$

for $\partial e(n)/\partial a_k$. Taking partial derivative of $E[e^2(n)]$ with respect to a_k gives rise to

$$\frac{\partial}{\partial a_k}E[e^2(n)] = 2E\left\{e(n)\frac{\partial e(n)}{\partial a_k}\right\} = 2r_{ee}(k) \quad \text{[since (27)]} \tag{28}$$

where $r_{ee}(k)$ is the autocorrelation function of $e(n)$. The optimum prediction error $e(n)$ must satisfy $\partial E[e^2(n)]/\partial a_k = 0$ that leads to the orthogonality principle

$$E[e(n)e(n-k)] = r_{ee}(k) = 0, \quad k = 1, 2, \ldots, p \tag{29}$$

which, however, form a set of nonlinear equations rather than a set of symmetric Toeplitz linear equations of $\mathbf{a}_p$, as formed by (21). The distinctions between the proposed LPE filter $\hat{V}_p(\omega)$ and the conventional LPE filter $\hat{A}_p(z)$ are summarized in the following remark.

R6) Both $\hat{V}_p(\omega)$, which is an IIR filter [see (26)], and $\hat{A}_p(z)$, which is an FIR filter, are Wiener filters with minimum phase. However, the optimum prediction error $e(n)$ is orthogonal to $\{e(n-1), e(n-2), \ldots, e(n-p)\}$ for the former and orthogonal to $\{x(n-1), x(n-2), \ldots, x(n-p)\}$ for the latter. Nevertheless, $e(n)$ will be a white process as p is sufficiently large, which implies that $\hat{V}_p(\omega) = \hat{A}_p(\omega)$ (identical whitening filter) for $p = \infty$ as long as $x(n)$ is a wide-sense stationary linear process.

Next, let us present the CR bounds associated with $\boldsymbol{\alpha}_p = (\alpha_1, \ldots, \alpha_p)^T$ and σ^2 when $x(n)$ is an FSBM(p, q) Gaussian process with $H(\omega)$ given by (6). The approximate probability density function given by (22) can also be used to obtain the CR bounds as

$$\mathbf{C}_{\hat{\boldsymbol{\alpha}}_p,\widehat{\sigma^2}} \geq \frac{1}{N}\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & 2\sigma^4\end{bmatrix} \tag{30}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix. The proof for the CR bounds given by (30) is given in Appendix B. Note that the CR bounds associated with AR parameters [see (24)] depend on correlations of $x(n)$, whereas those associated with α_i are uniform and independent of correlations of $x(n)$. The CR bound associated with σ^2 is the same for both FSBM and AR model.

Based on Theorem 1, an iterative algorithm is used to estimate or find an approximation to $H_{\rm MP}(\omega)$ with finite data $x(0), x(1), \ldots, x(N-1)$ as follows.

Algorithm 1—Estimation of $H_{\rm MP}(\omega)$:

S1) Search for the minimum of the objective function

$$J(\mathbf{a}_p) = \frac{1}{N}\sum_{n=0}^{N-1} e^2(n) \tag{31}$$

and the associated optimum $\hat{\mathbf{a}}_p$ by a gradient-type iterative optimization algorithm (such as the well-known Fletcher–Powell algorithm [26]).

S2) Obtain the estimates $\hat{H}_{\rm MP}(\omega)$ and $\widehat{\sigma^2}$ by

$$\hat{H}_{\rm MP}(\omega) = 1/\hat{V}_p(\omega) \quad (\text{or } \hat{\boldsymbol{\alpha}}_p = -\hat{\mathbf{a}}_p) \tag{32}$$

$$\widehat{\sigma^2} = J(\hat{\mathbf{a}}_p). \tag{33}$$

Two remarks regarding Algorithm 1 are given as follows.

R7) As with other nonlinear optimization algorithms, only a local minimum of $J(\mathbf{a}_p)$ can be found. The optimum prediction error $e(n) = x(n) * \hat{v}_p(n) \simeq u(n) * h_{\rm AP}(n)$ corresponds to amplitude equalized data by (32). The iterative Fletcher–Powell algorithm, which is summarized in Appendix C, needs the gradient of $J(\mathbf{a}_p)$ with respect to a_k [see (C.2)], which, by (27) and (31), can be easily seen to be

$$\frac{\partial J(\mathbf{a}_p)}{\partial a_k} = 2\hat{r}_{ee}(k) = \frac{2}{N}\sum_{n=0}^{N-1} e(n)e(n-k) \tag{34}$$

where $\hat{r}_{ee}(k)$ is the sample correlation function of $e(n)$ as defined in (34).

R8) It can be easily shown, from (22) and (23), that when $x(n)$ is Gaussian, maximizing $p(\mathbf{x})$ for finding an approximate maximum-likelihood (AML) estimate $\hat{\boldsymbol{\alpha}}_{p,\mathrm{AML}}$ is equivalent to minimizing

$$\begin{aligned} S(\boldsymbol{\alpha}_p) &= \frac{N}{2\pi}\int_{-\pi}^{\pi}\frac{I(\omega)}{|H(\omega)|^2}\,d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}|\,X(\omega)V_p(\omega)\,|^2\ d\omega \\ &\simeq \sum_{n=0}^{N-1}|e(n)|^2 \quad \text{(by Parseval's theorem)} \end{aligned} \tag{35}$$

when $H(\omega)$ is an FSBM(p,q) as given by (6). An AML estimate $\widehat{\sigma^2}_{\mathrm{AML}}$ can be obtained by letting $\partial \ln p(\mathbf{x})/\partial\sigma^2 = 0$ that leads to

$$\widehat{\sigma^2}_{\mathrm{AML}} = \frac{1}{N}\min\{S(\boldsymbol{\alpha}_p)\}. \tag{36}$$

It can be easily seen, from (31), (32), (33), (35) and (36), that both $\hat{\boldsymbol{\alpha}}_p = -\hat{\mathbf{a}}_p$ and $\widehat{\sigma^2}$ are also AML estimates when $x(n)$ is Gaussian since $J(\mathbf{a}_p) \to S(\boldsymbol{\alpha}_p)/N$ for large N.

Next, let us present how to estimate the order p of the FSBM for LPE filters. The well-known Akaike information criteria (AIC) [1], [2], [25] given by

$$\mathrm{AIC}(k) = -2\ln p\big(\mathbf{x}; \hat{\boldsymbol{\alpha}}_{k,\mathrm{AML}}, \widehat{\sigma^2}_{\mathrm{AML}}\big) + 2k \tag{37}$$

can be used for the estimation of p, assuming $N \gg p$. The optimum estimate, which is denoted $\hat{p}$, of p is the one such that AIC(k) is minimum for $k = \hat{p}$. Let

$$\varepsilon(k) = \widehat{\sigma^2} \to \widehat{\sigma^2}_{\mathrm{AML}} \quad \text{(for large } N) \quad \text{[by R8)]} \tag{38}$$

denote the minimum prediction error power of the LPE filter of order equal to k [see (31) and (33)]. Substituting (35) and (36) to (22) yields

$$\ln p\big(\mathbf{x}; \hat{\boldsymbol{\alpha}}_{k,\mathrm{AML}}, \widehat{\sigma^2}_{\mathrm{AML}}\big) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\varepsilon(k) - \frac{N}{2} \quad \text{(for large } N) \tag{39}$$

where we have used the approximation of $\widehat{\sigma^2}_{\mathrm{AML}}$ given by (38) in the derivation of (39). Substituting (39) back to (37) and, meanwhile, ignoring the constant terms, we obtain

$$\mathrm{AIC}(k) = N\ln\varepsilon(k) + 2k. \tag{40}$$

Remark that the same AIC(k) given by (40) [1], [2], [25] can be used for model-order estimation, regardless of whether the model used for LPE filters is FIR model or FSBM. Therefore, Algorithm 1 can be used for the design of the LPE filter modeled as an FSBM as well as the model order estimation using AIC(k) given by (40).

