Correspondence

An Improved Inverse Filtering Method for Parametric Spectral Estimation

Chong-Yung Chi and Diing Wang

Abstract—**For a wide-sense stationary process $x(k)$, it is well known that its power spectrum $P_{xx}(f)$ can be estimated by whitening the data with the inverse filter, $V(z) = 1/H(z)$, of the assumed minimum-phase rational model $H(z)$ associated with $x(k)$. However, the initial conditions for computing the output $e(k)$ of the recursive filter $V(z)$ are unknown and must be preassigned. In this correspondence, we propose an improved inverse filtering method which simultaneously estimates the coefficients of $V(z)$ as well as the initial conditions. The resultant power spectral estimator with the initial conditions being estimated outperforms that with the initial conditions wrongly set to zero as the time constant of $V(z)$ is comparable to the number of data. Finally, we show some simulation results which support that the performance of the former is superior to the performance of the latter.**

I. Introduction

Spectral estimation of a wide-sense stationary process $x(k)$ can be found in various science and engineering areas such as speech processing, image processing, radar, sonar, seismology, biomedicine, radio astronomy, and oceanography. Various parametric spectral estimators were developed which have less variance (i.e., less statistical fluctuations) and less bias (i.e., higher resolution) than classical Fourier-based spectral estimators as the number of data is finite and limited. References [2] and [3] offer details of the advantage of parametric spectral estimators with respect to classical ones. Autoregressive moving average (ARMA) spectral estimators form the most general class among parametric spectral estimators. They are based on the assumption that the data $x(k)$, $k = 0, 1, \cdots, N-1$ were generated from a minimum-phase recursive model as follows:

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

where $u(k)$ is a white noise sequence of zero mean and variance σ^2. This model has the following rational transfer function $H(z)$:

$$H(z) = \frac{B(z)}{A(z)} = \frac{1 + \sum_{i=1}^{q} b_i z^{-i}}{1 + \sum_{i=1}^{p} a_i z^{-i}}. \tag{2}$$

The estimated power spectrum, $\hat{P}_{xx}(f)$, of $x(k)$ is obtained by

$$\hat{P}_{xx}(f) = \hat{\sigma}^2 |\hat{H}(z = e^{j2\pi f})|^2 \tag{3}$$

where $\hat{\sigma}^2$ and $\hat{H}(z)$ denote the estimates of σ^2 and $H(z)$, respectively, obtained from data by a selected optimality criterion.

Manuscript received August 27, 1990; revised May 22, 1991. This work was supported by the National Science Council under Grant NSC79-0404-E-007-18.

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

IEEE Log Number 9200243.

Assume that (p, q) is known *a priori*. Let

$$\theta_a = (a_1, a_2, \cdots, a_p)' \tag{4}$$

and

$$\theta_b = (b_1, b_2, \cdots, b_q)'. \tag{5}$$

It is well known [1] that θ_a, θ_b, and σ^2 can be estimated by minimizing the following objective function:

$$J_r = \sum_{k=r}^{N-1} e^2(k) \tag{6}$$

where $e(k)$ is the output residual of the recursive whitening inverse filter $V(z) = 1/H(z)$ as follows:

$$e(k) = -\sum_{i=1}^{q} b_i e(k-i) + x(k) + \sum_{i=1}^{p} a_i x(k-i). \tag{7}$$

The inverse filter $V(z)$ is nothing but a linear prediction error [4] filter of infinite order. The initial conditions $\{e(k), r-q \le k \le r-1\}$ and $\{x(k), r-p \le k \le r-1\}$ must be preassigned in order to compute $e(k)$. However, the former is unknown and the latter is available only when $r \ge p$. One can judiciously select J_n with

$$n = \max\{p, q\} \tag{8}$$

to be the objective function so that the unknown initial conditions for computing $e(k)$ are limited to

$$\theta_0 = (e(n-1), e(n-2), \cdots, e(n-q))'. \tag{9}$$

Since θ_0 is unknown, it is often wrongly set to zero in parametric spectral estimators involving inverse filtering of data. It is well known, from linear systems theory, that $e(k)$ for $k \ge n$ is the superposition of the zero-input response due to θ_0 and the zero-state response due to the input $\{x(k), k \ge (n-p)\}$. When $B(z)$ has some roots close to the unit circle such that the time constant τ of $V(z)$ is comparable to N, the zero-input response of $e(k)$ is significant compared with the zero-state response of $e(k)$ for $n \le k \le N-1$. Surely, the estimation accuracy of both θ_a and θ_b heavily depends on the unknown θ_0 for this case.

