Signal Processing 24 (1991) 149-163
Elsevier

An adaptive maximum-likelihood deconvolution algorithm

Chong-Yung Chi and Wu-Ton Chen
Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, ROC

Received 6 July 1990
Revised 5 December 1990

Abstract. Kormylo and Mendel proposed a maximum-likelihood deconvolution (MLD) algorithm for estimating a desired sparse spike sequence $\mu(k)$, modelled as a Bernoulli-Gaussian (B-G) signal, which was distorted by a linear time-invariant system $v(k)$. Then Chi, Mendel and Hampson proposed another MLD algorithm which is a computationally fast MLD algorithm and has been successfully used to process real seismic data. In this paper, we propose an adaptive MLD algorithm, which allows $v(k)$ to be a slowly time-varying linear system, for estimating the B-G signal $\mu(k)$ from noisy data. Like the previous MLD algorithms, the proposed adaptive MLD algorithm can also recover the phase of $v(k)$ when $v(k)$ is time-invariant. Some simulation results are provided to support the proposed algorithm.

Zusammenfassung. Kormylo und Mendel haben einen Maximum-Likelihood Entfaltungs-(MLD)-Algorithmus für die Schätzung einer gewünschten, spärlichen impulsfolge $\mu(k)$ vorgeschlagen, die als ein Bernoulli-Gauß (B-G) Signal modelliert wird, das durch ein lineares, zeitinvarientes System $v(k)$ verzerrt wird. Danach haben Chi, Mendel und Hampson einen anderen MLD-Algorithmus vorgeschlagen, der ein rechentechnisch schneller MLD-Algorithmus ist und beim Verarbeiten echter seismischer Daten erfolgreich eingesetzt wurde. In diesem Beitrag schlagen wir einen adaptiven MLD-Algorithmus vor, der es erlaubt, daß $v(k)$ ein langsam zeitvariantes lineares System ist, und das B-G-Signal aus verrauschten Daten schätzt. Genau wie die vorhergehenden MLD-Algorithmen kann der vorgeschlagene adaptive MLD-Algorithmus auch die Phase von $v(k)$ gewinnen, wenn $v(k)$ zeitinvariant ist. Einige Simulationsergebnisse werden vorgestellt, die den vorgeschlagenen Algorithmus unterstützen.

Résumé. Kormylo et Mendel ont proposé un algorithme de déconvolution par maximum de vraisemblance (MLD) pour l'estimation d'une séquence $\mu(k)$ d'impulsions clairsemée, modélisée par un signal Bernoulli-gaussien (B-G) qui a été déformé par un système linéaire $v(k)$ invariant dans le temps. Puis Chi, Mendel et Hampson ont proposé un autre algorithme MLD qui est rapide au niveau du calcul et a été utilisé avec succès pour traiter des données sismiques réelles. Dans cet article, nous proposons un algorithme MLD adaptatif, ce qui permet à $v(k)$ d'être un système linéaire variant lentement dans le temps, pour l'estimation du signal B-G $\mu(k)$ à partir de données bruitées. De même que les algorithmes MLD précédents, l'algorithme MLD adaptatif proposé peut également recouvrer la phase de $v(k)$ quand $v(k)$ est invariant dans le temps. Nous présentons quelques résultats de simulation pour montrer l'intérêt de l'algorithme proposé.

Keywords. Maximum-likelihood deconvolution, block component method, nonminimum-phase linear systems, time-varying linear systems.

1. Introduction

Estimating a desired signal $\mu(k)$ from a given set of noisy data $z(k)$ based on the convolutional model

$$z(k) = y(k) + n(k) = \mu(k) * v(k) + n(k) = \sum_{i=0}^{\infty} v(i)\mu(k-i) + n(k) \quad (1)$$

is a deconvolution problem, where $y(k)$ is the noise-free measurement, $n(k)$ is measurement noise and $v(k)$ is the impulse response of a linear time-invariant signal distorting system which corresponds to such as the source wavelet in seismic deconvolution [3, 8, 11, 17, 18, 21, 23] and the channel impulse response in channel equalization [4, 20] (in communications). Conventionally,

$\mu(k)$, except for a scale factor, is estimated using a prediction error filter [21, 23] which assumes that $\mu(k)$ is a white noise and $v(k)$ is minimum-phase.

Regarding the model of the signal distorting system $v(k)$, two other equivalent forms of (1) are frequently used. The first one is a transfer function model. $V(z)$, the z-transform of $v(k)$, is modelled as an n-th order autoregressive moving average (ARMA) filter as follows:

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

where

$$A(z)=1+\alpha_1 z^{-1}+\cdots+\alpha_n z^{-n} \tag{3}$$

and

$$B(z)=\beta_1+\beta_2 z^{-1}+\cdots+\beta_n z^{-n+1}. \tag{4}$$

The other one is a state-variable model. The convolutional model (1) can be represented in an n-th order state-variable form as

$$\boldsymbol{x}(k)=\Phi\boldsymbol{x}(k-1)+\boldsymbol{\gamma}\mu(k), \tag{5}$$

$$z(k)=\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}(k)+n(k), \tag{6}$$

where $\boldsymbol{x}(k)$, $\boldsymbol{\gamma}$ and $\boldsymbol{h}$ are $n\times 1$ vectors, and Φ is an $n\times n$ matrix. Note that $v(k)=\boldsymbol{h}^{\mathrm{T}}\Phi^k\boldsymbol{\gamma}$ and that given $V(z)$ there exist many $(\Phi, \boldsymbol{\gamma}, \boldsymbol{h})$'s [12, 25]. These three models provide the designer of deconvolution algorithms with a flexibility of taking the most appropriate one at each specific signal processing stage during the algorithm design. Thus, the obtained deconvolution algorithm will be more efficient and applicable from practical viewpoints such as computational load and complexity of the obtained algorithm.

Kormylo and Mendel [9] and Kormylo [8] proposed a Bernoulli-Gaussian (B-G) model, which has been used in seismic deconvolution and biomedical ultrasonic imaging, for a sparse spike sequence as

$$\mu(k)=r(k)\cdot q(k), \tag{7}$$

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

$$P_r[q(k)]=\begin{cases}\lambda, & q(k)=1,\\ 1-\lambda, & q(k)=0.\end{cases} \tag{8}$$

The B-G model for a sparse spike sequence has led to many high-resolution deconvolution algorithms reported in the past decade such as [3, 5, 7, 9–11, 13, 14, 17, 18]. On the other hand, $n(k)$ is assumed to be zero-mean white Gaussian in almost all deconvolution problems.

The estimation of $\mu(k)$ when source wavelet $v(k)$ and statistical parameters λ, σ_r^2 and noise variance σ_n^2 are unknown is surely much more difficult than that when everything (i.e. $v(k)$, λ, σ_r^2 and σ_n^2) is known a priori. Among the existing deconvolution algorithms based on the B-G model, the ones which assume that everything is known form a major category and the ones which assume that everything is unknown form the other major category. The material reported in this paper belongs to the latter category and is more or less limited to the readers with background of deconvolution of B-G signals. In order to broaden the scope of the paper to more readers, let us briefly present the evolution of deconvolution of B-G signals associated with the former and then that associated with the latter in the following. Kwakernaak [13] reported a maximum-likelihood deconvolution (MLD) algorithm based on the Poisson-Gaussian model for $\mu(k)$ which was shown [8, 17] to be the same as the B-G model for $\lambda \ll 1$. His algorithm uses the convolutional model (1) in the whole signal processing procedure and is an adaptation of the familiar matched filtering technique. However, its performance is vulnerable to the overlapping of $r(\tau_i)v(k-\tau_i)$ in $z(k)$ where τ_i is associated with $q(\tau_i)=1$ that happens when the source wavelet is long and λ is not small. Based on the B-G model for $\mu(k)$, Kormylo and Mendel [9] and Kormylo [8] proposed a recursive fixed-lag MLD algorithm. Mahalanabis et al. [14] reported a recursive maximum a posteriori (MAP) algorithm whose performance is similar to that of Mendel and Kormylo's recursive algorithm with less computations. Kormylo and Mendel [10] pro-

posed a computationally efficient iterative MLD algorithm, called the single-most-likely-replacement (SMLR) algorithm, which is basically implemented by a Kalman filter type fixed-interval optimal smoother and outperforms the previous two recursive algorithms. The state-variable model for $v(k)$ is used in these algorithms. Goussard and Demoment [5] provided a recursive fixed-lag MAP algorithm and a recursive fixed-lag MLD algorithm, and both of them are based on a moving average (MA) degenerate state-variable representation for the source wavelet. They also indicated that the recursive fixed-lag MLD algorithm not only performs better than the recursive fixed-lag MAP algorithm but also shows better robustness. For computational feasibility, all the previous algorithms are suboptimal in nature and their performance was intuitively thought to depend on the length of source wavelet, since longer source wavelet leads to severer overlapping of $r(\tau_i)v(k-\tau_i)$ in $z(k)$, until Chi [1, 2] reported that the performance of the SMLR algorithm depends on the mainlobe width of the autocorrelation function of source wavelet rather than on the length of source wavelet.