IV. Estimation of FSBM Parameters

In this section, further with the assumption that $u(n)$ is non-Gaussian with nonzero Mth-order (≥ 3) cumulant [and thus $x(n)$ is also non-Gaussian], three iterative algorithms are to be presented for the estimation of parameters of the FSBM(p,q) given by (6).

The first two algorithms estimate the system amplitude using Algorithm 1 and system phase using the Chien *et al.* phase estimation algorithm [5] that maximizes a single absolute Mth-order (≥ 3) sample cumulant, which is denoted $|\hat{C}_M\{y(n)\}|$, of the phase equalized (allpass filtered) data

$$y(n) = x(n) * g_{\mathrm{AP}}(n) \tag{41}$$

where $g_{\mathrm{AP}}(n)$ is a qth-order allpass FSBM

$$G_{\mathrm{AP}}(\omega) = \exp\left\{ j\sum_{i=1}^{q}\gamma_i \sin(i\omega)\right\}. \tag{42}$$

It has been shown in [5] that the optimum $\hat{G}_{\mathrm{AP}}(\omega)$ turns out to be a phase equalizer except for an unknown time delay τ, i.e.,

$$\arg\{\hat{G}_{\mathrm{AP}}(\omega)\} = -\arg\{H(\omega)\} + \omega\tau \tag{43}$$

or the overall system $\hat{g}_{\mathrm{AP}}(n) * h(n)$ becomes a linear-phase system with the same magnitude response as $|H(\omega)|$. Because $|\hat{C}_M\{y(n)\}|$ is a highly nonlinear function of γ_i, we can use gradient-type iterative algorithms (such as the Fletcher–Powell algorithm) for finding the optimum γ_i. For instance

$$\hat{C}_3\{y(n)\} = \frac{1}{N}\sum_{n=0}^{N-1} y^3(n) \tag{44}$$

$$\hat{C}_4\{y(n)\} = \frac{1}{N}\sum_{n=0}^{N-1} y^4(n) - 3\left(\frac{1}{N}\sum_{n=0}^{N-1} y^2(n)\right)^2. \tag{45}$$

Gradient-type iterative algorithms require the computation of the gradient of $|\hat{C}_M\{y(n)\}|$ with respect to γ_i. It has also been proven in [5] that

$$\frac{\partial y(n)}{\partial \gamma_i} = \frac{1}{2}\{y(n+i) - y(n-i)\} \tag{46}$$

which is needed to compute the gradient of $|\hat{C}_M\{y(n)\}|$ with respect to γ_i (see [5] for details). However, when the order q is unknown, q must be estimated from measurements prior to the estimation of γ_i. An approach for the estimation of q is based on the cumulant variation rate (CVR) defined as

$$\mathrm{CVR}(k) = \frac{|\eta(k) - \eta(k-1)|}{|\eta(k-1)|} \times 100\% \tag{47}$$

where $\eta(k)$ is the maximum of $|\hat{C}_M\{y(n)\}|$ associated with the kth-order allpass FSBM $G_{\mathrm{AP}}(\omega)$ given by (42). The optimum estimate, which is denoted $\hat{q}$, is the smallest integer such that CVR(k) is below a threshold for all $k > \hat{q}$. Next, let us present the three algorithms for estimating the parameters of the FSBM(p,q) with (p,q) known in advance.

Algorithm 2—Separate Amplitude Estimation and Phase Estimation Based on the MG-PS Decomposition:

S1) Estimate $H_{\rm MP}(\omega)$ and σ^2 using Algorithm 1 to obtain $\hat{\alpha}_i, i = 1, 2, \ldots, p$. Thus, $\hat{H}_{\rm MG}(\omega)$ can be obtained from $\hat{\alpha}_i$.

S2) Find the optimum allpass FSBM $G_{\rm AP}(\omega)$ given by (42) using a gradient-type iterative algorithm such that $|\hat{C}_M\{y(n)\}|$ is maximum, where $y(n) = x(n) * g_{\rm AP}(n)$. Then, obtain $\hat{\beta}_i = -\hat{\gamma}_i$, $i = 1, 2, \ldots, q$, and $\hat{H}_{\rm PS}(\omega)$ since $\hat{G}_{\rm AP}(\omega) \simeq 1/H_{\rm PS}(\omega)$ [see (8)].

Algorithm 3—Amplitude Estimation Prior to Phase Estimation Based on the MP-AP Decomposition:

S1) Estimate $H_{\rm MP}(\omega)$ and σ^2 to obtain $\hat{\alpha}_i$, $i = 1, 2, \ldots, p$, and obtain the optimum prediction error $e(n) \simeq u(n) * h_{\rm AP}(n)$ [see R7)] using Algorithm 1.

S2) Find the optimum allpass FSBM $G_{\rm AP}(\omega)$ given by (42) using a gradient-type iterative algorithm such that $|\hat{C}_M\{y(n)\}|$ is maximum, where $y(n) = e(n) * g_{\rm AP}(n)$, and the order of the allpass FSBM $g_{\rm AP}(n)$ is equal to $\max\{p, q\}$. Then, obtain $\hat{\gamma}_i = -\hat{\alpha}_i - \hat{\beta}_i$, $i = 1, 2, \ldots, \max\{p, q\}$, and $\hat{H}_{\rm AP}(\omega)$ since $\hat{G}_{\rm AP}(\omega) \simeq 1/H_{\rm AP}(\omega)$ [see (15)].

The last algorithm (Algorithm 4) estimates the system amplitude and phase simultaneously using inverse filter criteria [27]–[30]. Chi and Wu [27] proposed a family of inverse filter criteria that includes Tugnait's criteria [28], Wiggins' criterion [29], and Shalvi and Weinstein's criterion [30] as special cases. The inverse filter $h_{\rm INV}(n)$ is estimated by maximizing

$$J_{r,m} = \frac{|\hat{C}_m\{\hat{u}(n)\}|^r}{|\hat{C}_r\{\hat{u}(n)\}|^m} \tag{48}$$

where r is even and $m > r$, and

$$\hat{u}(n) = x(n) * h_{\rm INV}(n). \tag{49}$$

It has been shown in [27] that $\hat{u}(n) = bu(n - \tau)$ when $h(n)$ is an arbitrary stable LTI system, where b is a scale factor, and τ is an unknown time delay. Next, let us present Algorithm 4.

Algorithm 4—Simultaneous Amplitude and Phase Estimation Based on the MG-PS Decomposition:

S1) Set integer $r \geq 2$ (even) and integer $m > r$. Let $H_{\rm INV}(\omega) = 1/H(\omega)$, where $H(\omega)$ is the FSBM(p, q) given by (6).

S2) Find the optimum $H_{\rm INV}(\omega)$ (i.e., α_i and β_i) using a gradient-type iterative algorithm such that $J_{r,m}$ is maximum. Then, σ^2 is estimated as the sample variance of the obtained optimum inverse filter output $\hat{u}(n)$.

Following the same procedure as when we proved $\partial e(n)/\partial a_k$ given by (27), it can be easily shown that

$$\frac{\partial \hat{u}(n)}{\partial \alpha_i} = -\frac{1}{2}\{\hat{u}(n+i) + \hat{u}(n-i)\} \tag{50}$$

$$\frac{\partial \hat{u}(n)}{\partial \beta_i} = -\frac{1}{2}\{\hat{u}(n+i) - \hat{u}(n-i)\} \tag{51}$$

which are needed for computing the gradient of $J_{r,m}$ with respect to α_i and β_i, respectively, required by the iterative gradient type algorithm in S2).