In this correspondence, in view of the case that the time constant τ of $V(z)$ is comparable to N, we propose a new iterative method, called an improved inverse filtering method, to simultaneously estimate θ_a, θ_b, and θ_0 by minimizing the objective function J_n. In Section II, we present this new method. Some simulation examples are then provided in Section III to support that its performance is superior to that associated with $\theta_0 = 0$. Finally, we draw some conclusions.

II. An Improved Inverse Filtering Method

Let us concatenate $e(k)$ given by (7) for $k = n, n+1, \cdots, N-1$ as the following vector form:

$$B_1 e_0 = -B_2 \theta_0 + A_1 x_{n-p} \tag{10}$$

where

$$x_j = (x(j), x(j+1), \cdots, x(N-1))' \tag{11}$$

1053-587X/92$03.00 © 1992 IEEE

$$\boldsymbol{e}_j = (e(n-j),\, e(n-j+1),\, \cdots,\, e(N-1-j))' \tag{12}$$

B_1 is an $(N-n) \times (N-n)$ matrix

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ b_1 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \cdots & & & & \cdots \\ b_q & & & \cdots & & & \\ 0 & b_q & & & \cdots & & \cdots \\ \vdots & & & & & \cdots & \cdots \\ 0 & \cdots & \cdots & b_q & \cdots & b_1 & 1 \end{bmatrix} \tag{13}$$

B_2 is an $(N-n) \times q$ matrix

$$B_2 = \begin{bmatrix} b_1 & b_2 & \cdots & b_q \\ b_2 & \cdots & b_q & 0 \\ \vdots & \vdots & & \\ b_{q-1} & b_q & \cdots & \\ b_q & 0 & \cdots & \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{14}$$

and A_1 is an $(N-n) \times (N-n+p)$ matrix

$$A_1 = \begin{bmatrix} a_p & a_{p-1} & \cdots & a_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & a_p & \cdots & \cdots & a_1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & & & & & & \\ \vdots & \vdots & & \cdots & & \cdots & & \cdots & \vdots \\ 0 & 0 & \cdots & \cdots & & 0 & a_p & \cdots & a_1 \quad 1 \end{bmatrix}. \tag{15}$$

$$\boldsymbol{e}_0 = -B_1^{-1}B_2\boldsymbol{\theta}_0 + B_1^{-1}A_1\boldsymbol{x}_{n-p}. \tag{16}$$

Notice, from (16), that the first term on the right-hand side is the zero-input response and the other term is the zero-state response. The objective function J_n (see (6)) can also be expressed as the following vector form:

$$J_n = \boldsymbol{e}_0'\boldsymbol{e}_0. \tag{17}$$

We, now, present an iterative block component method (BCM) [5] to minimize J_n. This iterative BCM consists of the following five-step procedure:

S1) find an initial guess for $\hat{\boldsymbol{\theta}}_b$;
S2) update $\hat{\boldsymbol{\theta}}_a$ using (21) below;
S3) update $\hat{\boldsymbol{\theta}}_0$ using (18) below;
S4) update $\hat{\boldsymbol{\theta}}_b$ using (22) below;
S5) if J_n converges, then stop; otherwise, go to S2).

Notice that for the case that $q = 0$ (and, consequently $n = p$), the proposed method reduces to the well-known covariance method [4] for AR spectral estimation, hence S3) and S4) are no longer needed. For the case that $p = 0$ (and, consequently, $n = q$), the proposed method becomes a MA spectral estimator with no need of S2). For the case that $p > 0$ and $q > 0$, the proposed method is an ARMA spectral estimator. The initial guess for $\hat{\boldsymbol{\theta}}_b$ can be obtained using any other appropriate parametric spectral estimation method. Whenever a block of unknown quantities (associated with S2)–S4)) is updated, J_n is guaranteed to decrease with the other unknown quantities fixed.