On the other hand, estimation of $\mu(k)$ when everything is unknown is so-called blind deconvolution. Based on the state-variable model (5) and (6), Kormylo and Mendel [11] proposed a high-resolution maximum-likelihood deconvolution (MLD) algorithm [11, 17, 18], which includes estimation of $v(k)$, detection of $q(k)$, estimation of $r(k)$ and estimation of statistical parameters λ, σ_r^2 and σ_n^2, for estimating $\mu(k)=r(k)\cdot q(k)$ using measurements $z(1), z(2), \ldots, z(N)$. Under the same assumptions about $\mu(k)$, $n(k)$ and $v(k)$, Chi et al. [3] proposed a computationally fast MLD algorithm, which has been successfully used to process real seismic data [3, 18]. These algorithms are also Kalman filter based nonlinear signal processing algorithms and can, in particular, recover the phase of $v(k)$. Thus, they outperform the conventional predictive deconvolution filter partly because a more accurate model for $\mu(k)$ is used and partly because $v(k)$ might not be minimum-phase in practice. Kollias and Halkias [7] combined the recursive instrumental variables (RIV) methods for the estimation of ARMA parameters associated with $v(k)$ with the fixed-lag minimum-variance deconvolution (MVD) algorithm [17, 18] for the estimation of $\mu(k)$. Their RIV-MVD algorithm, which is also an adaptive algorithm, seems to be promising. However, to the authors' knowledge, it is never reported that the RIV-MVD algorithm has been successfully used to process real seismic data. Various other deconvolution algorithms based on the B-G model for $\mu(k)$ which are not mentioned here are still a lot and most of them are listed in the bibliography on deconvolution of [18].

However, in practice, $v(k)$ is time-varying due to frequency absorption incurred during the wavelet propagation from source to sensor. Many adaptive deconvolution algorithms for processing nonstationary seismic data were also reported in the open literature [6, 7, 15, 16, 19, 24, 26]. Prediction error based adaptive deconvolution algorithms such as [6, 15, 19, 26] and Kalman filters based adaptive deconvolution algorithms such as [16, 24] are the two major categories of existing adaptive deconvolution algorithms, whereas they are based on the conventional white noise model instead of a B-G model for $\mu(k)$ and either a time-varying $v(k)$, or a time-varying ARMA model or a time-varying state-variable model for the source wavelet. In this paper, we propose an adaptive MLD algorithm, which allows $v(k)$ to be a slowly time-varying linear system, with the same assumptions about $\mu(k)$ and $n(k)$ made by Kormylo, Mendel and Chi.

In Section 2, we present the proposed adaptive MLD algorithm. We then show some simulation results to support this algorithm in Section 3. Finally, we provide a discussion and some conclusions.

2. An adaptive MLD algorithm

The proposed MLD algorithm is basically a noncausal block signal processing algorithm. A

block of $z(i)$, $i=k, k+1, \ldots, k+2L-1$, where L is a positive integer, is processed to yield $\hat{\mu}(i)$ for $i=k, k+1, \ldots, k+L-1$. Then the next block of $z(i)$, $i=k+L, k+L+1, \ldots, k+3L-1$ is processed to yield $\hat{\mu}(i)$ for $i=k+L, k+L+1, \ldots, k+2L-1$. In other words, the size of signal processing block is $2L$ and the contiguous blocks have a 50% overlap. $\hat{\mu}(i)$ for $i \geq k+2L$ are obtained so on and so forth.

Assume that $v(k)$ is time-invariant (equivalently α_i and β_i are constant) within any signal processing block. Let $\boldsymbol{\theta}=(\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n)^{\mathrm{T}}$, $\boldsymbol{z}_k=(z(1), z(2), \ldots, z(k), \ldots, z(k+2L-1))^{\mathrm{T}}$, $\boldsymbol{r}_k=(r(1), r(2), \ldots, r(k), \ldots, r(k+2L-1))^{\mathrm{T}}$ and $\boldsymbol{q}_k=(q(1), q(2), \ldots, q(k), \ldots, q(k+2L-1))^{\mathrm{T}}$. The likelihood function to be maximized is defined to be

$$
\begin{aligned}
&S_k\{\boldsymbol{r}_k, \boldsymbol{q}_k, \boldsymbol{\theta}, \lambda, \sigma_{\mathrm{n}}^2 | \boldsymbol{z}_k\} \\
&\quad = p(\boldsymbol{z}_k, \boldsymbol{r}_k, \boldsymbol{q}_k | \boldsymbol{\theta}, \lambda, \sigma_{\mathrm{n}}^2) \\
&\quad = p(\boldsymbol{z}_k | \boldsymbol{r}_k, \boldsymbol{q}_k, \boldsymbol{\theta}, \sigma_{\mathrm{n}}^2) \cdot p(\boldsymbol{r}_k) \cdot P_r(\boldsymbol{q}_k | \lambda) \\
&\quad = \prod_{i=1}^{k+2L-1} p(e(i) | \boldsymbol{\theta}, \sigma_n^2) \prod_{i=1}^{k+2L-1} p(r(i)) \\
&\qquad \times \prod_{i=1}^{k+2L-1} P_r(q(i) | \lambda) \\
&\quad = \prod_{i=1}^{k+2L-1} \frac{1}{(2\pi\sigma_{\mathrm{n}}^2)^{1/2}} \exp\left\{-\frac{e(i)^2}{2\sigma_{\mathrm{n}}^2}\right\} \\
&\qquad \times \prod_{i=1}^{k+2L-1} \frac{1}{(2\pi\sigma_r^2)^{1/2}} \exp\left\{-\frac{r(i)^2}{2\sigma_r^2}\right\} \\
&\qquad \times \prod_{i=1}^{k+2L-1} \lambda^{q(i)}(1-\lambda)^{1-q(i)}, \qquad (9)
\end{aligned}
$$

where

$$
e(i) = z(i) - y(i) = z(i) - \mu(i) * v(i). \qquad (10)
$$

The parameter σ_r^2, which is not identifiable because of $y(k)=(s\mu(k)) * (v(k)/s)$ for any $s \neq 0$, is, therefore, assumed to be known a priori. The proposed adaptive algorithm tries to search for $\hat{\boldsymbol{r}}_k$, $\hat{\boldsymbol{q}}_k$, $\hat{\boldsymbol{\theta}}_{k+L-1}$, $\hat{\lambda}(k+L-1)$ and $\hat{\sigma}_{\mathrm{n}}^2(k+L-1)$ such that S_k is maximum when $\boldsymbol{r}_k=\hat{\boldsymbol{r}}_k$, $\boldsymbol{q}_k=\hat{\boldsymbol{q}}_k$, $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{k+L-1}$, $\lambda=\hat{\lambda}(k+L-1)$ and $\sigma_{\mathrm{n}}^2=\hat{\sigma}_{\mathrm{n}}^2(k+L-1)$, under the 'adaptiveness constraint':

(C1) $\hat{r}(i)$, $\hat{q}(i)$ (equivalently $\hat{\mu}(i)$) and $\hat{y}(i)$ (equivalently $\hat{e}(i)$) for $i \leq k-1$ be fixed.

Our approach for finding a local maximum of S_k is a block component method (BCM) [3, 11, 17, 18] given as follows:

(s1) Iter = 0

(s2) Iter = Iter + 1

(s3) Update $\hat{\boldsymbol{q}}_k$; update $\hat{\lambda}(k+L-1)$; update $\hat{\boldsymbol{r}}_k$; update $\hat{\boldsymbol{\theta}}_{k+L-1}$; update $\hat{\sigma}_{\mathrm{n}}^2(k+L-1)$

(s4) If Iter $< M$ go to (s2),

where M is the allowed maximum number of iterations and is set ahead of time. Whenever a block of parameters in (s3) is updated, S_k is guaranteed to increase with the other parameters fixed. The desired estimates of $\hat{\mu}(i)=\hat{r}(i)\hat{q}(i)$, $k \leq i \leq k+L-1$, can be obtained from the first L elements of $\hat{\boldsymbol{r}}_k$ and $\hat{\boldsymbol{q}}_k$. We, next, present how to update $\hat{\boldsymbol{r}}_k$, $\hat{\boldsymbol{q}}_k$, $\hat{\boldsymbol{\theta}}_{k+L-1}$, $\hat{\lambda}(k+L-1)$ and $\hat{\sigma}_{\mathrm{n}}^2(k+L-1)$, respectively, by processing the measurement block of $z(k), z(k+1), \ldots, z(k+2L-1)$.

