A FAST APPROACH TO IDENTIFICATION USING DECONVOLUTION

Chong-Yung Chi

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109

and

Jerry M. Mendel

Department of Electrical Engineering
University of Southern California
Los Angeles, California 90089

Abstract

In this paper, we propose a fast approach to impulse response and noise-variance identification for a finite-order, linear, time-invariant, single-input/single-output system, whose input driving noise is white (stationary or nonstationary) and measurement noise is stationary, white and Gaussian. Our algorithm is an iterative block component method that includes two stages, deconvolution and prediction-error identification. Experiences with our method indicates that it works well and saves about an order of magnitude in computation. Analyses and examples are given in this paper to support this claim.

I. Introduction

In this paper we are interested in identification of a finite-order, linear, time-invariant single-input/single-output discrete-time system, with the convolution representation

$$z(k) = \sum_{j=1}^{k} V(k-j)u(j) + n(k) . \qquad (1)$$

In (1) $z(k)$ is the observed output for $k = 1, 2, \ldots, N$; $V(k)$ is the system's impulse response (IR); $n(k)$ is zero-mean, white Gaussian observation noise with variance R; and $u(k)$ is a zero-mean, non-stationary white driving noise sequence, where

$$E\{u^2(k)\} = C\,q(k) \qquad (2)$$

in which $q(k)$ accounts for the time-variability of $E\{u^2(k)\}$ and C represents an additional scale factor. We seek estimates of $V(k)$ and R, denoted $\hat{V}(k)$ and $\hat{R}$, respectively, and are especially interested in the situation when $V(k)$ is non-minimum phase.

Many techniques have been developed for solving this identification problem, although most are limited to minimum phase systems. The recent studies in [3], [5] and [10] are for non-minimum phase systems. Our results are closely related to those in [3] and [9], in which it is assumed that input sequence $u(k)$ cannot be observed.

One popular technique for identifying $V(k)$ and R is to begin with a state-space model for Eq. (1):

$$\underline{x}(k) = \Phi\,\underline{x}(k-1) + \underline{\gamma}\,u(k) \qquad (3)$$

and

$$z(k) = \underline{h}'\,\underline{x}(k) + n(k) , \qquad (4)$$

where

$$\Phi = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \hline & & & \\ I_{n-1} & & & \underline{0} \end{bmatrix} \qquad (5)$$

$$\underline{\gamma} = \mathrm{col}\,(1, 0, \ldots, 0) \qquad (6)$$

and

$$\underline{h} = \mathrm{col}\,(\beta_1, \beta_2, \ldots, \beta_n) . \qquad (7)$$

It is well known that

$$V(k) = \underline{h}'\,\Phi^k\,\underline{\gamma} \qquad \forall\ k = 0, 1, \ldots ; \qquad (8)$$

hence, if we compute maximum-likelihood (ML) estimates of ARMA parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$, $\beta_1, \beta_2, \ldots, \beta_n$, then we can compute a ML estimate, $\hat{V}(k)$, of $V(k)$, from

$$\hat{V}(k) = \hat{\underline{h}}'\,\hat{\Phi}^k\,\underline{\gamma} \qquad \forall\ k = 0, 1, \ldots , \qquad (9)$$

where $\hat{\underline{h}} = \mathrm{col}\,(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n)$ and $\hat{\Phi}$ is defined similarly.

Mehra [11] and Schweppe [12], as well as others, showed how to obtain ML parameter estimates for all of the unknown parameters in Eqs. (3) and (4) when $u(k)$ cannot be measured [e.g., $u(k)$ can be thought of as a non-invasive input, such as wind acting on a launch vehicle] and is assumed to be white gaussian with variance σ_u^2. Let $\underline{\delta}$ denote the complete set of parameters in this case; then $\underline{\delta} = \mathrm{col}\,(\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n, \sigma_u^2, R)$. The log-likelihood function in this case can be expressed as

$$L\{\underline{\delta}|\underline{z}\} = \ell n\, p(\underline{z}|\underline{\delta}) = -\frac{1}{2}\sum_{j=1}^{N}\left[\tilde{z}_\delta^2(j|j-1)/\eta_\delta(j) + \ell n\,\eta_\delta(j) + \ell n\,2\pi\right] \qquad (10)$$

where

$$\underline{z} = \mathrm{col}\,[z(1), z(2), \ldots, z(N)] , \qquad (11)$$

$$\tilde{z}_\delta(j|j-1) = z(j) - \hat{z}_\delta(j|j-1) \qquad (12)$$

and

$$\eta_\delta(j) = \mathrm{Var}\,[\tilde{z}_\delta(j|j-1)] . \qquad (13)$$

Signals $\tilde{z}_\delta(j|j-1)$ and $\eta_\delta(j)$ are computed from a Kalman filter, and are subscripted $-\delta$ to denote a dependence of these quantities on the unknown δ-parameters. Kormylo and Mendel [3], [9] (see also Mendel [4]), as well as others, use a Marquardt-Levenberg algorithm to find $\hat{\underline{\delta}}$ [18]. That algorithm is