In our iterative BCM, both $\hat{\boldsymbol{\theta}}_a$ and $\hat{\boldsymbol{\theta}}_0$ are updated by closed-form formulas, whereas $\hat{\boldsymbol{\theta}}_b$ must be updated by a numerical search algorithm no matter whether the model is MA or ARMA. Therefore, an initial guess for $\hat{\boldsymbol{\theta}}_b$ is enough to initialize the proposed BCM. Next, we present how to update $\hat{\boldsymbol{\theta}}_0$, $\hat{\boldsymbol{\theta}}_a$, and $\hat{\boldsymbol{\theta}}_b$, respectively.

A. Updating $\hat{\boldsymbol{\theta}}_0$

From (18), one can see that $\boldsymbol{e}_0$ is linearly related to $\boldsymbol{\theta}_0$. Associated with this linear model, minimizing J_n given by (17) is a typical least squares estimation problem. Thus, we have [6] the least squares estimate $\hat{\boldsymbol{\theta}}_0$ as follows:

$$\hat{\boldsymbol{\theta}}_0 = \{B_2'B_1'^{-1}B_1^{-1}B_2\}^{-1}\{B_2'B_1'^{-1}\}B_1^{-1}A_1\boldsymbol{x}_{n-p}. \tag{18}$$

B. Updating $\hat{\boldsymbol{\theta}}_a$

The vector equation (16) can also be expressed as

$$\boldsymbol{e}_0 = B_1^{-1}(-B_2\boldsymbol{\theta}_0 + X\boldsymbol{\theta}_a + \boldsymbol{x}_n) \tag{19}$$

where $\boldsymbol{x}_n$ is given by (11) and X is an $(N-n) \times p$ matrix

$$X = \begin{pmatrix} x(n-1) & x(n-2) & \cdots & x(n-p) \\ x(n) & x(n-1) & \cdots & x(n-p+1) \\ x(n+1) & x(n) & \cdots & x(n-p+2) \\ \vdots & \vdots & \cdots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-1-p) \end{pmatrix}. \tag{20}$$

Again, from (19), we see that $\boldsymbol{e}_0$ is also linearly related to $\boldsymbol{\theta}_a$. Thus, we have the least squares estimate $\hat{\boldsymbol{\theta}}_a$ as follows:

$$\hat{\boldsymbol{\theta}}_a = \{X'B_1'^{-1}B_1^{-1}X\}^{-1}X'B_1'^{-1}B_1^{-1}(B_2\boldsymbol{\theta}_0 - \boldsymbol{x}_n). \tag{21}$$

C. Updating $\hat{\boldsymbol{\theta}}_b$

Since $e(k)$ (see (10)) for $k \geq n$ is a nonlinear function of $\boldsymbol{\theta}_b$, a popular approach for obtaining $\hat{\boldsymbol{\theta}}_b$ is to use a gradient-type iterative algorithm to update $\hat{\boldsymbol{\theta}}_b(i)$, in order to decrease J_n at every iteration.

We update $\hat{\boldsymbol{\theta}}_b(i+1)$ from $\hat{\boldsymbol{\theta}}_b(i)$ using a modified Newton–Raphson type algorithm as follows:

$$\hat{\boldsymbol{\theta}}_b(i+1) = \hat{\boldsymbol{\theta}}_b(i) - \rho H_i^{-1}\boldsymbol{g}_i \tag{22}$$

where $0 < \rho \leq 1$, $\boldsymbol{g}_i$ denotes the gradient

$$\boldsymbol{g}_i = \left.\frac{\partial J_n}{\partial \boldsymbol{\theta}_b}\right|_{\boldsymbol{\theta}_b = \hat{\boldsymbol{\theta}}_b(i)} = 2\sum_{k=n}^{N-1} e(k)\left.\frac{\partial e(k)}{\partial \boldsymbol{\theta}_b}\right|_{\boldsymbol{\theta}_b = \hat{\boldsymbol{\theta}}_b(i)} \tag{23}$$

and H_i denotes an approximate, i.e., pseudo-Hessian matrix [7]

$$H_i = \left.\frac{\partial^2 J_n}{\partial \boldsymbol{\theta}_b^2}\right|_{\boldsymbol{\theta}_b = \hat{\boldsymbol{\theta}}_b(i)} \cong 2\sum_{k=n}^{N-1}\left[\frac{\partial e(k)}{\partial \boldsymbol{\theta}_b}\right]\left[\frac{\partial e(k)}{\partial \boldsymbol{\theta}_b}\right]'\Bigg|_{\boldsymbol{\theta}_b = \hat{\boldsymbol{\theta}}_b(i)} \tag{24}$$

which is obtained by dropping the term involving the second derivative of $e(k)$ with respect to $\boldsymbol{\theta}_b$ in order to reduce computational load. One can easily show, by taking the partial derivative of (10) with respect to b_j, that

$$\frac{\partial \boldsymbol{e}_0}{\partial b_j} = -B_1^{-1}\boldsymbol{e}_j. \tag{25}$$