2.1. Estimation of λ and σ_{n}^2

Setting the partial derivatives of S_k with respect to λ and σ_{n}^2 equal to zero yields

$$
\hat{\lambda}(k+L-1) = \frac{1}{k+2L-1}\left\{\sum_{i=1}^{k-1} \hat{q}(i) + \sum_{i=k}^{k+2L-1} q(i)\right\} \qquad (11)
$$

and

$$
\hat{\sigma}_{\mathrm{n}}^2(k+L-1) = \frac{1}{k+2L-1}\left\{\sum_{i=1}^{k-1} \hat{e}(i)^2 + \sum_{i=k}^{k+2L-1} e(i)^2\right\}, \qquad (12)
$$

respectively, where we have used $q(i)=\hat{q}(i)$ and $e(i)=\hat{e}(i)$ for $i \leq k-1$.

We basically follow the same procedure for amplitude estimation, detection of $q(i)$ and estimation of $\boldsymbol{\theta}$ as reported in [3, 11, 17, 18] with some necessary modifications in view of the constraint (C1).

2.2. Detection of q(i)

The well-known SMLR algorithm [3, 10, 11, 17, 18] is an efficient algorithm for the detection of $q(i)$ for $k \leq i \leq k+L-1$. Let $\Lambda(j)$ denote the following likelihood ratio:

$$\Lambda(j) = \frac{S_k\{\hat{\boldsymbol{r}}_j, \boldsymbol{q}_j, \boldsymbol{\theta}, \lambda, \sigma_n^2 | z_k\}}{S_k\{\hat{\boldsymbol{r}}_r, \boldsymbol{q}_r, \boldsymbol{\theta}, \lambda, \sigma_n^2 | z_k\}}, \tag{13}$$

where $\boldsymbol{q}_r = (q_r(1) = \hat{q}(1), \quad q_r(2) = \hat{q}(2), \ldots, q_r(k-1) = \hat{q}(k-1), q_r(k), \ldots, q_r(k+2L-1))^T$, in which we have used $q_r(i) = \hat{q}(i)$ for $i \leq k-1$, is a reference sequence, $\boldsymbol{q}_j = (q_r(1), \ldots, q_r(j-1), 1-q_r(j), q_r(j+1), \ldots, q_r(k+2L-1))^T$ is a test sequence which differs from $\boldsymbol{q}_r$ only at a single time location j, and $\hat{\boldsymbol{r}}_j$ as well as $\hat{\boldsymbol{r}}_r$ (see (17) below) are the corresponding optimum amplitude estimates associated with $\boldsymbol{q}_j$ and $\boldsymbol{q}_r$, respectively. The iterative SMLR detection procedure is summarized as follows:

(a) Compute $\ln \Lambda(j)$ for $j = k, \quad k+1, \ldots, k+2L-1$.

(b) Assume that $\ln \Lambda(j') = \max\{\ln \Lambda(j) | k \leq j \leq k+2L-1\}$; if $\ln \Lambda(j') > 0$, update $q_r(j')$ by $1-q_r(j')$ and go to (a).

When $\ln \Lambda(j) \leq 0$ for all $k \leq j \leq k+2L-1$, the detection of $q(i)$ is finished and $\hat{q}(i) = q_r(i)$ for $k \leq i \leq k+L-1$. The by-product $q_r(i)$ for $k+L \leq i \leq k+2L-1$ together with $q(i) = 0$ for $k+2L \leq i \leq k+3L-1$ can be used as the initial conditions for the next signal processing block associated with S_{k+L}.

$\ln \Lambda(j)$ for $j = k, k+1, \ldots, k+2L-1$ have been shown to be [3, 10]

$$\ln \Lambda(j) = \frac{1}{2} \frac{\sigma_r^2 f_j^2 \rho_j}{1+\sigma_r^2 a_j \rho_j} + \rho_j \ln \frac{\lambda}{1-\lambda}, \tag{14}$$

where

$$f_j = \boldsymbol{v}_j^T \Omega_r^{-1} z_k, \tag{15}$$

$$a_j = \boldsymbol{v}_j^T \Omega_r^{-1} \boldsymbol{v}_j, \tag{16}$$

$\rho_j = 1-2q_r(j)$, $\boldsymbol{v}_j = (0, 0, \ldots, v(0), v(1), \ldots, v(k-j))^T$ and $\Omega_r = E[z_k \, z_k^T | \boldsymbol{q}_r]$. f_j and a_j for $k \leq i \leq k+2L-1$ can also be obtained by processing the signal block of $z(k), z(k+1), \ldots, z(k+2L-1)$ using a Kalman filter type optimal smoother associated with (5) and (6), which is summarized in Appendix A, with the initial conditions $\hat{\boldsymbol{x}}(k-1|k-1)$ and $P(k-1|k-1)$ where $\hat{\boldsymbol{x}}(k-1|k-1)$ is the filtered estimate of $\boldsymbol{x}(k-1)$, $P(k-1|k-1)$ is the error covariance matrix of $\hat{\boldsymbol{x}}(k-1|k-1)$ and both of them are available prior to time point k.

2.3. Amplitude estimation

It is well-known that the ML estimate $\hat{r}(i)$ [3, 11, 17, 18] is given by

$$\hat{r}(i) = \sigma_r^2 q(i) f_i. \tag{17}$$

Therefore, obtaining $\hat{r}(i)$ using (17) with $q(i) = \hat{q}(i)$ and f_i given by (15) is trivial.

2.4. Estimation of wavelet parameters

Maximizing S_k (see (9)) with respect to $\boldsymbol{\theta}$ under the constraint (C1) is equivalent to minimizing the following nonlinear objective function:

$$J(\boldsymbol{\theta}) = \sum_{i=k}^{k+2L-1} \tfrac{1}{2} e(i)^2. \tag{18}$$

We use a Newton-Raphson type algorithm to search for a local optimal $\boldsymbol{\theta}$. With the initial $\hat{\boldsymbol{\theta}}_{k+L-1}^0 = \hat{\boldsymbol{\theta}}_{k-1}$, $\hat{\boldsymbol{\theta}}_{k+L-1}^j$ at the j-th iteration is updated by

$$\hat{\boldsymbol{\theta}}_{k+L-1}^j = \hat{\boldsymbol{\theta}}_{k+L-1}^{j-1} - \rho H_{j-1}^{-1} \boldsymbol{g}_{j-1}, \tag{19}$$

where $0 < \rho \leq 1$ is a constant, $\boldsymbol{g}_{j-1}$ and H_{j-1} denote the gradient and the approximate Hessian matrix for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{k+L-1}^{j-1}$, respectively, as follows:

$$\boldsymbol{g}_{j-1} = \left.\frac{\partial J}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{k+L-1}^{j-1}} = \sum_{i=k}^{k+2L-1} e(i) \left.\frac{\partial e(i)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{k+L-1}^{j-1}}, \tag{20}$$

$$H_{j-1} = \left.\frac{\partial^2 J}{\partial \boldsymbol{\theta}^2}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{k+L-1}^{j-1}} \simeq \sum_{i=k}^{k+2L-1} \left(\frac{\partial e(i)}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial e(i)}{\partial \boldsymbol{\theta}}\right)^T \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{k+L-1}^{j-1}}, \tag{21}$$

where the term including the second derivative of $e(i)$ with respect to $\boldsymbol{\theta}$ in (21) is neglected. In order to compute $\boldsymbol{g}_{j-1}$ and H_{j-1}, we need $e(i)$, $\partial e(i)/\partial\alpha_m$ and $\partial e(i)/\partial\beta_m$ for $m=1,2,\ldots,n$, and how to compute them is described in the following.