$$\hat{\underline{\delta}}_{i+1} = \hat{\underline{\delta}}_i + (H_i + D_i)^{-1}\,\underline{g}_i \qquad (14)$$

where

$$\underline{g}_i = L_\delta(\underline{\delta}|\underline{z}) = [\partial L(\underline{\delta}|\underline{z})/\partial\underline{\delta}]\big|_{\underline{\delta}=\hat{\underline{\delta}}_i} \qquad (15)$$

and

$$H_i \simeq L_{\delta\delta}(\underline{\delta}|\underline{z}) = [\partial^2 L(\underline{\delta}|\underline{z})/\partial\underline{\delta}^2]\big|_{\underline{\delta}=\hat{\underline{\delta}}_i} . \qquad (16)$$

0191-2216/83/0000-1347 $1.00 © 1983 IEEE

Matrix H_i is a pseudo-Hessian that can be computed once $\underline{g}_i$ has been computed; matrix D_i is diagonal and is chosen so that H_i+D_i is positive definite and $L(\hat{\underline{\delta}}_{i+1}|\underline{z}) > L(\hat{\underline{\delta}}_i|\underline{z})$.

In order to compute $\underline{g}_i$ we must simulate Kalman filter sensitivity systems, usually, one for each unknown element in $\underline{\delta}$. These calculations, as well as the Kalman filter itself, are computationally intensive, and in some applications, such as seismic deconvolution and wavelet estimation, have been rejected as impractical.

We remind the readers, at this point, that our overall goal is to obtain $\hat{V}(k)$, so that $\hat{\underline{\delta}}$ for us is merely a means to an end. In other situations, $\hat{\underline{\delta}}$ may be the end result iself (e.g., postflight data analysis). In the latter situations, when $u(k)$ cannot be observed, the preceding approach may be the only feasible one for computing $\hat{\underline{\delta}}$.

A second popular technique for identifying V(k) and R is to begin with a difference-equation model for Eq. (1), assuming first that

$$V(z) = B(z)/A(z) \tag{17}$$

where

$$A(z) = 1 - \sum_{i=1}^{n} \alpha_i z^{-i} \tag{18}$$

and

$$B(z) = \sum_{i=1}^{n} \beta_i z^{-i} . \tag{19}$$

In these equations, z^{-1} is the unit delay operator. The difference equation model for Eq. (1) is

$$A(z)z(k) = B(z)u(k) + A(z)n(k) . \tag{20}$$

A wide range of parameter estimation techniques can be used to identify $\alpha_1, \alpha_2, \ldots, \alpha_n$, $\beta_1, \beta_2, \ldots, \beta_n$ and R, when u(k) can be measured [i.e., u(k) can be thought of as an invasive input, such as a test signal]. Let $\underline{\theta}^1 = \text{col}\,(\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n, R)$. If we can compute ML estimates of $\underline{\theta}^1$ then we can compute the ML estimate, $\hat{V}(k)$, of V(k), from

$$\hat{V}(k) = Z^{-1}\left\{ \sum_{i=1}^{n} \hat{\beta}_i z^{-i} / [1 - \sum_{i=1}^{n} \hat{\alpha}_i z^{-i}] \right\} . \tag{21}$$

We direct our attention at a prediction-error method for obtaining $\hat{\underline{\theta}}^1$. Signal $\varepsilon(k)$ denotes the prediction error,

$$\varepsilon(k) = z(k) - \sum_{j=1}^{k} V(k-j)\,u(j) . \tag{22}$$

From Eqs. (22) and (17), we see that $\varepsilon(k)$ can be expressed as

$$A(z)\,[z(k) - \varepsilon(k)] = B(z)u(k) , \tag{23}$$

which represents a finite-difference equation for computing $\varepsilon(k)$.

The log-likelihood function for this problem can be expressed as

$$L\{\underline{\theta}^1|\underline{z}, \underline{u}\} = \ell n\, p(\underline{z}|\underline{\theta}^1, \underline{u}) =$$
$$-\frac{1}{2} \sum_{k=1}^{N} [\varepsilon^2(k)/R + \ell n\, R + \ell n\, 2\pi] . \tag{24}$$

As in the preceding problem, we can use the Marquardt-Levenberg algorithm, (14), to compute $\hat{\underline{\theta}}_i^1$. Åström and Bohlin [7] show that the gradient $\underline{g}_i$ can be computed from two relatively simple finite-difference equations. These calculations, as well as Eq. (23), are not computationally intensive, so that this prediction-error method is quite practical.