From (23)–(25), one can easily see that the jth component of $\boldsymbol{g}_i$ and the (j, k)th component of H_i are given by

$$[\boldsymbol{g}_i]_j = 2\boldsymbol{e}_0'\left.\frac{\partial \boldsymbol{e}_0}{\partial b_j}\right|_{\boldsymbol{\theta}_b = \hat{\boldsymbol{\theta}}_b(i)} = -2\boldsymbol{e}_0'B_1^{-1}\boldsymbol{e}_j|_{\boldsymbol{\theta}_b = \hat{\boldsymbol{\theta}}_b(i)} \tag{26}$$

and

$$[H_i]_{jk} \cong 2\left[\frac{\partial \boldsymbol{e}_0}{\partial b_j}\right]^t \left[\frac{\partial \boldsymbol{e}_0}{\partial b_k}\right]\Bigg|_{\theta_b = \hat{\theta}_b(i)} = 2\boldsymbol{e}_j' B_1'^{-1} B_1^{-1} \boldsymbol{e}_k|_{\theta_b = \hat{\theta}_b(i)} \quad (27)$$

respectively.

Some remarks are worth to be mentioned as follows. First of all, B_1 is an invertible lower triangular matrix and thus any matrix multiplications involving B_1^{-1} can be efficiently performed by forward substitutions rather than general matrix multiplications. Second, the roots of $\hat{B}(z)$ (or the poles of $\hat{V}(z)$) associated with $\hat{\theta}_b(i+1)$ must be checked if they are inside the unit circle because the recursive filtering by $\hat{V}(z)$ for computing such as $\boldsymbol{e}_0$, $\boldsymbol{g}_i$, and H_i is implicitly performed in the matrix computations involving B_1^{-1}. In other words, if $\hat{\theta}_b(i+1)$ obtained using (22) with $\rho = 1$ does not lead to the decrease of J_n with a minimum-phase $\hat{B}(z)$, one then has to reduce the value of ρ appropriately.

After $\hat{\theta}_a$, $\hat{\theta}_b$, and $\hat{\theta}_0$ are obtained, the least squares estimate of σ^2 can be easily seen to be equal to

$$\hat{\sigma}^2 = \frac{1}{N-n} J_n(\hat{\theta}_a, \hat{\theta}_b, \hat{\theta}_0). \quad (28)$$

III. Simulation Examples

Two simulation examples are to be presented to support that the performance of the proposed iterative BCM with θ_0 to be estimated is superior to that as θ_0 is set to zero (i.e., S3) removed) for the same data. The synthetic data $x(k)$ used were Gaussian and the data length was $N = 100$. For each simulation example, 30 realizations of spectral estimates were obtained. The 30 estimates are plotted in an overlaid fashion to indicate the variability of the resultant spectral estimator. The average (shown as a thin dot-dashed curve for θ_0 to be estimated and a thick dot-dashed curve for $\theta_0 = 0$) of 30 estimates is also shown together with the true power spectral density (shown as a solid curve). The initial guess for $\hat{\theta}_b$ was obtained using the Durbin's algorithm [8] either from $x(k)$ for the case of $p = 0$ (MA case) or from the residual series $y(k) = x(k) * \hat{a}_k$ for the case of $p > 0$ (ARMA case) where $\hat{a}_0 = 1$ and $\hat{a}_k$, $k = 1, 2, \cdots, p$, were obtained using the least squares modified Yule-Walker equations method [9], [10].