From (1), (2) and (10), we have that

$$A(q)[z(i)-e(i)]=A(q)y(i)=B(q)\mu(i), \quad (22)$$

in which we have used q^{-1} to denote a unit delay operator. Notice that $\{y(i)=\hat{y}(i)\,|\,k-n\leq i\leq k-1\}$ (due to (C1)) and $\{\mu(i)\,|\,k-n+1\leq i\leq k+2L-1\}$, where $\mu(i)=\hat{\mu}(i)$ for $i\leq k-1$ (due to (C1)), are needed for computing $y(i)$ (or $e(i)=z(i)-y(i)$), for $k\leq i\leq k+2L-1$ with (22). Taking the partial derivative of (22) with respect to α_m and β_m, for $m=1,2,\ldots,n$, we find

$$A(q)\frac{\partial e(i)}{\partial\alpha_m}=z(i-m)-e(i-m)=y(i-m) \quad (23)$$

and

$$A(q)\frac{\partial e(i)}{\partial\beta_m}=-\mu(i-m+1). \quad (24)$$

Remark that $\{y(i)\,|\,k-m\leq i\leq k+2L-1-m\}$, where $y(i)=\hat{y}(i)$ for $i\leq k-1$, is needed for computing $\partial e(i)/\partial\alpha_m$ with (23) and that $\{\mu(i)\,|\,k-m+1\leq i\leq k+2L-m\}$, where $\mu(i)=\hat{\mu}(i)$ for $i\leq k-1$, is needed for computing $\partial e(i)/\partial\beta_m$ with (24). On the other hand, the initial conditions for computing $\partial e(i)/\partial\alpha_m$ and $\partial r(i)/\partial\beta_m$ are zero. At each iteration, updating $\hat{\boldsymbol{\theta}}$ using (19) with $\rho=1$ normally leads to a decrease of J along with a stable recursive filter $1/\hat{A}(z)$ which is needed for computing $e(i)$, the gradient and the approximate Hessian matrix; otherwise, a smaller ρ must be considered.

Two remarks are worth to be mentioned here. The first one is that when $2L=N$ (total number of data), the previous adaptive MLD algorithm with $k=1$ when M is large is the same as the algorithm of Chi et al. [3] which also maximizes S_1 using the previous BCM until convergence. Secondly, to initialize the proposed adaptive MLD algorithm, the initial conditions for $(\hat{\boldsymbol{\Phi}}(0),\hat{\boldsymbol{\gamma}}(0),\hat{\boldsymbol{h}}(0))$, $\hat{\sigma}_n^2(0)$ as well as $\hat{\lambda}(0)$, must be given in advance. Of course, the closer the initial conditions to the optimal solutions, the faster the time-varying $v(k)$ will be tracked by the previous adaptive MLD algorithm. It is, therefore, appropriate that one obtains the optimum solutions associated with S_1 using the algorithm of Chi et al. until S_1 converges and then switches the mode to the previous adaptive MLD algorithm for S_i, $i\geq 2$. Next, we present some computer simulations to support the proposed adaptive MLD algorithm.

3. Simulation examples

Two simulation examples are to be presented. The first one is associated with a second order nonminimum-phase time-invariant $v(k)$ for which $\boldsymbol{\Phi}_1$, $\boldsymbol{h}_1$, $\boldsymbol{\gamma}_1$ for the 'controllable canonical form' are

$$\boldsymbol{\Phi}_1=\begin{bmatrix}0 & 1\\ -\alpha_2 & -\alpha_1\end{bmatrix}=\begin{bmatrix}0 & 1\\ -0.5 & 1\end{bmatrix},$$

$$\boldsymbol{h}_1=\begin{bmatrix}\beta_2\\ \beta_1\end{bmatrix}=\begin{bmatrix}1.3\\ 1\end{bmatrix} \quad \text{and} \quad \boldsymbol{\gamma}_1=\begin{bmatrix}0\\ 1\end{bmatrix}.$$

Note that the zero of $V(z)$ located at $z=-\beta_2/\beta_1=-1.3$ is outside the unit circle. For the second example, the true time-varying $\boldsymbol{\Phi}(k)$, $\boldsymbol{\gamma}(k)$ and $\boldsymbol{h}(k)$ of the state-variable model (5) and (6) are given as follows:

$$\boldsymbol{\Phi}(k)=\begin{cases}\boldsymbol{\Phi}_1, & k\leq 120,\\ \boldsymbol{\Phi}_1+[(\boldsymbol{\Phi}_2-\boldsymbol{\Phi}_1)\\ \qquad\times(k-120)/1280], & 120<k<1400,\\ \boldsymbol{\Phi}_2, & k\geq 1400,\end{cases}$$

$$\boldsymbol{h}(k)=\begin{cases}\boldsymbol{h}_1, & k\leq 120,\\ \boldsymbol{h}_1+[(\boldsymbol{h}_2-\boldsymbol{h}_1)\\ \qquad\times(k-120)/1280], & 120<k<1400,\\ \boldsymbol{h}_2, & k\geq 1400,\end{cases}$$

$$\boldsymbol{\gamma}(k)=\boldsymbol{\gamma}_1$$

where

$$\boldsymbol{\Phi}_2=\begin{bmatrix}0 & 1\\ -0.4 & 1.2\end{bmatrix}, \qquad \boldsymbol{h}_2=\begin{bmatrix}1.1\\ 1\end{bmatrix}.$$

Notice that the source wavelet for $120 < k < 1400$ is slowly changing from the one (solid line in Fig. 1(a)) associated with (Φ_1, γ_1, h_1) for $k \leq 120$ to the one (solid line in Fig. 7(a)) associated with $(\Phi_2, \gamma_2 = \gamma_1, h_2)$ for $k \geq 1400$ from which high frequency attenuation in this purely artificial source wavelet can be observed. For each example, we generated the noise-free data $y(k)$ by convolving the selected wavelet $v(k)$ with a B-G signal with parameters $\lambda = 0.1$ and $\sigma_r^2 = 1$. We then added a pseudo-Gaussian random noise sequence to $y(k)$ to form the synthetic data $z(k)$ with signal-to-noise

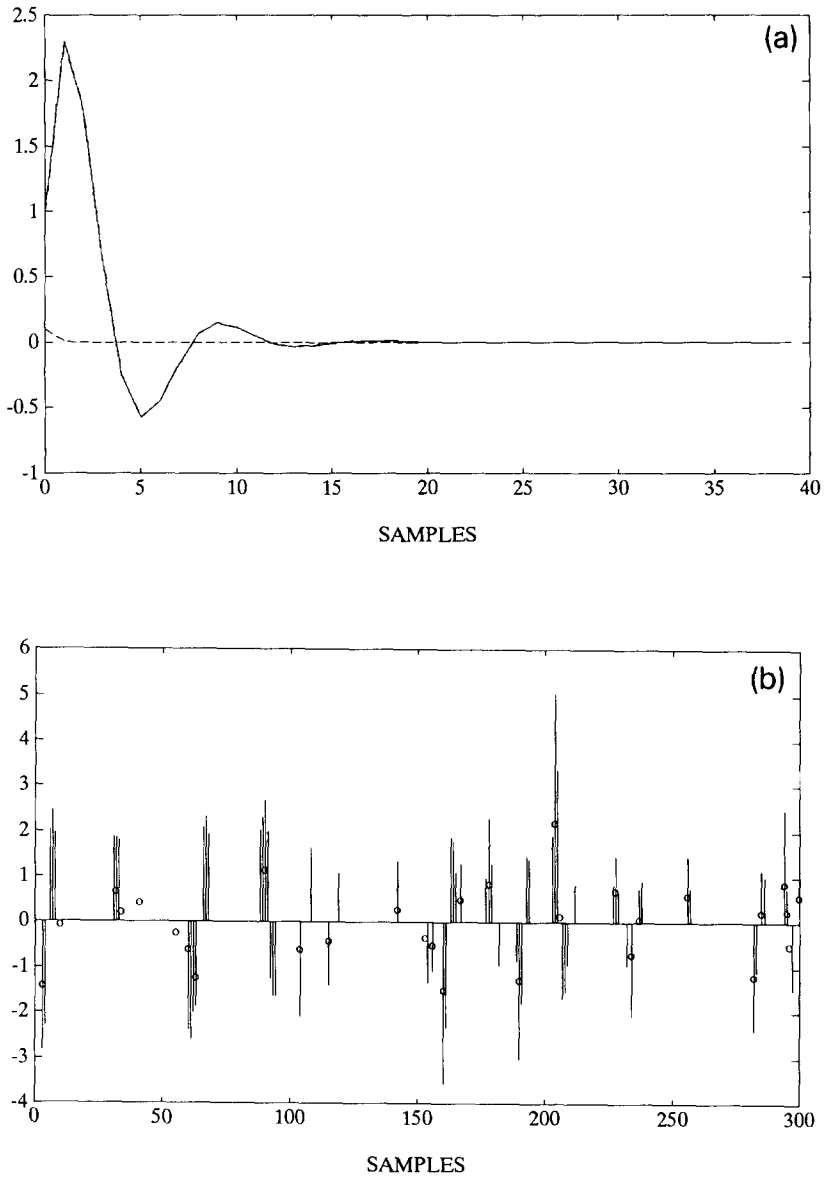

Fig. 1. Simulation results associated with Example 1. (a) True time-invariant wavelet (solid line) and estimated wavelet (dashed line) associated with S_1; (b) true $\mu(k)$ (circles) and estimated $\mu(k)$ (bars) for $1 \leq k \leq 300$.

ratio (SNR) equal to 10. We then processed $z(k)$ using the proposed adaptive MLD algorithm with $L=30$ and $M=1$.