The reader has probably observed by now that we have described two quite different identification problems. In the first one, input u(k) cannot be measured and computation is intensive. In the second one, input u(k) can be measured and computation is non-intensive. In this paper we show how to transform the first problem into the second one, using deconvolution. By means of deconvolution we remove the effect of IR V(k) and noise from z(k), so that one is left with an estimate of u(k). In this paper, we do not estimate the variances σ_u^2 or C q(k) of the input. They are assumed known or fixed. The reason for this will be explained in a later section.

The approach that we suggest for identification of $\underline{\theta}^1$, called a block component method (BCM), is: (1) Obtain an initial estimate of $\underline{\theta}^1$, denoted $\hat{\underline{\theta}}_0^1$; (2) Obtain $\hat{u}_1(k)$ by means of deconvolution, using $\hat{\underline{\theta}}_0^1$ in place of $\underline{\theta}^1$; (3) Obtain $\hat{\underline{\theta}}_1^1$ by means of a prediction-error algorithm, using $\hat{u}_1(k)$ in place of u(k); (4) Obtain $\hat{u}_2(k)$ by means of deconvolution, using $\hat{\underline{\theta}}_1^1$ in place of $\underline{\theta}^1$; (5) Obtain $\hat{\underline{\theta}}_2^1$ by means of a prediction-error algorithm, using $\hat{u}_2(k)$ in place of u(k); etc. Our BCM algorithm continues until the likelihood function reaches a local maximum.

The proposed algorithm is iterative but not recursive. Many recursive algorithms are available in the system identification literature (e.g., recursive instrumental variables [13], recursive generalized least squares [14], recursive maximum-likelihood [15]). These algorithms update system parameters each time a new measurement becomes available. We update system parameters using all of the data at every iteration.

II. Maximum-Likelihood Deconvolution

According to Mendel [1], "Deconvolution is the signal processing procedure for removing the effects of a signal-distorting-system (e.g., communication channel, seismic source wavelet) from a desired signal (e.g., message, reflectivity sequence)." The signal-distortion-system is represented by IR V(k), and, in this paper is assumed unknown, so that it must be identified.

In this section, however, V(k), R and system order n are assumed known a priori, and we show how to obtain the maximum-likelihood estimate of u(k), $\hat{u}_{ML}(k|N)$, using measurements z(1), z(2), ..., z(N). We discuss two cases: first when, u(k) is white and Gaussian, and second when, u(k) is Bernoulli-Gaussian.

In the first case, when u(k) is white and Gaussian with variance σ_u^2, we define the likelihood function $S\{\underline{u}|\underline{z}, \underline{\theta}, R, \sigma_u^2\}$ to be maximized as

$$S\{\underline{u}|\underline{z}, \underline{\theta}, R, \sigma_u^2\} = p(\underline{z}|\underline{u}, \underline{\theta}, R)p(\underline{u}|\sigma_u^2) \tag{25}$$

where

$$\underline{u} = \text{col}\,(u(1), u(2), \ldots, u(N)) \tag{26}$$

and

$$\underline{\theta} = \text{col}\,(\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n)\,. \tag{27}$$

Note that $\underline{\theta}$ includes only IR parameters. We want to find the optimum estimate $\hat{\underline{u}}_{ML}$ such that $S\{\underline{u}|\underline{z}, \underline{\theta}, R, \sigma_u^2\}$ reaches a maximum when $\underline{u} = \hat{\underline{u}}_{ML}$.

When $\underline{\theta}$, R and σ_u^2 are assumed known then $\underline{z}$ and $\underline{u}$ are jointly Gaussian; thus, the maximum-likelihood estimate $\underline{u}_{ML}$ is the same as the minimum-variance estimate $\hat{\underline{u}}_{MV}$, and

$$\hat{\underline{u}}_{ML} = \hat{\underline{u}}_{MV} = E\{\underline{u}|\underline{z}\}\,. \tag{28}$$

Because both $\underline{u}$ and $\underline{z}$ are jointly gaussian, we can express (28) as [4]

$$\hat{\underline{u}}_{ML} = E\{\underline{u}\,\underline{z}'\}\,[E\{\underline{z}\,\underline{z}'\}]^{-1}\underline{z}\,. \tag{29}$$

We compute the components of $\hat{\underline{u}}_{ML}$ using Mendel's minimum-variance deconvolution (MVD) algorithm [4].

$$\hat{u}_{MV}(k|N) = \sigma_u^2\,\underline{y}'\,\underline{r}(k|N)\,, \tag{30}$$

where $\underline{r}(k|N)$ satisfies the following time-varying backwards state equation which is driven by the innovations process, $\tilde{z}(k|k-1)$ that is obtained from a Kalman filter:

$$\underline{r}(k|N) = \Phi_b(k)\,\underline{r}(k+1|N) + \underline{h}\eta^{-1}(k)\tilde{z}(k|k-1) \tag{31}$$

where $k = N, N-1, \ldots, 1$, $\underline{r}(N+1|N) = \underline{0}$,

$$\Phi_b(k) = [I_n - K(k)\,\underline{h}']'\,\Phi'\,, \tag{32}$$

K(k) is the Kalman gain and $\eta(k)$ is the variance of $\tilde{z}(k|k-1)$.