Next, let us discuss the computational complexity of the proposed three algorithms. The computation of $e(n)$ [see (26)], $y(n)$ [see (41)] and $\hat{u}(n)$ [see (49)] can be efficiently performed using FFT because the FSBM model is a parametric model in the frequency domain. All the gradient-type iterative optimization algorithms used in the three algorithms have a computationally efficient parallel structure (FIR filter banks with only two nonzero coefficients $1/2$ or $-1/2$) in computing the partial derivative of the allpass filter output with respect to γ_i as given by (46) for Algorithms 2 and 3 and the partial derivative of the inverse filter output with respect to α_i and β_i as given by (50) and (51) for Algorithm 4. However, the gradient computation associated with Algorithm 1 [see (34)] does not need any further processing to the prediction error $e(n)$. Let us conclude this section with the following remark discussing the use of the three algorithms when some prior information about the FSBM is known ahead of time, but the order (p, q) is not known.

R9) When the order (p, q) of the FSBM is unknown, p can be estimated using AIC(k) given by (40), and q can be estimated using CVR(k) given by (47). On the other hand, when some prior information is known *a priori*, the proposed three algorithms can be simplified. For instance, when the unknown system is known to be minimum-phase, S2) in Algorithms 2 and 3 is redundant since $\beta_i = -\alpha_i$ and $p = q$ for this case [see R5)] and the FSBM $H(\omega)$ used by Algorithm 4 can be reduced to $H_{\rm MP}(\omega)$. When the unknown system is an allpass system, the S1) in the former is redundant and the FSBM $H(\omega)$ used by the latter can be reduced to $H_{\rm PS}(\omega)$ since $\alpha_i = 0$ for all i.

V. Simulation Results and Experimental Results

In this section, let us show some simulation results (Example 1) for the CR bounds as well as the order estimation of the FSBM and performance evaluation of the proposed algorithms followed by some simulation results (Example 2) for seismic deconvolution using the proposed algorithms. Then, some experimental results for speech deconvolution with real speech data (Example 3) are presented to support the efficacy of the proposed algorithms. Next, let us turn to Example 1.

Example 1: This example includes two parts. Part 1 presents some simulation results for the CR bounds associated with amplitude parameters α_i, $i = 1, 2, \ldots, p$ of the FSBM(p, q), the estimation of p, and the performance evaluation of Algorithm 1. Part 2 presents some simulation results for the estimation of q as well as the performance evaluation of Algorithms 2–4. Next, let us present Parts 1 and 2, respectively.

In the simulation of Part 1, the driving input $u(n)$ was assumed to be a zero-mean white Gaussian random sequence and a nonminimum-phase FSBM$(3, 4)$ given by

$$H_{\rm MG}(\omega) = \exp\{1.1535\cos(\omega) - 0.4054\cos(2\omega) - 0.3138\cos(3\omega)\} \tag{52}$$

$$H_{\rm PS}(\omega) = \exp\{j[-0.9112\sin(\omega) + 0.5234\sin(2\omega) + 0.5290\sin(3\omega) + 0.2348\sin(4\omega)]\} \tag{53}$$

TABLE I
SIMULATION RESULTS OF PART 1 OF EXAMPLE 1. AIC(k), $k = 1 \sim 6$ AND $N = 512,\ 1024,\ 2048$ AND 4096

N \ p	1	2	3	4	5	6
512	120.0040	49.5771	3.8451	4.7642	5.8875	6.5956
1024	255.9041	107.3531	4.6679	5.5394	6.1014	7.1759
2048	507.7294	211.0764	9.6831	10.2516	11.0237	12.0000
4096	961.2921	365.2427	–24.4231	–23.2485	–22.0741	–20.4869

TABLE II
SIMULATION RESULTS OF PART 1 OF EXAMPLE 1. AVERAGES AND RMS ERRORS OF THIRTY INDEPENDENT AMPLITUDE PARAMETER ESTIMATES $\hat{\alpha}_i$, $i = 1 \sim 3$ USING ALGORITHM 1

	N	α_1 (1.1535)	α_2 (–0.4054)	α_3 (–0.3138)	$\frac{1}{\sqrt{N}}$
RMS error	512	0.0451	0.0414	0.0363	0.0442
	1024	0.0284	0.0322	0.0326	0.0312
	2048	0.0251	0.0209	0.0201	0.0221
	4096	0.0140	0.0152	0.0163	0.0156
Average	512	1.1712	–0.4011	–0.3114	
	1024	1.1487	–0.4071	–0.3238	
	2048	1.1500	–0.4083	–0.3185	
	4096	1.1504	–0.4076	–0.3130	

TABLE III
SIMULATION RESULTS OF PART 2 OF EXAMPLE 1. CVR(k) (%), $k = 1 \sim 8$ AND $N = 512,\ 1024,\ 2048$ AND 4096

N \ q	1	2	3	4	5	6	7	8
512	17.40	77.13	37.08	3.01	0.49	0.47	0.39	0.28
1024	13.06	81.07	36.91	3.16	0.29	0.18	0.20	0.31
2048	14.21	83.61	34.03	2.88	0.19	0.13	0.01	0.22
4096	9.74	89.79	33.41	2.82	0.01	0.15	0.02	0.06

TABLE IV
SIMULATION RESULTS OF PART 2 OF EXAMPLE 1. RMS ERRORS OF 30 INDEPENDENT AMPLITUDE PARAMETER ESTIMATES $\hat{\alpha}_i$, $i = 1 \sim 3$ AND PHASE PARAMETER ESTIMATES $\hat{\beta}_i$, $i = 1 \sim 4$ USING ALGORITHMS 2 AND 3 WITH $M = 3$, AND ALGORITHM 4 WITH $r = 2$ AND $m = 3$

Algorithm	N	α_1 (1.1535)	α_2 (–0.4054)	α_3 (–0.3138)	β_1 (–0.9112)	β_2 (0.5234)	β_3 (0.5290)	β_4 (0.2348)
2	512	0.0403	0.0391	0.0423	0.1660	0.1044	0.0892	0.1030
	1024	0.0228	0.0263	0.0351	0.1290	0.0751	0.0772	0.0464
	2048	0.0172	0.0197	0.0235	0.1199	0.0650	0.0499	0.0482
	4096	0.0112	0.0146	0.0194	0.0407	0.0266	0.0354	0.0480
3	512	0.0403	0.0391	0.0423	0.0660	0.0735	0.0677	0.0523
	1024	0.0228	0.0263	0.0351	0.0441	0.0503	0.0492	0.0288
	2048	0.0172	0.0197	0.0235	0.0263	0.0269	0.0327	0.0311
	4096	0.0112	0.0146	0.0194	0.0211	0.0211	0.0217	0.0239
4	512	0.0679	0.0749	0.0617	0.0657	0.0771	0.0696	0.0559
	1024	0.0396	0.0423	0.0411	0.0431	0.0503	0.0485	0.0285
	2048	0.0283	0.0256	0.0308	0.0257	0.0266	0.0320	0.0314
	4096	0.0185	0.0186	0.0213	0.0210	0.0211	0.0216	0.0239

was used to generate the synthetic data $x(n)$ (SNR $= \infty$). Then, Algorithm 1 was employed to estimate α_i. Simulation results obtained from 30 independent runs are shown in Tables I and II. Table I shows simulation results for AIC(k), $k = 1 \sim 6$ and $N = 512$, 1024, 2048, and 4096, where each AIC(k) was obtained by substituting the average of the obtained 30 $\varepsilon(k)$ into (40). We can see, from Table I, that AIC(k) is minimum for $k = p = 3$, indicating the true order $p = 3$. These simulation results support that AIC(k) can be used for the estimation of p. On the other hand, assuming that the true $p = 3$ was known, the averages of the obtained thirty estimates $\hat{\alpha}_i$, $i = 1 \sim p = 3$, and the associated root mean-square (RMS) errors are shown in Table II, together with the square root of the CR bounds $1/\sqrt{N}$ [by (30)]. From these two tables, we can see that all the estimates $\hat{\alpha}_i$, $i = 1 \sim 3$ are unbiased with RMS errors close to the square root of the CR bounds $1/\sqrt{N}$. These simulation results justify that Algorithm 1 is an AML estimator when data $x(n)$ are Gaussian.