For the first example, the initial conditions for $(\hat{\Phi}(0), \hat{\gamma}(0), \hat{h}(0))$ used were

$$\hat{\Phi}(0)=\begin{bmatrix} 0 & 1 \\ -0.01 & 0.1 \end{bmatrix}, \qquad \hat{h}(0)=\begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$$

and $\hat{\gamma}(0)=\gamma_1$. Note that $\hat{\Phi}(0)$, $\hat{h}(0)$ are very different from Φ_1 and h_1, respectively. $\hat{\sigma}_n^2(0)$ as well as $\hat{\lambda}(0)$ were equal to the true values of σ_n^2 and λ, respectively. The simulation results are shown in Figs. 1 through 3. In Figs. 1(a), 2(a) and 3(a), solid lines indicate true $v(k)$ and dashed lines depict estimated $v(k)$'s associated with S_1, S_{901} and S_{1501}, respectively. The estimated $\mu(k)$'s for

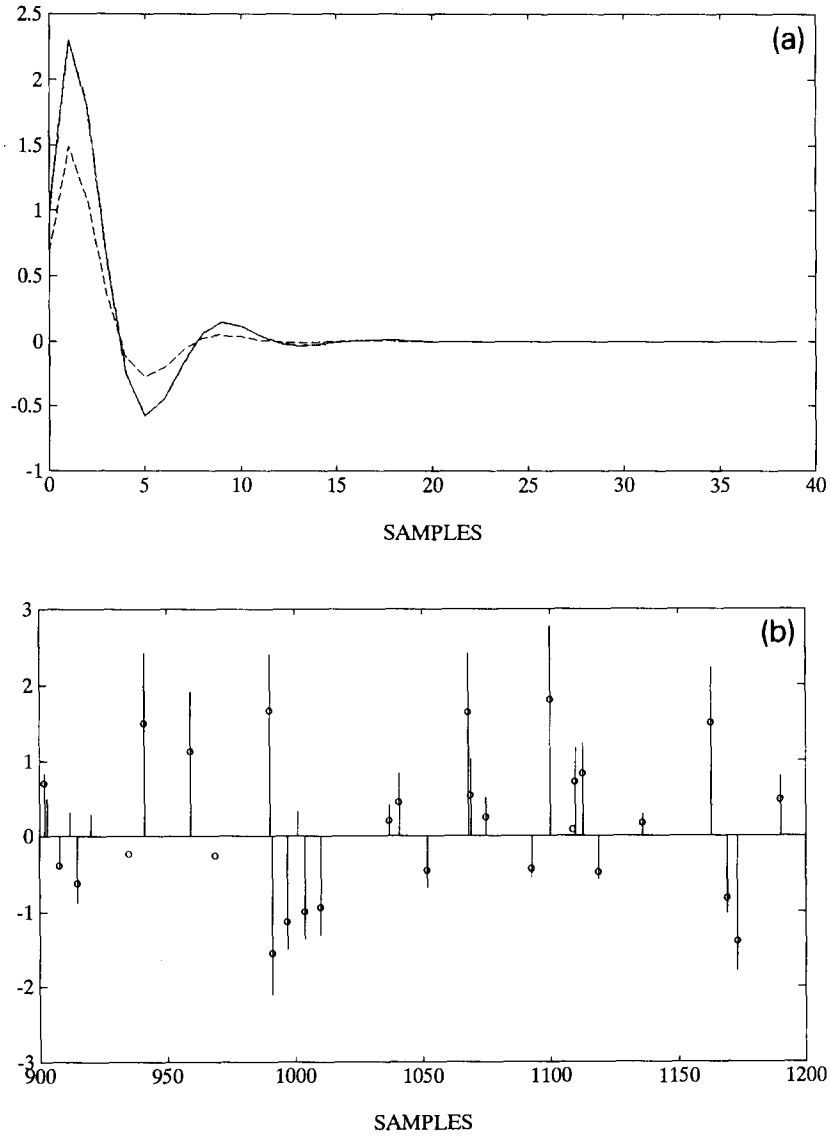

Fig. 2. Simulation results associated with Example 1. (a) True time-invariant wavelet (solid line) and estimated wavelet (dashed line) associated with S_{901}; (b) true $\mu(k)$ (circles) and estimated $\mu(k)$ (bars) for $901 \leq k \leq 1200$.

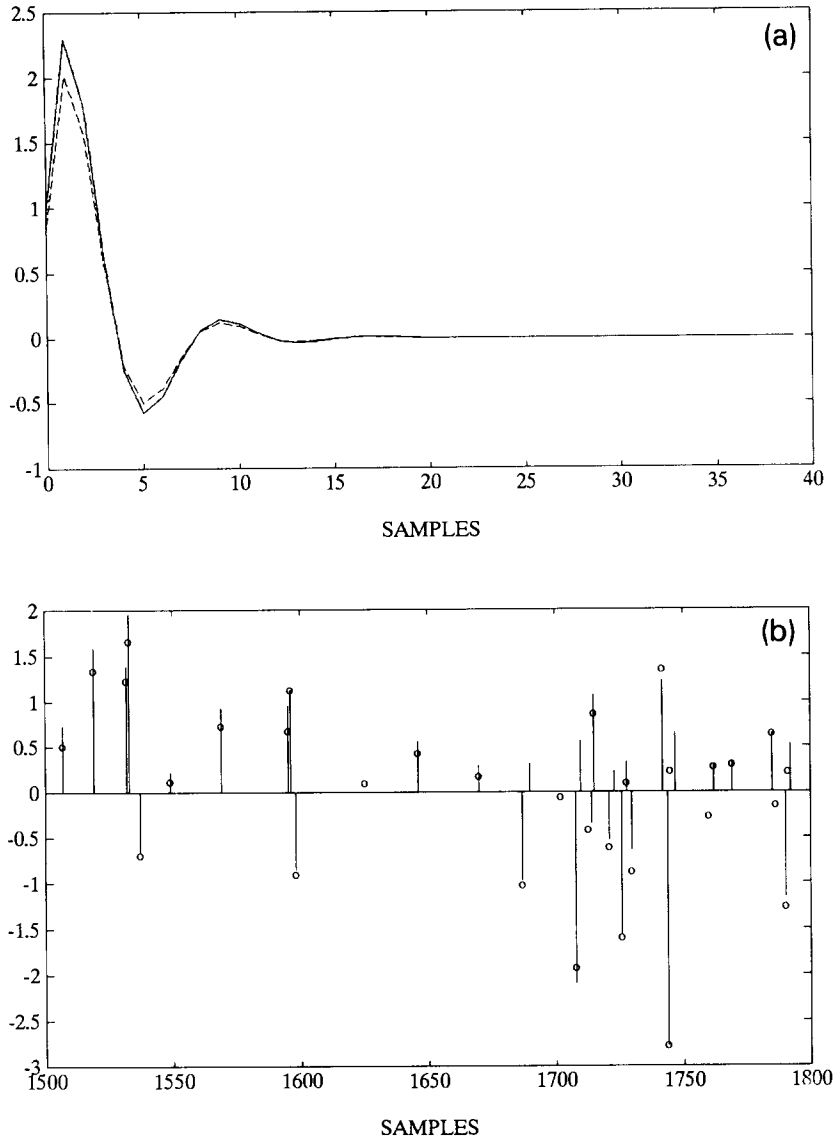

Fig. 3. Simulation results associated with Example 1. (a) True time-invariant wavelet (solid line) and estimated wavelet (dashed line) associated with S_{1501}; (b) true $\mu(k)$ (circles) and estimated $\mu(k)$ (bars) for $1501 \leq k \leq 1800$.