In the second case, we assume that u(k) is given by the following product model

$$u(k) = r(k)\,q(k) \tag{33}$$

where r(k) is zero mean, white, and Gaussian, with

$$E\{r^2(k)\} = C\,, \tag{34}$$

and q(k) is Bernoulli with

$$Pr[q(k)] = \begin{cases} \lambda & : q(k) = 1 \\ 1-\lambda & : q(k) = 0 \end{cases}\,. \tag{35}$$

When q(k) is known a priori, then input u(k) is a non-stationary white Gaussian process with variance

$$E\{u^2(k)|q(k)\} = q^2(k)\,E\{r^2(k)\} = Cq(k)\,. \tag{36}$$

In this case, we define the likelihood function to be maximized as

$$S\{\underline{r}|\underline{z}, \underline{q}, \underline{\theta}, R, C\} = p(\underline{z}|\underline{r}, \underline{q}, \underline{\theta}, R)\,p(\underline{r}|C) \tag{37}$$

where

$$\underline{r} = \text{col}\,(r(1), r(2), \ldots, r(N)) \tag{38}$$

and

$$\underline{q} = \text{col}\,(q(1), q(2), \ldots, q(N))\,. \tag{39}$$

Using the same procedure as described above, Mendel [4] and Kormylo and Mendel [2] prove that

$$\hat{\underline{r}}_{ML} = \hat{\underline{r}}_{MV} \tag{40}$$

and

$$\hat{\underline{u}}_{ML} = \hat{\underline{r}}_{ML}\,. \tag{41}$$

In this case, maximum-likelihood deconvolution (MLD) is the same as MVD, and we compute $\hat{\underline{u}}_{ML}$ using Mendel's MVD algorithm.

When q(k) is unknown and random, the deconvolution problem becomes more complicated. Its solution involves detection and estimation and is fully described in Mendel [4].

III. Prediction-Error Identification

In this section, we discuss the computations of the gradient $\underline{g}_i$ and the pseudo-Hessian H_i for the prediction-error method when u(k) is a test input signal.

Equation (24) can be expressed as

$$L\{\underline{\theta}, R|\underline{z}, \underline{u}\} = -\frac{1}{2}\sum_{k=1}^{N}[\varepsilon^2(k)/R + \ell n\,R + \ell n\,2\pi]$$

$$= -\frac{1}{R}\,J(\underline{\theta}) - \frac{N}{2}\,\ell n\,R - \frac{N}{2}\,\ell n\,2\pi \tag{42}$$

where

$$J(\underline{\theta}) = \frac{1}{2}\sum_{k=1}^{N}\varepsilon^2(k)\,. \tag{43}$$

From (43) and (22), we see that $J(\underline{\theta})$ is greater than zero and does not depend on R. Maximizing $L\{\underline{\theta}, R|\underline{z}, \underline{u}\}$ can be accomplished by first finding the minimum of $J(\underline{\theta})$ with respect to $\underline{\theta}$, and then finding $\hat{R}$. The latter is accomplished by taking the derivative of (42) with respect to R and setting it equal to zero. Doing this, one finds

$$\hat{R} = \frac{2}{N}\{\min_{\underline{\theta}}\,J(\underline{\theta})\}\,. \tag{44}$$

From (23), we see that $\varepsilon(k)$ is a linear function of z(k) and u(k), but is not a linear function of $\underline{\theta}$; therefore, $J(\underline{\theta})$ is also a nonlinear function of $\underline{\theta}$. We use the Marquardt-Levenberg algorithm to update $\hat{\underline{\theta}}_i$ in order to decrease $J(\underline{\theta})$. To do this, we need to compute the gradient and pseudo-Hessian of $J(\underline{\theta})$ with respect to $\underline{\theta}$. For our present purpose, we redefine the gradient $\underline{g}_i$ as

$$\underline{g}_i = J_\theta \triangleq [\partial J(\underline{\theta})/\partial\underline{\theta}]\Big|_{\underline{\theta}=\hat{\underline{\theta}}_i} \tag{45}$$

and the pseudo-Hessian H_i as

$$H_i \cong J_{\theta\theta} = [\partial^2 J(\underline{\theta})/\partial\underline{\theta}^2]\Big|_{\underline{\theta}=\hat{\underline{\theta}}_i} \tag{46}$$

Using a procedure similar to the one in [7], we determine that the elements of $\underline{g}_i$ can be computed by solving the difference equations

$$A(z)\,\varepsilon_a(k) = \varepsilon(k-1) - z(k-1) \tag{47}$$

and

$$A(z)\,\varepsilon_b(k) = -u(k) \tag{48}$$

where

$$\varepsilon_a(k) = \partial\varepsilon(k)/\partial\alpha_1 \quad \text{and} \quad \varepsilon_b(k) = \partial\varepsilon(k)/\partial\beta_1 \tag{49}$$