In the simulation of Part 2, the procedure for generating the synthetic data $x(n)$ is the same as in Part 1, except that the driving input $u(n)$ was assumed to be a zero-mean exponentially distributed random sequence. Thirty independent estimates $\hat{\beta}_i$ were obtained using the Chien *et al.* phase equalization algorithm [Step S2) of Algorithm 2] with $M = 3$. Table III shows the simulation results for CVR(k)(%), $k = 1 \sim 8$ and $N = 512$, 1024, 2048, and 4096, where each CVR(k) was obtained by substituting the average of the obtained 30 $\eta(k)$ into (47). We can see, from Table III, that CVR(k) $\leq 1\%$ for $k > 4$, indicating the true order $q = 4$, although CVR(k) does not decrease monotonically with k. On the other hand, assuming that the true $p = 3$ and $q = 4$ were known, we also obtained 30 estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$ using Algorithms 2, 3 (with $M = 3$) and 4 (with $r = 2$ and $m = 3$). RMS errors of the obtained 30 estimates $\hat{\alpha}_i$, $i = 1 \sim p = 3$, and $\hat{\beta}_i$, $i = 1 \sim q = 4$ are shown in Table IV. Note from Table IV that the RMS errors of $\hat{\alpha}_i$ associated with Algorithms 2 and 3 are the same because Step S1) is the same for these two algorithms. RMS errors of $\hat{\alpha}_i$ and $\hat{\beta}_i$ decrease with N for all the three algorithms. The RMS errors of $\hat{\alpha}_i$ associated with Algorithms 2 and 3 are smaller than those associated with Algorithm 4. The RMS errors of $\hat{\beta}_i$ associated with Algorithm 3 are close to those associated with Algorithm 4 and, meanwhile, smaller than those associated with Algorithm 2. Nevertheless, these simulation results support the efficacy of the proposed Algorithms 2–4 and indicate that Algorithm 3 is preferable to the other two algorithms for this case (SNR $= \infty$).

Example 2: The simulation results, that is for seismic deconvolution, were obtained with $u(n)$ assumed to be a Bernoulli–Gaussian sequence (taken from [12]), a sparse spike train for modeling reflectivity sequences of the Earth, and the system (source wavelet) $h(n)$ to be a nonminimum-phase causal ARMA$(3,3)$ system (taken from [12])

$$H(z) = \frac{1 + 0.1z^{-1} - 3.2725z^{-2} + 1.411\,25z^{-3}}{1 - 1.9z^{-1} + 1.1525z^{-2} - 0.1625z^{-3}} \tag{54}$$

instead of an FSBM. Fig. 1(a) shows the synthetic data $x(n)$ for $N = 512$ and SNR $= 20$ dB (white Gaussian noise). Algorithms 2, 3 (with $M = 4$) and 4 (with $r = 2$ and $m = 4$) were employed to process the synthetic data $x(n)$, including the order estimation. The estimated order $(\hat{p}, \hat{q}) = (12, 8)$ was obtained for this case using AIC(k) and CVR(k). The simulation results associated with Algorithm 3 are shown in Fig. 1(b)–(e). Fig. 1(b) and (c) show the magnitude response and phase response of the estimated FSBM$(12, 8)$ (dash line) and those of the true ARMA$(3, 3)$ system (solid line), respectively. Fig. 1(d) and (e) show the (noncausal) estimate

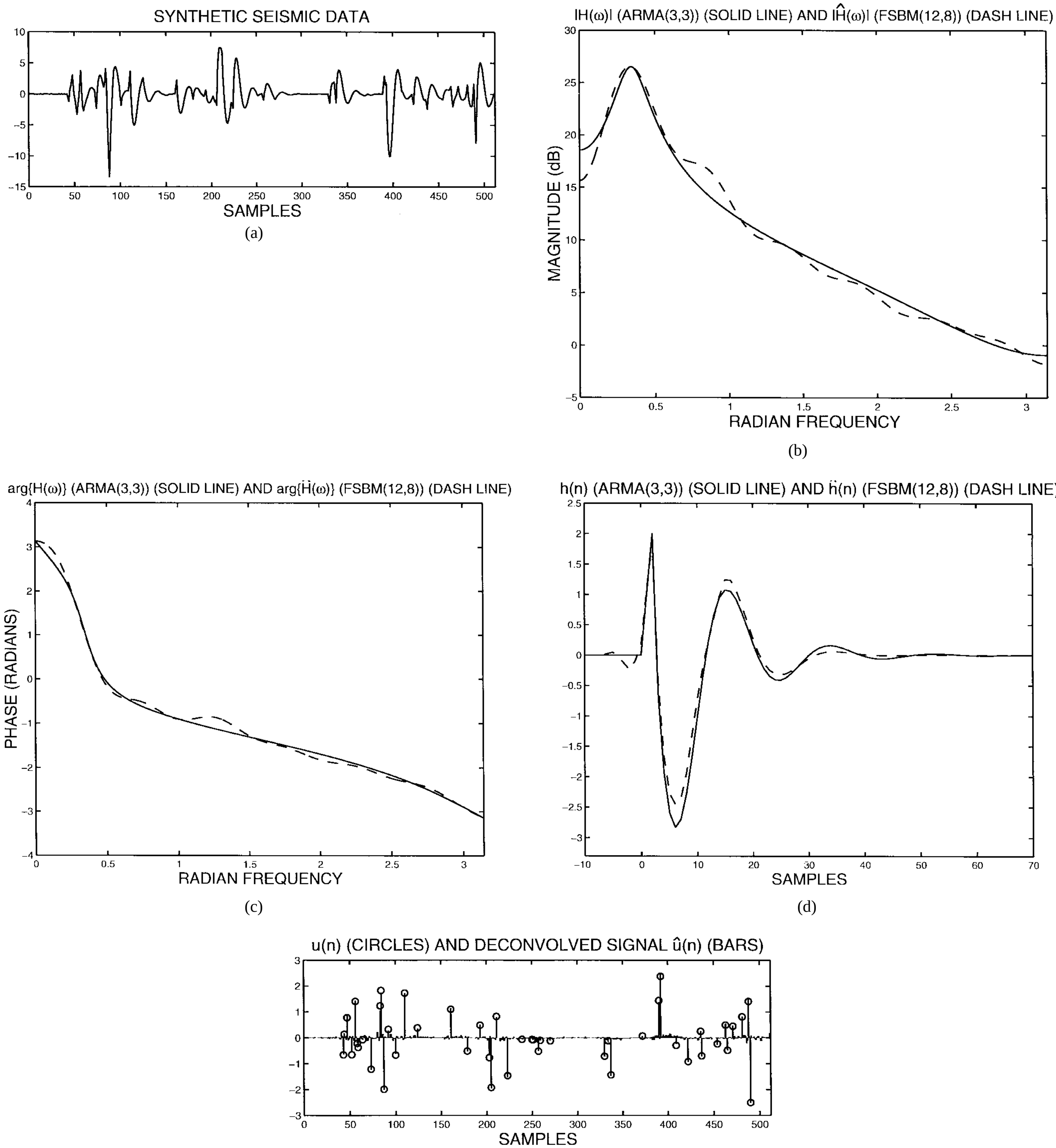

Fig. 1. Simulation results of Example 2 for Algorithm 3 with $M = 4$. (a) Synthetic seismic data $x(n)$ for $N = 512$ and SNR $= 20$ dB. (b) Magnitude response and (c) phase response of the estimated FSBM$(12, 8)$ (dash line) and those of the true ARMA$(3, 3)$ system (solid line). (d) Source wavelet $h(n)$ (solid line) and estimate $\hat{h}(n)$ (dash line). (e) Input $u(n)$ (circles) and deconvolved signal $\hat{u}(n)$ (bars).