$1 \leq k \leq 300, 901 \leq k \leq 1200$ and $1501 \leq k \leq 1800$ are shown in Figs. 1(b), 2(b) and 3(b), respectively, where circles denote true spikes and bars denote estimates. From Fig. 1, one can see that $\hat{v}(k)$ is not close to $v(k)$ at all due to bad initial conditions used for $(\hat{\Phi}(0), \hat{\gamma}(0), \hat{h}(0))$, and thus the estimated $\mu(k)$ is terribly away from the true $\mu(k)$. From Fig. 2(a), one can see that $\hat{v}(k)$ tracks $v(k)$ well except for a scale factor $\hat{v}(k)/v(k) \approx 0.65$. This fact is consistent with Fig. 2(b) which shows that $\hat{\mu}(k)$ overshoots $\mu(k)$ by about 35%. Figure 3(b) shows that $\hat{\mu}(k)$ overshoots $\mu(k)$ by about 15% for $1500 \leq k \leq 1590$, which is also consistent with the scale factor $\hat{v}(k)/v(k) \approx 0.85$ (see Fig. 3(a)), and the amount of overshoot is smaller for $k \geq 1591$ which implies that the scale factor $\hat{v}(k)/v(k)$ associated

with S_i for $i \geq 1591$ is greater than 0.85. Again, we draw, from Fig. 3, the same conclusion that $\hat{v}(k)$ tracks $v(k)$ well except for a scale factor.

For the second example, we assume that the proposed adaptive MLD algorithm had processed all $z(k)$ for $-\infty < k < 0$ corresponding to the case of Example 1, and all unknown parameters were correctly estimated although their estimation accuracies depend on SNR. We would like to show that the algorithm can then track $v(k)$ and estimate $\mu(k)$ well from $k = 0$ on. Therefore, we processed $z(k)$ using the proposed adaptive algorithm with the initial conditions $(\hat{\Phi}(0), \hat{\gamma}(0), \hat{h}(0)) = (\Phi_1, \gamma_1, h_1)$, $\hat{\sigma}_n^2(0)$ as well as $\hat{\lambda}(0)$ equal to the true values of σ_n^2 and λ, respectively.

The simulation results are shown in Figs. 4 through 7. In Figs. 4(a), 5(a), 6(a) and 7(a), solid

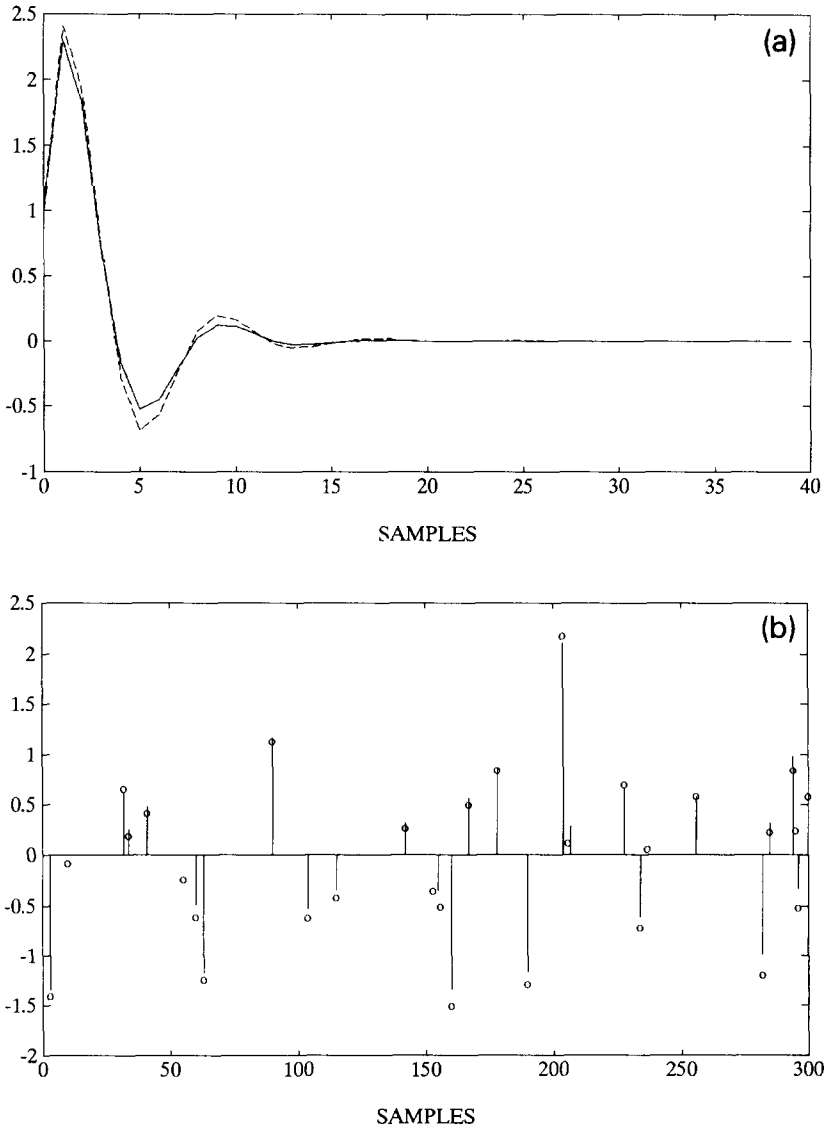

Fig. 4. Simulation results associated with Example 2. (a) True time-varying wavelet (solid line) and estimated wavelet (dashed line) associated with S_{181}; (b) true $\mu(k)$ (circles) and estimated $\mu(k)$ (bars) for $1 \leq k \leq 300$.

lines and dashed lines depict true time-varying $v(k)$'s (associated with the center time point of signal processing block) and estimated $v(k)$'s associated with S_i for $i = 181$, 721, 1081 and 1891, respectively. In Figs. 4(b), 5(b), 6(b) and 7(b), we also show $\hat{\mu}(k)$'s for $1 \leq k \leq 300$, $601 \leq 900$, $901 \leq 1200$ and $1801 \leq 2100$, respectively. From Figs. 4 through 6, one can see that $\hat{v}(k)$'s (dashed lines) track the time-varying $v(k)$ well except for a different scale factor, which is also consistent with the corresponding amount of undershoot of $\hat{\mu}(k)$ with respect to $\mu(k)$, to each figure. Figure 7(a), where $v(k)$ is actually time-invariant but not minimum-phase (with a zero located at $z = -1.1$), shows that $\hat{v}(k)$ tracks $v(k)$ well except for a scale factor $\hat{v}(k)/v(k) \approx 1.2$ which, again, is consistent with the fact (see Fig. 7(b)) that $\hat{\mu}(k)$ undershoots $\mu(k)$ by about 20%.