Additionally, $\partial\varepsilon(k)/\partial\alpha_i = \varepsilon_a(k-i+1)$ and $\partial\varepsilon(k)/\partial\beta_i = \varepsilon_b(k-i+1)$, $i = 2, \ldots, n$. (50)

Furthermore, an element of the pseudo-Hessian can be computed from one of the following three

equations:

$$\sum_{k=1}^{N}(\partial\varepsilon(k)/\partial\alpha_i)(\partial\varepsilon(k)/\partial\alpha_j) = \sum_{k=1}^{N}\varepsilon_a(k-i+1)\varepsilon_a(k-j+1)\ , \quad (51)$$

$$\sum_{k=1}^{N}(\partial\varepsilon(k)/\partial\beta_i)(\partial\varepsilon(k)/\partial\beta_j) = \sum_{k=1}^{N}\varepsilon_b(k-i+1)\varepsilon_b(k-j+1) \quad (52)$$

and

$$\sum_{k=1}^{N}(\partial\varepsilon(k)/\partial\alpha_i)(\partial\varepsilon(k)/\partial\beta_j) = \sum_{k=1}^{N}\varepsilon_a(k-i+1)\varepsilon_b(k-j+1) \quad (53)$$

where $1 \le i \le n$ and $1 \le j \le n$.

Observe that we only have to solve three difference equations--(23), (46) and (47)--in order to compute J, $\underline{g}_i$ and H_i.

IV. Block Component Method for Identification

In this section, we show, by means of deconvolution, how to convert the identification problem when $u(k)$ cannot be measured to the identification problem when $u(k)$ can be measured.

When $u(k)$ is Bernoulli-Gaussian and q and C are known a priori, we determine $\underline{\theta}$, R and $\underline{r}$ by maximizing

$$S\{\underline{r}, \underline{\theta}, R|\underline{z}, \underline{q}, C\} = p(\underline{z}|\underline{r}, \underline{q}, \underline{\theta}, R)p(\underline{r}|C)\ . \quad (54)$$

Kormylo and Mendel have shown that the overall optimal solution for estimating $\underline{r}$, $\underline{\theta}$ and R is to first estimate $\underline{\theta}$ and R and then estimate $\underline{r}$. Their approach for estimating $\underline{\theta}$ and R was developed by treating $\underline{r}$ as unknown driving noise [though $\underline{q}$ is known], and, as explained in Section I, this approach is computationally intensive.

Observe that when $\underline{r}$ is known (i.e., $\underline{u}$ is known) maximizing (53) is equivalent to maximizing $p(\underline{z}|\underline{r}, \underline{q}, \underline{\theta}, R)$ or $\ell n\ p(\underline{z}|\underline{r}, \underline{q}, \underline{\theta}, R)$ which is exactly the same as the log-likelihood function given by (42). Additionally, when $\underline{\theta}$ and R are known, maximizing (54) is equivalent to the MLD problem described in Section II.

Based on these observations, we suggest the block component method (BCM), depicted in Fig. 1, for computing a local optimum of the likelihood function $S\{\underline{r}, \underline{\theta}, R|\underline{z}, \underline{q}, C\}$.

The first block in Fig. 1 provides initial values, $\underline{\hat{\theta}}_0$ and $\hat{R}_0$, for $\underline{\hat{\theta}}$ and $\hat{R}$. Generally, we choose the initial value of $\underline{\hat{\theta}}$ by determining the minimum-phase IR associated with covariance information for the measurements. The second block in Fig. 1, provides estimates of $\underline{\hat{r}}$; it requires values for $\underline{\hat{\theta}}$ and $\hat{R}$. The third block provides $\underline{\hat{\theta}}$ and $\hat{R}$. Of course, $\underline{\hat{r}}$, obtained from the MVD filter, is used to update $\underline{\hat{\theta}}$ and R.

Each block in our BCM guarantees that $S\{\underline{r}, \underline{\hat{\theta}}, R|\underline{z}, \underline{q}\}$ increases at every iteration. When $S\{\underline{r}, \underline{\hat{\theta}}, R|\underline{z}, \underline{q}\}$ reaches a local maximum, the algorithm stops.

When $u(k)$ is white and gaussian with variance σ_u^2, we determine $\underline{\theta}$, R and $\underline{u}$ by maximizing

$$S\{\underline{u}, \underline{\theta}, R|\underline{z}, \sigma_u^2\} = p(\underline{z}|\underline{u}, \underline{\theta}, R)p(\underline{u}|\sigma_u^2)\ . \quad (55)$$

This is just a special case of the preceding situation in which we set $q(k) = 1$ for all k, $\sigma_u^2 = C$ and $\underline{r} = \underline{u}$; thus, we do not have to discuss the BCM for this case.