$\hat{h}(n)$ (dash line) and the deconvolved signal $\hat{u}(n)$ (bars) [i.e., the optimum allpass filter output signal $y(n)$ obtained in Step S2) of Algorithm 3], respectively, where the scale factor and the time delay between $\hat{h}(n)$ and $h(n)$ (solid line) and those between $\hat{u}(n)$ and $u(n)$ (circles) were artificially removed. We can see that $\hat{h}(n)$ and $\hat{u}(n)$ are good approximations of $h(n)$ and $u(n)$, respectively. Both amplitude and phase response of $\hat{h}(n)$ are also good approximations of those of $h(n)$. The results obtained using Algorithms 2 and 4 are also similar to those shown in Fig. 1 and therefore are omitted.

We also performed the same simulation except that the true system was a nonminimum-phase causal ARMA$(3, 4)$ system

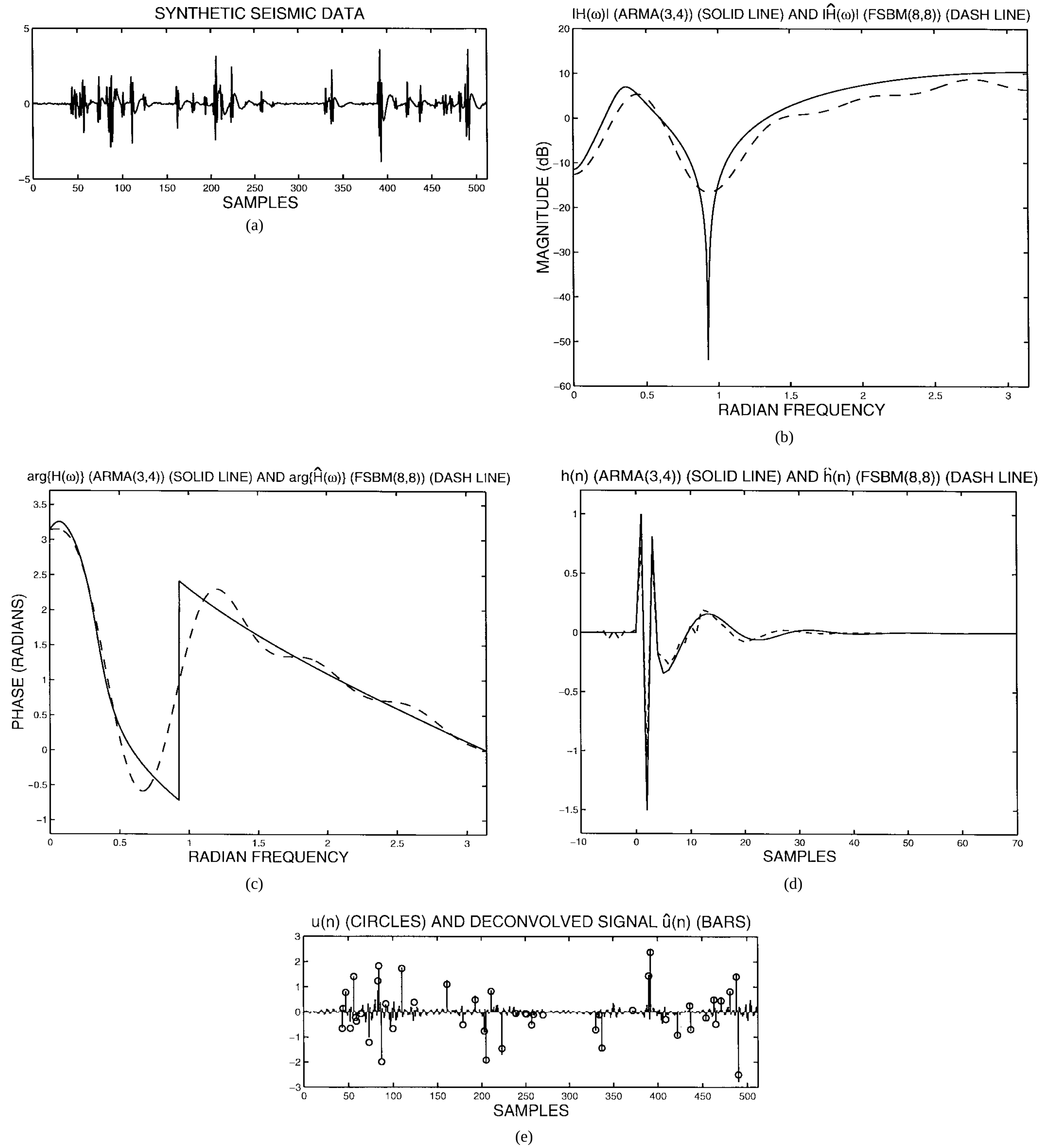

Fig. 2. Simulation results of Example 2 for Algorithm 3 with $M = 4$. (a) Synthetic seismic data $x(n)$ for $N = 512$ and SNR $= 20$ dB. (b) Magnitude response and (c) phase response of the estimated FSBM$(8, 8)$ (dash line) and those of the true ARMA$(3, 4)$ system (solid line). (d) Source wavelet $h(n)$ (solid line) and estimate $\hat{h}(n)$ (dash line). (e) Input $u(n)$ (circles) and deconvolved signal $\hat{u}(n)$ (bars).

(taken from [5])

$$H(z) = \frac{1 - 3.4z^{-1} + 4.81z^{-2} - 3.604z^{-3} + 1.17z^{-4}}{1 - 1.9z^{-1} + 1.1525z^{-2} - 0.1625z^{-3}} \tag{55}$$

which has a pair of zeros $(0.6 \pm j0.8)$ $(= e^{\pm j0.9237})$ on the unit circle. The estimated order $(\hat{p}, \hat{q}) = (8, 8)$ was obtained for this case. The results corresponding to Fig. 1(a)–(e) are shown in Fig. 2(a)–(e), respectively, for the case using Algorithm 3. Again, $\hat{h}(n)$ and $\hat{u}(n)$ are good approximations of $h(n)$ and $u(n)$, respectively, and both amplitude response and phase response of $\hat{h}(n)$ are also good approximations of those of