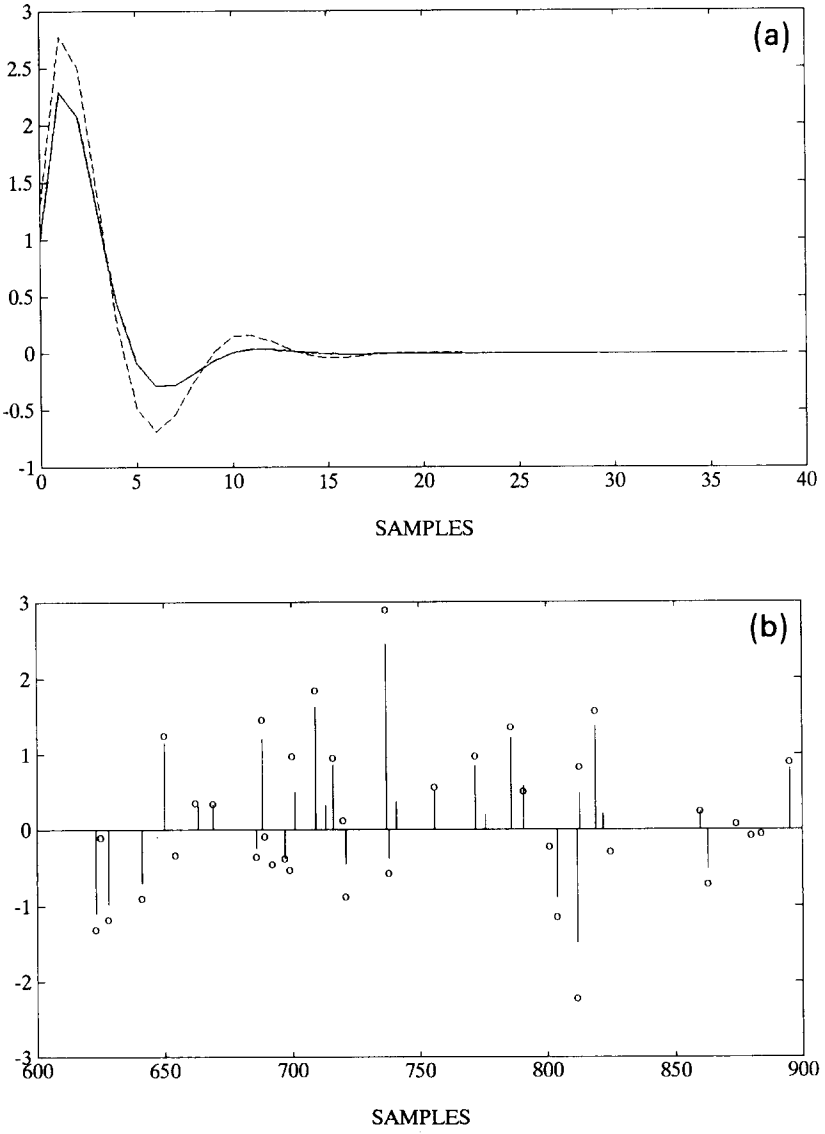

Fig. 5. Simulation results associated with Example 2. (a) True time-varying wavelet (solid line) and estimated wavelet (dashed line) associated with S_{721}; (b) true $\mu(k)$ (circles) and estimated $\mu(k)$ (bars) for $601 \leq k \leq 900$.

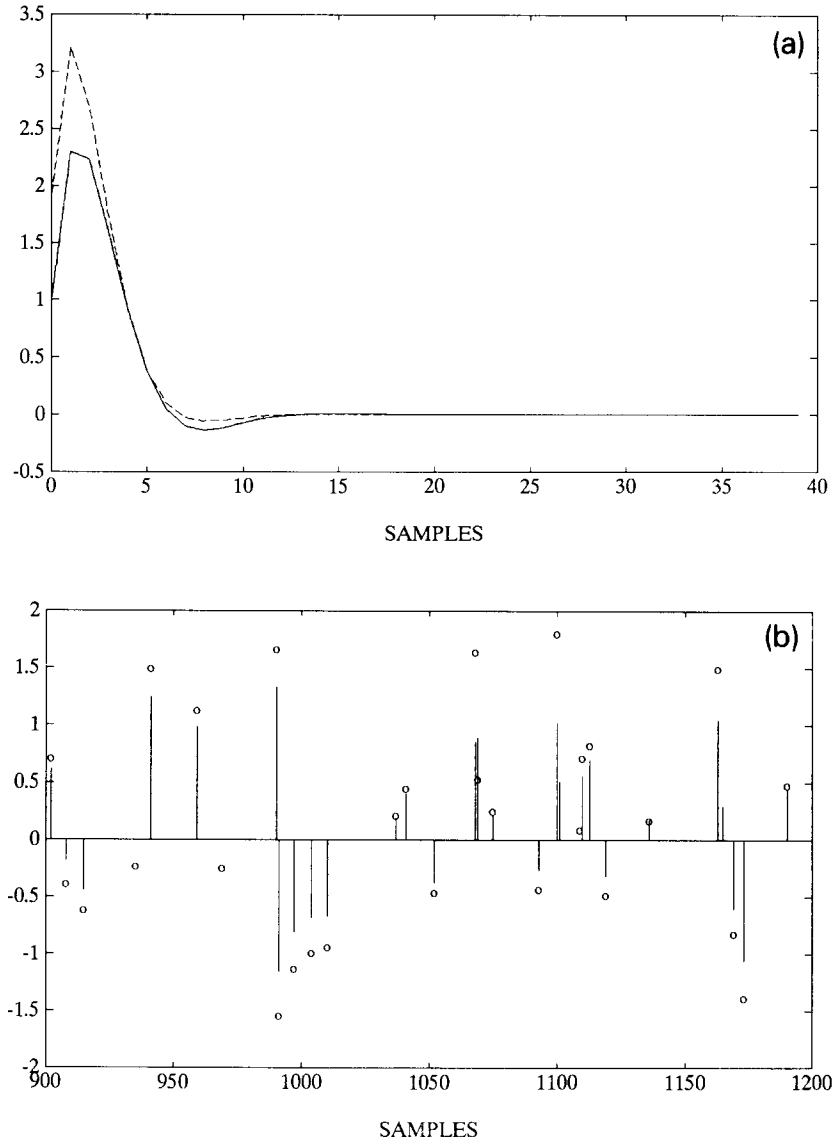

Fig. 6. Simulation results associated with Example 2. (a) True time-varying wavelet (solid line) and estimated wavelet (dashed line) associated with S_{1081}; (b) true $\mu(k)$ (circles) and estimated $\mu(k)$ (bars) for $901 \leq k \leq 1200$.

The previous two simulation examples demonstrated that the proposed adaptive MLD algorithm works well when $v(k)$ is slowly time-varying. The estimated $v(k)$ tracks $v(k)$, which is not necessarily minimum-phase when it is time-invariant, well except for a scale factor. However, the scale factor is usually existent in other deconvolution algorithms. For the case that $v(k)$ is time-invariant (Example 1 and Example 2 for $k \leq 120$ and $k \geq 1400$), if the initial conditions are closer to the optimum solutions, the proposed adaptive MLD algorithm can track $v(k)$ well by spending less recursions (equivalently less data). The scale factor for this case is also closer to unity when more data are used. As a final remark, we also performed many other simulations with the order n of the

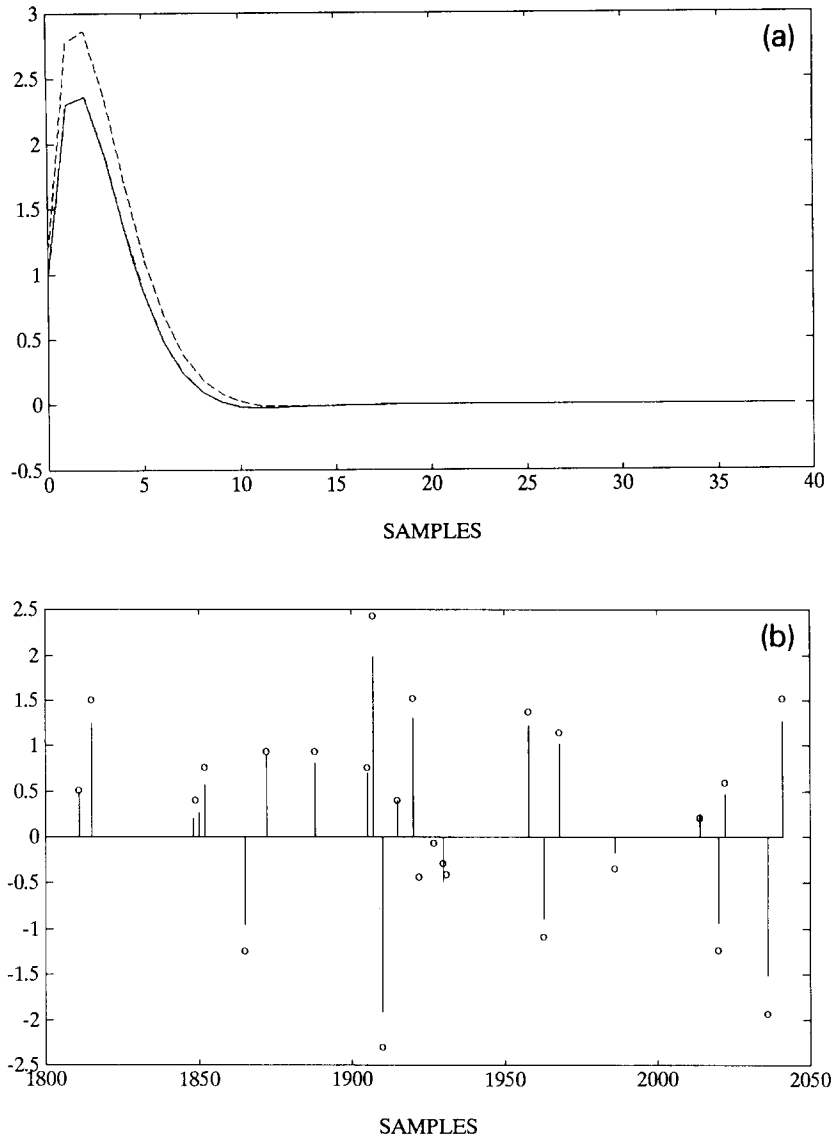

Fig. 7. Simulation results associated with Example 2. (a) True time-invariant wavelet (solid line) and estimated wavelet (dashed line) associated with S_{1891}; (b) true $\mu(k)$ (circles) and estimated $\mu(k)$ (bars) for $1801 \leq k \leq 2100$.

source wavelet larger than 2 which also led to the same conclusions obtained from the previous two examples.