In our BCM, we assume that C is known, and, therefore, we do not estimate it. If C is unknown, we set it equal to a positive number, $\hat{C}$. The reason for doing this is that a scale factor cannot be resolved from the convolutional model $z(k) = u(k) * V(k) + n(k) = (Ku(k)) * (\frac{1}{K}V(k)) + n(k)$ when both $u(k)$ and $V(k)$ are unknown. In our BCM, this always occurs because even if $q(k)$ is known $\underline{u}$ is estimated via MVD [9].

For a discussion on the identifiability of $V(k)$ and R when $\underline{u}$ is known, see [16]. Because $V(k)$ and R are identifiable when $u(k)$ is known, if, in our Fig. 1 BCM, $\hat{u}(k)$ converges to the true $u(k)$, then $\hat{V}(k)$ [which is characterized by $\underline{\hat{\theta}}$] converges to true IR $V_T(k)$ as $N \to \infty$. In other words, when $u(k)$ is known, the performance of the block that updates $\underline{\theta}$ and R depends on N. In [17] it is shown that $\underline{\hat{u}}_{ML}$ is more likely to be close to $\underline{u}$ for high SNR; thus, the performance of our BCM depends on SNR, which determines the performance of the MVD filter, and N, which determines the performance of the prediction-error identification method.

V. Computational Advantages

Kormylo and Mendel [3] estimate the wavelet and noise variance using the approach that treats the input $u(k)$ [$= r(k)q(k)$] as an unknown driving noise, although $q(k)$ is assumed to be known. Their maximum-likelihood algorithm is an iterative one that needs computationally intensive Kalman filter sensitivity equations. Recently, Goutsias, Mendel and Chi [8] made a computational comparison of their algorithm and ours (Fig. 1). We present some of these results below.

Assume that $\underline{\xi}$, $\underline{\gamma}$ and $\underline{h}$ are in phase-variable form. The total number of computations to update $\underline{\hat{\theta}}$ and $\hat{R}$ one time, for the Kormylo-Mendel (i.e., KM) algorithm, is approximately

$$T_1 = \frac{N}{2}(12n^3+93n^2+113n+36) + \frac{1}{6}(8n^3+36n^2+66n+25)\ \text{flops}, \quad (56)$$

where one flop indicates one real multiplication and one addition. The total number of computations to update $\underline{\hat{\theta}}$ and $\hat{R}$ one time, for the Chi-Mendel (i.e., CM) algorithm is

$$T_2 = 10nN + \frac{2}{3}n(2n^2+9n+6)\ \text{flops}\ . \quad (57)$$

Note that $T_2 << T_1$.

In the CM algorithm, we do not have to update $\underline{\theta}$ and R just once during each iteration. We define an iteration for our block component method as one in which $\underline{\hat{r}}$ is estimated one time using an MVD filter and $\underline{\hat{\theta}}$ and $\hat{R}$ are updated L times using the Marquardt-Levenberg algorithm. The total number of computations for an MVD filter is about the same as that for two Kalman filters, and is

$$T_3 = N(5n^2+13n)\ \text{flops}\ . \quad (58)$$

The total number of computations for one iteration of the CM algorithm is

$$T = LT_2 + T_3\ . \quad (59)$$

To illustrate these numbers, assume $N = 300$ and $n = 4$; then, $T_1 = 411,830$ flops, $T_2 = 12,197$ flops and $T_3 = 39,600$ flops. The ratio T_1/T is about

8 for L = 1 and 4 for L = 5. We have observed, from simulations, that when L > 1, our BCM converges in fewer iterations than when L = 1.

VI. Examples

In this section, we provide an example which demonstrates that our proposed Section IV BCM works well.

Synthetic noisy data were generated by convolving an input sequence, $u(k)$, with a wavelet, $V_T(k)$, and adding white gaussian noise to the result. In every iteration of our BCM we estimate $\hat{r}$ once using an MVD filter and update $\hat{\xi}$ and $\hat{R}$ L = 5 times using the Marquardt-Levenberg algorithm. We used a 4th-order non-minimum phase ARMA wavelet whose transfer function (taken from [4]) is

$$V_T(z)=[-0.76286+1.5884z^{-1}-0.82356z^{-2}+0.000222419z^{-3}]/$$
$$[1-2.2633z^{-1}+1.7734z^{-2}-0.49803z^{-3}+0.045546z^{-4}]. \quad (60)$$

Figure 2 depicts the synthetic noisy data for which SNR = 10. True parameter values for R, C and λ are 0.1698×10^{-3}, 0.0225 and 0.05, respectively. The total number of measurements is N = 300. Initial values for $\hat{R}$ and $\hat{C}$ were chosen to be 0.2×10^{-3} and 0.015, respectively. For reasons explained in Section IV parameter $\hat{C}$ was fixed at its initial value.