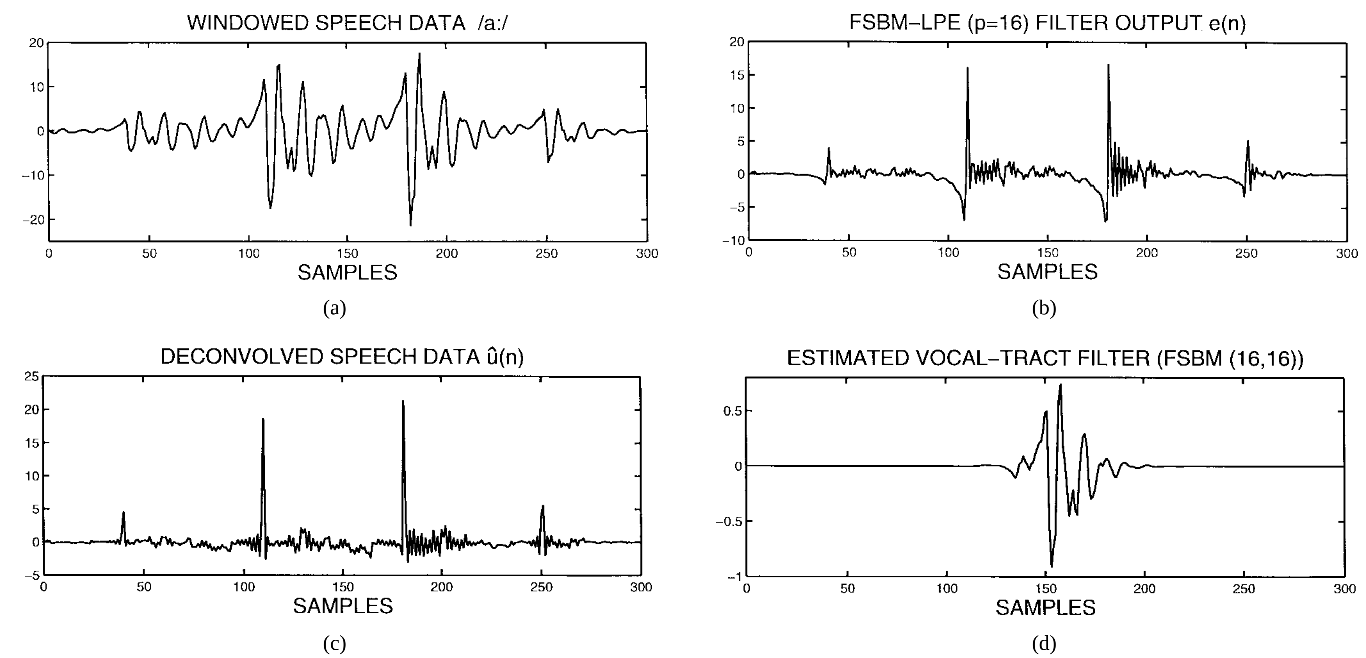

Fig. 3. Experimental results of Example 3. Speech deconvolution with the vocal tract filter modeled as an FSBM$(16, 16)$ using Algorithm 2 with $M = 3$. (a) Windowed speech data (by Hamming window) of sound $/a:/$ uttered by a man (sampling rate equal to 8 kHz). (b) Predictive deconvolved speech signal $e(n)$. (c) Deconvolved signal $\hat{u}(n)$ obtained by the inverse filter $1/\hat{H}(\omega)$. (d) Impulse response of the estimated vocal tract filter $\hat{H}(\omega)$.

$h(n)$. These simulation results support that Algorithm 3 also works well for this case that the system magnitude response has a broad dynamic range and the system phase response has a pair of discontinuities of magnitude equal to π at $\omega = \pm 0.9237$ rad due to the pair of zeros on the unit circle. As mentioned in R4), the approximation of FSBM(p, q) to the true ARMA$(3, 4)$ system is never perfect for this case, even when $p = q = \infty$. This also implies that the FSBM(p, q) is merely a stable approximation to any arbitrary LTI systems, regardless of pole and zero locations. The results obtained using Algorithms 2 and 4 are also similar to those shown in Fig. 2 and, therefore, are omitted.

Example 3: This example presents some experimental results for speech deconvolution with real voiced speech data (taken from [5]), which were obtained from a sound $/a:/$ uttered by a man through a 16-bit A/D converter with a sampling rate of 8 kHz. It is well known [3] that a voiced speech signal can be modeled as (1), where $u(n)$ is a pseudo-periodic impulse train (also called excitation signal), and $h(n)$ is the vocal tract filter (with possibly nonminimum phase). The speech data multiplied by a Hamming window, which are shown in Fig. 3(a), were processed using Algorithm 2 with the cumulant order $M = 3$ and the vocal tract filter modeled as an FSBM$(16, 16)$. Fig. 3(b) shows the LPE filter output $e(n)$ [obtained by Algorithm 1 in Step (S1) of Algorithm 2] that, as predicted, approximates a pseudo-periodic impulse train, except for some phase distortion because the vocal tract filter is not minimum phase for this case. Fig. 3(d) shows the impulse response of the estimated nonminimum-phase vocal tract filter $\hat{H}(\omega)$ (FSBM$(16, 16)$), which shows considerable resemblance to the windowed speech data of one pitch period [see Fig. 3(a)], and Fig. 3(c) shows the deconvolved speech signal $\hat{u}(n)$ by processing the windowed speech data with the inverse filter $1/\hat{H}(\omega)$ (FSBM$(16, 16)$). It can be seen, from these figures, that $\hat{u}(n)$ approximates a pseudo-periodic impulse train better than $e(n)$ with the pitch period of 70 samples (i.e., 8.75 ms) because both amplitude and phase of the vocal tract filter were equalized by the inverse filter $1/\hat{H}(\omega)$. The corresponding results for $M = 4$ and those obtained using Algorithms 3 and 4 are quite similar to those shown in Fig. 3 and, therefore, are omitted.

Only a single frame of speech data was involved in the preceding experimental results. Next, let us present some experimental results with a set of real speech data (with sampling frequency 11 025 Hz) shown in Fig. 4(a) (taken from [31]) that contains a speech segment 0.3 s long, of a female saying "why." The speech data were divided into 20 frames of 160 samples each (corresponding to 14.5 ms). Then, Algorithms 2, 3 (with $M = 3$), and 4 (with $r = 2$ and $m = 3$) were employed to process the speech data frame by frame with the vocal tract filter modeled as an FSBM$(16, 16)$ for each frame. The obtained deconvolved signals $\hat{u}(n)$ associated with Algorithms 2–4 are shown in Fig. 4(b)–(d), respectively. We can see, from these figures, that all the deconvolved signals approximate a pseudo-periodic impulse train. Over the 20 frames, the pitch period smoothly decreases from 67 samples to 43 samples, and variations of impulse magnitudes of $\hat{u}(n)$ from frame to frame can be clearly observed. We would like to mention that the total energy of each frame of the deconvolved signal shown in Fig. 4(d) was normalized by the same energy of each frame of the deconvolved signal shown in Fig. 4(c) because the unknown scale factor in the deconvolved signal associated with Algorithm 4 can be very different over different frames. As a final remark, the FSBM model used for speech signal processing may lead to simple efficient coding and compression schemes when the deconvolved signal

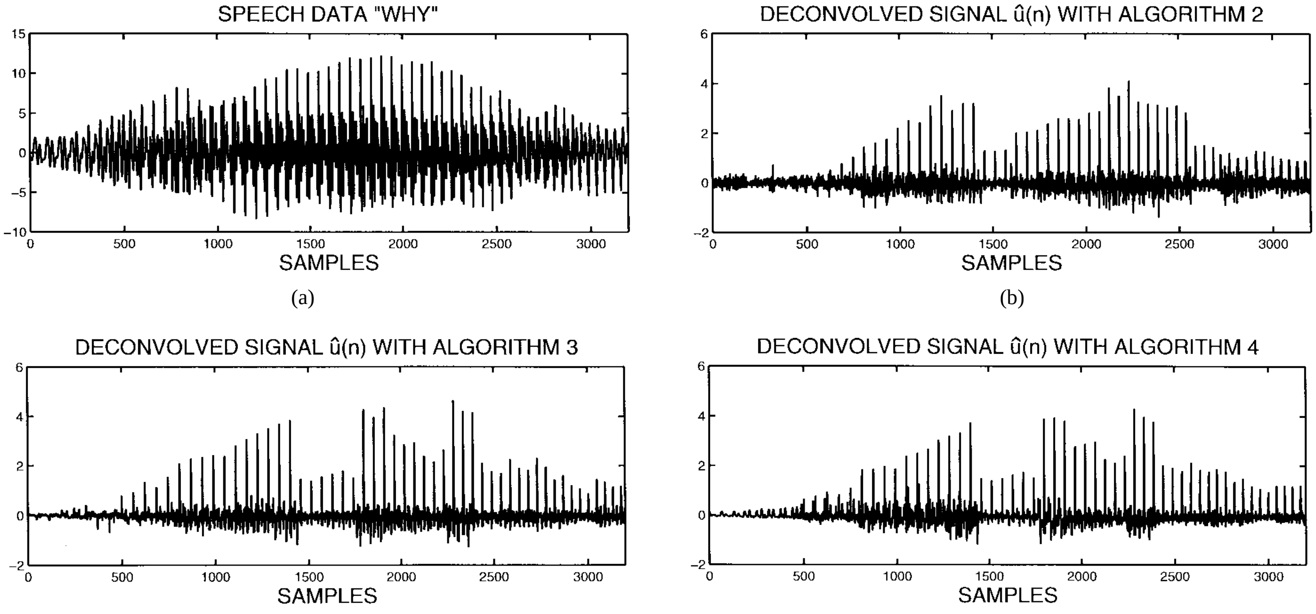

Fig. 4. Experimental results of Example 3. Speech deconvolution with the vocal tract filter modeled as an FSBM$(16, 16)$ using Algorithms 2, 3 (with $M = 3$) and 4 (with $r = 2$ and $m = 3$). (a) Speech data "why" uttered by a woman (taken from [31]) (sampling rate equal to 11 025 Hz). (b)–(d) Deconvolved signals $\hat{u}(n)$ associated with Algorithms 2, 3 and 4, respectively.