4. Discussion and conclusions

We have presented a noncausal adaptive MLD algorithm for estimating a desired sparse spike sequence $\mu(k)$, modelled as a B-G signal, which was distorted by a slowly time-varying linear system $v(k)$. At each time point k, all the unknown quantities are updated based on the likelihood function S_k (see (9)) by a BCM under the adaptiveness constraint (C1). Measurements are processed block by block with a 50% overlap. Like the MLD algorithms reported in [3, 11, 17, 18], our adaptive

MLD algorithm can also recover $v(k)$ to a scale factor when $v(k)$ is time-invariant but not necessarily minimum-phase. The larger $2L$ (size of signal processing block), the better the performance of the adaptive MLD algorithm is when $v(k)$ is time-invariant. However, when $v(k)$ is slowly time-varying, $2L$ must be chosen for which $v(k)$ can be thought of as a time-invariant system within each signal processing block. The more overlap in contiguous blocks and the larger M (maximum iterations of BCM), the better the performance is at the expense of more computations. The overlap of 50% is a good choice from the viewpoints of computational load and performance by our experience.

Appendix A. Optimal smoother for computing f_j given by (15) and a_j given by (16)

The well-known Kalman filter associated with (5) and (6) is used to convert $z(j)$, $j=k, k+1, \ldots, k+2L-1$, into the innovations process, denoted $\tilde{z}(j|j-1)$, as follows:

$$\hat{\boldsymbol{x}}(j|j-1)=\Phi\hat{\boldsymbol{x}}(j-1|j-1), \tag{A.1}$$

$$P(j|j-1)=\Phi P(j-1|j-1)\Phi^{\mathrm{T}}+\sigma_r^2 q_r(j)\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathrm{T}}, \tag{A.2}$$

$$\tilde{z}(j|j-1)=z(j)-\boldsymbol{h}^{\mathrm{T}}\hat{\boldsymbol{x}}(j|j-1), \tag{A.3}$$

$$\eta(j)=\boldsymbol{h}^{\mathrm{T}}P(j|j-1)\boldsymbol{h}+\sigma_{\mathrm{n}}^2, \tag{A.4}$$

$$K(j)=P(j|j-1)\boldsymbol{h}\eta(j)^{-1}, \tag{A.5}$$

$$P(j|j)=[I-K(j)\boldsymbol{h}^{\mathrm{T}}]P(j|j-1), \tag{A.6}$$

where I is an $n\times n$ identity matrix. We then process $\tilde{z}(j|j-1)$ by the anticausal filter

$$\boldsymbol{w}(j|k+2L-1)=\Phi_{\mathrm{p}}(j)\boldsymbol{w}(j+1|k+2L-1)+\boldsymbol{h}\eta^{-1}(j)\tilde{z}(j|j-1), \tag{A.7}$$

and compute the covariance, $S_w(j|k+2L-1)$, of $\boldsymbol{w}(j|k+2L-1)$ by

$$S_w(j|k+2L-1)=\Phi_{\mathrm{p}}(j)S_w(j+1|k+2L-1)\Phi_{\mathrm{p}}^{\mathrm{T}}(j)+\boldsymbol{h}\eta^{-1}(j)\boldsymbol{h}^{\mathrm{T}}, \tag{A.8}$$

$j=k+2L-1,\ k+2L-2,\ldots,k$, where $\Phi_{\mathrm{p}}^{\mathrm{T}}(j)=\Phi[I-K(j)\boldsymbol{h}^{\mathrm{T}}]$, $\boldsymbol{w}(k+2L|k+2L-1)=\boldsymbol{0}$ (zero vector) and $S_w(k+2L|k+2L-1)=[0]$ (zero matrix). Finally,

$$f_j=\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{w}(j|k+2L-1) \tag{A.9}$$

and

$$a_j=\boldsymbol{\gamma}^{\mathrm{T}}S_w(j|k+2L-1)\boldsymbol{\gamma}. \tag{A.10}$$

Acknowledgments

We appreciate the valuable suggestions by the anonymous reviewers. The research described in this paper was performed at the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, ROC, and was supported by the National Science Council under Grant NSC79-0404-E007-11.

References

[1] C.-Y. Chi, "Performance of maximum-likelihood deconvolution for Bernoulli-Gaussian processes", *Proc. 28th IEEE Conf. on Decision and Control*, Tampa, Florida, 13-15 December 1989, pp. 2588-2589.

[2] C.-Y. Chi, "Performance of the SMLR deconvolution algorithm", *IEEE Trans. Acoust. Speech Signal Process.*, 1991, to appear.

[3] C.-Y. Chi, J.M. Mendel and D. Hampson, "A computationally-fast approach to maximum-likelihood deconvolution", *Geophysics*, Vol. 49, 1984, pp. 550-565.

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

[5] Y. Goussard and G. Demoment, "Recursive deconvolution of Bernoulli-Gaussian processes via a MA repre-

sentation", *IEEE Trans. Geosci. Remote Sensing*, Vol. 27, No. 4, July 1989, pp. 384-394.

[6] L.J. Griffith, F.R. Smolka and L.D. Trebly, "Adaptive deconvolution: A new technique for processing time-varying seismic data", *Geophysics*, Vol. 42, 1977, pp. 742-759.

[7] S.D. Kollias and C.C. Halkias, "An instrumental variable approach to minimum-variance seismic deconvolution", *IEEE Trans. Geosci. Remote Sensing*, Vol. GE-23, 1985, pp. 778-788.

[8] J. Kormylo, Maximum-likelihood seismic deconvolution, Ph.D. dissertation, Dept. of Electrical Engineering, University of Southern California, Los Angeles, CA, 1979.

[9] J. Kormylo and J.M. Mendel, "On maximum-likelihood detection and estimation of reflection coefficients", presented at *48th Annual Meeting of the Society of Exploration Geophysicists*, San Francisco, CA, 1978.

[10] J. Kormylo and J.M. Mendel, "Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes", *IEEE Trans. Inform. Theory*, Vol. IT-28, 1982, pp. 482-488.

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

[12] S.Y. Kung, "A new identification and model reduction algorithm via singular value decomposition", presented at *12th Asilomar Conf. Circuits, Systems, and Computers*, Pacific Grove, CA, 1978.

[13] H. Kwakernaak, "Estimation of pulse heights and arrival times", *Automatica*, Vol. 16, 1980, pp. 367-377.

[14] A.K. Mahalanabis, S. Prasad and K.P. Mohandis, "Recursive decision directed estimation of reflection coefficients for seismic data deconvolution", *Automatica*, Vol. 18, 1982, pp. 721-726.

[15] A.K. Mahalanabis, S. Prasad and K.P. Mohandas, "Deconvolution of nonstationary seismic data using adaptive lattice filters", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-31, 1983, pp. 591-598.

[16] A.K. Mahalanabis, S. Prasad and K.P. Mohandas, "On the application of the fast Kalman algorithm to adaptive deconvolution of seismic data", *IEEE Trans. Geosci. Remote Sensing*, Vol. GE-21, 1983, pp. 426-433.

[17] J.M. Mendel, *Optimal Seismic Deconvolution: An Estimation-based Approach*, Academic Press, New York, 1983.

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

[19] S. Prasad and A.K. Mahalanabis, "Adaptive filter structures for deconvolution of seismic signals", *IEEE Trans. Geosci. Remote Sensing*, Vol. GE-18, July 1980, pp. 267-273.

[20] J.G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1989.

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

[22] P. Rousseaux and J. Troquet, "Deconvolution of time-varying systems by Kalman filtering: Its application to the computation of the active state in the muscle", *Signal Processing*, Vol. 10, No. 3, April 1986, pp. 291-301.

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

[24] C.S. Sims and M.R. D'Mellow, "Adaptive deconvolution of seismic signals", *IEEE Trans. Geosci. Remote Sensing*, Vol. GE-16, 1978, pp. 99-103.

[25] S.A. Tretter, *Introduction To Discrete-time Signal Processing*, Wiley, New York, 1976.

[26] R.J. Wang, "Adaptive predictive deconvolution of seismic data", *Geophys. Pros.*, Vol. 25, 1977, pp. 342-381.