Figure 3 depicts both the initial IR $V_0(k)$, and the true IR $V_T(k)$. IR $V_0(k)$ was obtained using a singular value decomposition (SVD) method [6]; its transfer function is

$$V_0(z)=[-1.047377+2.521712z^{-1}-2.104901z^{-2}+0.6229523z^{-3}]/$$
$$[1-2.735755z^{-1}+2.784949z^{-2}-1.190092z^{-3}+0.163848z^{-4}]. \quad (61)$$

One can check that $V_0(k)$ is minimum phase. Observe, also, that although the initial wavelet looks similar to the true wavelet, its phase is incorrect.

Using the ARMA parameters of $V_0(k)$ as the initial parameters, $\hat{\xi}_0$, in our BCM, the BCM converged in 6 iterations. The wavelet $\hat{V}(k)$ associated with $\hat{\xi}$ is non-minimum phase and has the transfer function

$$\hat{V}(z)=[-1.016298+2.23804z^{-1}-1.347001z^{-2}+0.124965z^{-3}]/$$
$$[1-2.370875z^{-1}+2.015682z^{-2}-0.6872525z^{-3}+0.09224902z^{-4}]. \quad (62)$$

Figure 4 depicts $\hat{V}(k)$ and $V_T(k)$. Observe how close $\hat{V}(k)$ is to $V_T(k)$. Figure 5 depicts $\hat{u}(k)$ and $u(k)$. Because $q(k)$ has been assumed to be known for all k, $\hat{u}$ only has non-zero values at those time points where $u(k)$ does. We see that $\hat{u}(k)$ is also very close to $u(k)$.

The final estimate for R is $\hat{R} = 0.1545\times10^{-3}$. Quantity $\hat{R}$ also represents the mean-square error of the estimated output.

The CPU time on a PDP-10/KI computer for this example was 16 seconds. For other examples, see [16].

VII. Conclusions

We have proposed a fast (computationally nonintensive) maximum-likelihood algorithm that can estimate a non-minimum (or minimum) phase IR for a linear time-invariant system. Our algorithm is an iterative block component method that includes two stages, one stage of which is deconvolution and the other stage of which is prediction-error identification. The user may wish, and can, replace the prediction-error algorithm with his favorite ML identification algorithm.

In time-series identification problems where there are no true parameters the Mehra/Schweppe approach to parameter estimation, which is computationally intensive, can be replaced by our BCM.

Experiences with our method indicates that it works well and saves about an order of magnitude in computation.

Acknowledgement

The work reported on in this paper was performed at the University of Southern California, Los Angeles, California, under National Science Foundation Grant NSF ECS-8200882, and under the support of the sponsors of the USC Geo-Signal Processing Program.

References

[1] J. M. Mendel, "Minimum-variance deconvolution," IEEE Trans. on Geoscience and Remote Sensing, vol. GE-19, pp. 161-171, 1981.

[2] J. Kormylo and J. M. Mendel, "Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes," IEEE Trans. on Infor. Theory, vol. IT-28, No. 3, pp. 482-488, May 1982.

[3] J. Kormylo and J. M. Mendel, "Identifiability of non-minimum phase linear stochastic systems," IEEE Trans. on Auto. Control, 1983.

[4] J. M. Mendel, "Optimal Seismic Deconvolution: an estimation-based approach," Academic Press, Inc., New York, 1983.

[5] J. D. Scargle, "Studies in astronomical time series analysis: Modeling random processes in the time domain," NASA TM81148, 1979.

[6] 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.

[7] K. J. Åström and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records," presented at the IFAC Symp. Theory of Self-Adaptive Cont. Syst., Teddington, England, 1965.

[8] J. Goutsias, J. M. Mendel and C. Y. Chi, "Computational aspects of maximum likelihood deconvolution," Report No. GEO-R-4, November, 1982, University of Southern California, Los Angeles, CA.

[9] J. Kormylo and J. M. Mendel, "Maximum-likelihood seismic deconvolution," IEEE Trans. on Geoscience and Remote Sensing, vol. GE-21, pp. 72-82, 1983.

[10] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a non-minimum phase system: Blind adjustment of linear equalizer in data communication," IEEE Trans. on Auto. Control, vol. AC-25, no. 3, pp. 637-646, Oct. 1980.

[11] R. K. Mehra, "Identification of stochastic linear dynamic systems using Kalman filter representation," AIAA Journal, vol. 9, no. 6, pp. 774-783, Dec. 1971.

[12] F. C. Schweppe, "Evaluation of likelihood functions for gaussian signals," IEEE Trans. on Info. Theory, vol. IT-11, no. 1, pp. 61-70.

[13] P. C. Young, "Some observations on instrumental variable methods of time series analysis," Centre Resource Environ. Studies, Rep. CRES/SS/RR1, Australian National Univ., 1975.