(the excitation signal) approximates to a neat pseudo-periodic impulse train.

VI. Conclusion

We have presented an FSBM(p, q) for (or as) an approximation to an arbitrary nonminimum-phase LTI system for applications in the statistical signal processing areas mentioned in the introduction. Based on the FSBM, an LPE filter for amplitude estimation together with the CR bounds, an algorithm (Algorithm 1) for obtaining the optimum LPE filter, and three algorithms (Algorithms 2–4) for estimating FSBM parameters, including the estimation of the order (p, q), were presented. Two simulation examples (Examples 1 and 2) and some experimental results with real speech data (Example 3) were presented to support the efficacy of the proposed algorithms using the proposed FSBM. However, Algorithm 3 is preferable to Algorithms 2 and 4 because for the amplitude parameter estimation, Algorithms 2 and 3 are the same and slightly superior to Algorithm 4, whereas for the phase parameter estimation, Algorithms 3 and 4 have similar performance and slightly outperform Algorithm 2.

The ARMA model has been widely used in statistical signal processing, and most signal processing algorithms using the ARMA model are suitable for the time domain implementation (recursive equations). On the other hand, signal processing algorithms using the proposed FSBM are suitable for the frequency domain implementation. The proposed FSBM may be a more suitable choice than the ARMA model for some signal processing applications such as deconvolution and channel equalization, system identification, speech coding and compression, time delay estimation, and signal detection and classification. The results presented in this paper are merely an introduction to the FSBM, and its usefulness needs to be studied further.

Appendix A
Proof of Theorem 1

The power spectrum of $e(n)$ can be easily seen to be

$$P_{ee}(\omega) = P_{xx}(\omega)|V_p(\omega)|^2 = \sigma^2 |H(\omega)V_p(\omega)|^2 = \sigma^2 |G(\omega)|^2 \tag{A.1}$$

where

$$G(\omega) = H_{\mathrm{MP}}(\omega)V_p(\omega) \quad (\text{since } |H(\omega)| = |H_{\mathrm{MP}}(\omega)|) \tag{A.2}$$

is also a causal minimum-phase system with leading coefficient $g(0) = 1$ since both $H_{\mathrm{MP}}(\omega)$ and $V_p(\omega)$ are causal minimum-phase filters with the same leading coefficient $h_{\mathrm{MP}}(0) = v_p(0) = 1$. Therefore

$$E[e^2(n)] = \sigma^2 \sum_{n=0}^{\infty} |g(n)|^2 \ge \sigma^2 |g(0)|^2 = \sigma^2 \tag{A.3}$$

which holds only when $g(n) = \delta(n)$. Thus, the optimum minimum-phase LPE filter $\hat{V}_p(\omega) = 1/H_{\mathrm{MP}}(\omega)$ when $p \ge p^\star$ [by (A.2)] and $\min\{E[e^2(n)]\} = \sigma^2$. □

Appendix B
Proof of the CR Bounds Given by (30)

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}_p^T, \sigma^2)^T$. It is well known that the CR bounds are equal to $\mathbf{I}_{\boldsymbol{\theta}}^{-1}$, where $\mathbf{I}_{\boldsymbol{\theta}}$ is the Fisher information matrix defined as [1]

$$\mathbf{I}_{\boldsymbol{\theta}} = -E\left\{\frac{\partial^2 \ln p(\mathbf{x})}{\partial \boldsymbol{\theta}^2}\right\}. \tag{B.1}$$

The log function of $p(\mathbf{x})$ given by (22) with $|H(\omega)|$ substituted by (7) yields

$$\ln p(\mathbf{x}) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{N}{2\sigma^2}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} I(\omega) \times \exp\left\{-2\sum_{i=1}^{p} \alpha_i \cos(i\omega)\right\} d\omega\right\}. \tag{B.2}$$

Taking partial derivative of $\ln p(\mathbf{x})$ given by (B.2) with respect to α_k yields

$$\frac{\partial \ln p(\mathbf{x})}{\partial \alpha_k} = \frac{N}{2\sigma^2}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} 2\cos(k\omega) I(\omega) \times \exp\left\{-2\sum_{i=1}^{p}\alpha_i \cos(i\omega)\right\} d\omega\right\}. \quad \text{(B.3)}$$

Similarly, it can be shown that

$$\frac{\partial \ln p(\mathbf{x})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{N}{2\sigma^4}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} I(\omega) \times \exp\left\{-2\sum_{i=1}^{p}\alpha_i \cos(i\omega)\right\} d\omega\right\}. \quad \text{(B.4)}$$

The second-order partial derivatives are

$$\frac{\partial^2 \ln p(\mathbf{x})}{\partial \alpha_k \partial \alpha_l} = -\frac{N}{2\sigma^2}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} 4\cos(k\omega)\cos(l\omega) I(\omega) \times \exp\left\{-2\sum_{i=1}^{p}\alpha_i \cos(i\omega)\right\} d\omega\right\} \quad \text{(B.5)}$$

$$\frac{\partial^2 \ln p(\mathbf{x})}{\partial \alpha_k \partial \sigma^2} = -\frac{N}{2\sigma^4}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} 2\cos(k\omega) I(\omega) \times \exp\left\{-2\sum_{i=1}^{p}\alpha_i \cos(i\omega)\right\} d\omega\right\} \quad \text{(B.6)}$$

$$\frac{\partial^2 \ln p(\mathbf{x})}{\partial \sigma^2 \partial \sigma^2} = \frac{N}{2\sigma^4} - \frac{N}{\sigma^6}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} I(\omega) \times \exp\left\{-2\sum_{i=1}^{p}\alpha_i \cos(i\omega)\right\} d\omega\right\}. \quad \text{(B.7)}$$

It has been shown in [1] that the average periodogram converges to the true power spectrum for large N, i.e.,

$$E[I(\omega)] = \sigma^2 |H(\omega)|^2 = \sigma^2 \exp\left\{2\sum_{i=1}^{p}\alpha_i \cos(i\omega)\right\}. \quad \text{(B.8)}$$

Then, from (B.5)–(B.8), we get

$$E\left\{\frac{\partial^2 \ln p(\mathbf{x})}{\partial \alpha_k \partial \alpha_l}\right\} = -N\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} 2\cos(k\omega)\cos(l\omega)\, d\omega\right\} = -N\delta(k-l) \quad \text{(B.9)}$$

$$E\left\{\frac{\partial^2 \ln p(\mathbf{x})}{\partial \alpha_k \partial \sigma^2}\right\} = -\frac{N}{2\sigma^2}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} 2\cos(k\omega)\, d\omega\right\} = 0 \quad \text{(B.10)}$$

$$E\left\{\frac{\partial^2 \ln p(\mathbf{x})}{\partial \sigma^2 \partial \sigma^2}\right\} = \frac{N}{2\sigma^4} - \frac{N}{\sigma^4}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega\right\} = -\frac{N}{2\sigma^4}. \quad \text{(B.11)}$$

Substituting (B.9)–(B.11) into (B.1) yields

$$\mathbf{I}_{\boldsymbol{\theta}} = N\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & \frac{1}{2\sigma^4}\end{bmatrix} \quad \text{(B.12)}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix, and $\mathbf{0}$ is a $p \times 1$ zero vector. Therefore, the CR bounds $\mathbf{I}_{\boldsymbol{\theta}}^{-1}$ are equal to the right-hand side of (30). □

APPENDIX C
FLETCHER–POWELL ALGORITHM

Assume that $J(\boldsymbol{\theta})$ is the objective function to be minimized and that it is a function of parameter vector $\boldsymbol{\theta}$. At the $(i+1)$th iteration, $\hat{\boldsymbol{\theta}}(i+1)$ is updated by

$$\hat{\boldsymbol{\theta}}(i+1) = \hat{\boldsymbol{\theta}}(i) - \lambda \mathbf{R}_i \mathbf{d}_i \quad \text{(C.1)}$$

where λ is a step-size parameter, and

$$\mathbf{d}_i = \left.\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\theta=\hat{\theta}(i)} \quad \text{(C.2)}$$

and

$$\mathbf{R}_{i+1} = \mathbf{R}_i + \frac{1}{\mathbf{r}_i^T \mathbf{s}_i}\left[\left\{1 + \frac{\mathbf{s}_i^T \mathbf{R}_i \mathbf{s}_i}{\mathbf{r}_i^T \mathbf{s}_i}\right\}\mathbf{r}_i \mathbf{r}_i^T - \mathbf{r}_i \mathbf{s}_i^T \mathbf{R}_i - \mathbf{R}_i \mathbf{s}_i \mathbf{r}_i^T\right] \quad \text{(C.3)}$$

where

$$\mathbf{r}_i = \hat{\boldsymbol{\theta}}(i+1) - \hat{\boldsymbol{\theta}}(i) \quad \text{(C.4)}$$

and

$$\mathbf{s}_i = \mathbf{d}_{i+1} - \mathbf{d}_i. \quad \text{(C.5)}$$

The initial matrix $\mathbf{R}_0$ can be any positive definite matrix (e.g., an identity matrix), which always leads to a positive definite $\mathbf{R}_i$ for $i > 0$, provided that λ is chosen such that $J(\hat{\boldsymbol{\theta}}(i+1)) < J(\hat{\boldsymbol{\theta}}(i))$. At each iteration, λ can be chosen as $\lambda = (1/2)^k$ for $k = 0, 1, 2, \ldots$ until $J(\hat{\boldsymbol{\theta}}(i+1)) < J(\hat{\boldsymbol{\theta}}(i))$. The iterative algorithm ends up with the optimum estimate $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(i)$ as the objective function $J(\hat{\boldsymbol{\theta}}(i))$ converges.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their valuable comments to the paper during the revision and Mr. C.-Y. Chen and Mr. C.-F. Rau for their helps in revising the simulation and experimental results.

REFERENCES

[1] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[2] M. H. Hayes, *Statistical Signal Processing and Modeling*. New York: Wiley, 1996.
[3] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[4] S. A. Dianat and M. R. Raghuveer, "Fast algorithms for phase and magnitude reconstruction from bispectra," *Opt. Eng.*, vol. 29, no. 5, pp. 504–-512, May 1990.
[5] H.-M. Chien, H.-L. Yang, and C.-Y. Chi, "Parametric cumulant based phase estimation of 1-D and 2-D nonminimum phase systems by allpass filtering," *IEEE Trans. Signal Processing*, vol. 45, pp. 1742–1762, July 1997.
[6] A. V. Oppenheim, "A speech analysis-synthesis system based on homomorphic filtering," *J. Acoust. Soc. Amer.*, vol. 45, pp. 458–465, Feb. 1969.
[7] A. V. Oppenheim and L. R. Rabiner, "System for automatic Formant analysis of voiced speech," *J. Acoust. Soc. Amer.*, vol. 47, no. 2, pt. 2, pp. 634–648, Feb. 1970.
[8] S. Subramaniam, A. P. Petropulu, and C. Wendt, "Cepstrum-based deconvolution for speech dereverberation," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 392–396, Sept. 1996.
[9] M. Akay, *Biomedical Signal Processing*. New York: Academic, 1994.
[10] T. J. Ulrych, "Application of homomorphic deconvolution to seismology," *Geophys.*, vol. 36, no. 4, pp. 650–660, 1971.
[11] B. Buttkus, "Homomorphic filtering—Theory and practice," *Geophys. Prospect.*, vol. 23, pp. 712–748, 1975.
[12] C.-Y. Chi and J.-Y. Kung, "A new identification algorithm for allpass systems by higher-order statistics," *Signal Process.*, vol. 41, no. 2, pp. 239–256, Jan. 1995.

[13] A. P. Petropulu and C. L. Nikias, "Blind deconvolution using signal reconstruction from partial higher order cepstral information," *IEEE Trans. Signal Processing*, vol. 41, pp. 2069–2095, June 1993.

[14] D. Hatzinakos, "Nonminimum phase channel deconvolution using the complex cepstrum of the cycle autocorrelation," *IEEE Trans. Signal Processing*, vol. 42, pp. 3026–3042, Nov. 1994.

[15] A. G. Bessios and C. L. Nikias, "POTEA: The power cepstrum and tricoherence equalization algorithm," *IEEE Trans. Commun.*, vol. 43, pp. 2667–2671, Nov. 1995.

[16] D. Hatzinakos and C. L. Nikias, "Blind equalization using a tricepstrum-based algorithm," *IEEE Trans. Commun.*, vol. 39, pp. 669–682, May 1991.

[17] D. Hatzinakos, "Blind equalization based on prediction and polyspectra principles," *IEEE Trans. Commun.*, vol. 43, nos. 2/3/4, pp. 178–181, Feb./Mar./Apr., 1995.

[18] S. Haykin, Ed., *Blind Deconvolution*. Englewood, Cliffs, NJ: Prentice-Hall, 1994.

[19] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis*. Englewood, Cliffs, NJ: Prentice-Hall, 1993.

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

[21] J. K. Tugnait, "Consistent parameter estimation for noncausal autoregressive models via higher-order statistics," *Automatica*, vol. 26, pp. 51–61, 1990.

[22] ——, "Identification of nonminimum phase linear stochastic systems," *Automatica*, vol. 22, pp. 457–464, 1986.

[23] 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, pp. 981–985, Feb. 1993.

[24] 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, Jan. 1988.

[25] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.

[26] D. M. Burley, *Studies in Optimization*. New York: Falsted, 1974.

[27] C.-Y. Chi and M.-C. Wu, "Inverse filter criteria for blind deconvolution and equalization using two cumulants," *Signal Process.*, vol. 43, no. 1, pp. 55–63, Apr. 1995.

[28] J. K. Tugnait, "Estimation of linear parametric models using inverse filter criteria and higher order statistics," *IEEE Trans. Signal Processing*, vol. 41, pp. 3196–3199, Nov. 1993.

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

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

[31] B. Porat, *A Course in Digital Signal Processing*. New York: Wiley, 1997.

Chong-Yung Chi (S'83–M'83–SM'89) was born in Taiwan, R.O.C., 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. Since August 1989, he has been a Professor with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan. He has published more than 80 technical papers in radar remote sensing, system identification and estimation theory, deconvolution and channel equalization, digital filter design, spectral estimation, and higher order statistics (HOS)-based signal processing. His research interests include signal processing for wireless communications, statistical signal processing, and digital signal processing and their applications.

Dr. Chi has served as a reviewer for several international journals and conferences, such as the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *Electronic Letters*, since 1983. He is also an active member of New York Academy of Sciences, an active member of Society of Exploration Geophysicists, a member of European Association for Signal Processing, and an active member of the Chinese Institute of Electrical Engineering. He is also a Technical Committee Member of both 1997 and 1999 IEEE Signal Processing Workshop on Higher Order Statistics.