[14] R. Hastings-James and M. W. Sage, "Recursive generalized least-squares procedure for on-line identification of process parameters," Proc. IEE, 116, No. 12, 1969.

[15] I. Soderstrom, L. Ljung and and I. Gustavsson, "A comparative study of recursive identification methods," Report 7427, Lund Inst. Tech., Dept. Auto. Control, 1974.

[16] C. Y. Chi, "Single-channel and multichannel deconvolution," Ph.D. dissertation, Univ. of Southern California, 1983.

[17] C. Y. Chi and J. M. Mendel, "Performance of MVD filters," in preparation.

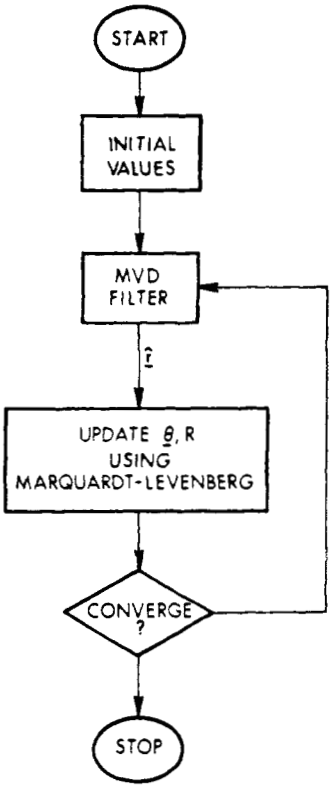

Fig. 1. Block component method.

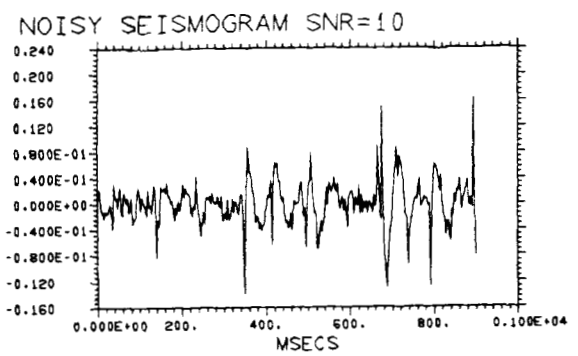

Fig. 2. Synthetic noisy data with SNR = 10.

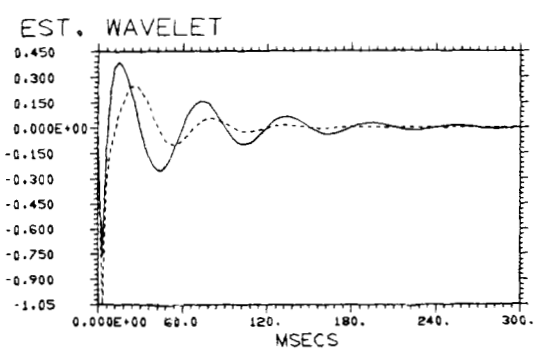

Fig. 3. Initial 4^{th}-order wavelet. Solid line depicts true wavelet, and dashed line depicts initial wavelet.

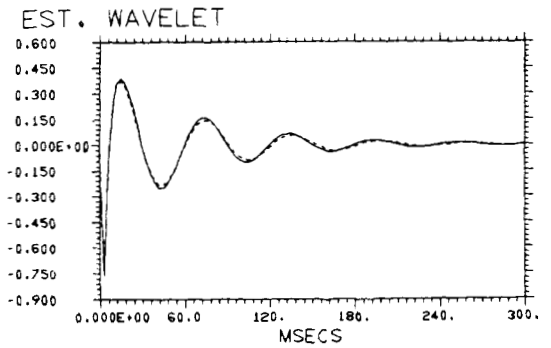

Fig. 4. Estimated wavelet. Solid line depicts true wavelet, and dashed line depicts scaled estimated wavelet $K\hat{V}(k)$ ($K = 0.75$).

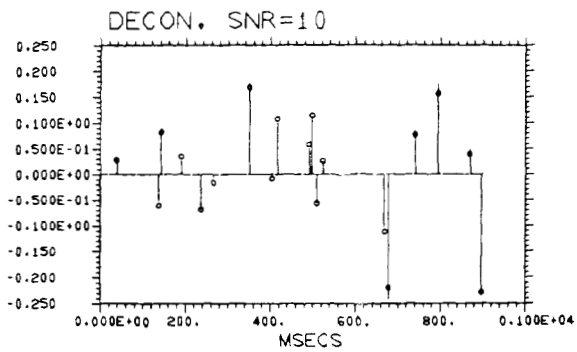

Fig. 5. Estimated input signal. Circles depict true input signal, and bars depict scaled estimated input signal $\frac{1}{K}\hat{u}(k)$ ($K = 0.75$).