A. Example 1. (MA Case)

The transfer function $H_1(z)$ was a MA model with $q = 2$ as follows:

$$H_1(z) = 1 - 0.3673z^{-1} + 0.9604z^{-2}$$

which has a pair of strong zeros located at $z = 0.98e^{\pm j2\pi(0.22)}$. The simulation results are shown in Fig. 1(a) where $\hat{\theta}_0$ was estimated and Fig. 1(b) where $\hat{\theta}_0 = 0$. A spectral null can be observed from these two figures. As we predicted for this case where the time constant τ_1 of $V_1(z) = 1/H_1(z)$ is about $49 \approx N/2$, the variance associated with the results shown in Fig. 1(a) is much smaller than that shown in Fig. 1(b). From Fig. 1(c) one can observe that the bias associated with the results shown in Fig. 1(a) is also smaller than the bias associated with the results shown in Fig. 1(b) with a maximum of 5-dB bias improvement at the spectral null.

B. Example 2. (ARMA Case)

$H_2(z)$ was selected to be an ARMA model with $(p, q) = (2, 2)$ as follows:

$$H_2(z) = \frac{1 - 1.5102z^{-1} + 0.9604z^{-2}}{1 - 0.4467z^{-1} + 0.81z^{-2}}$$

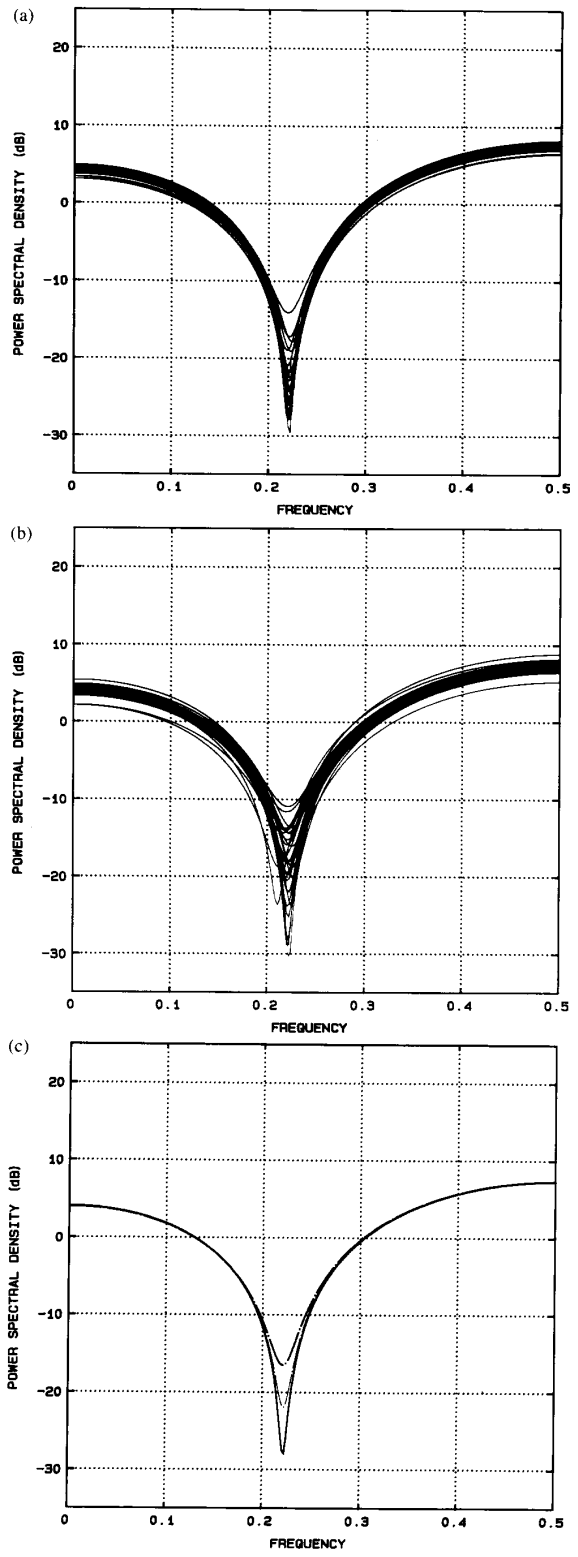

Fig. 1. Simulation results for example 1 (MA case). (a) Overlaid realizations associated with θ_0 being estimated; (b) overlaid realizations associated with $\theta_0 = 0$. (c) Average (thin dot-dashed curve) of realizations associated with (a); average (thick dot-dashed curve) of realizations associated with (b); and the true power spectral density (solid curve).

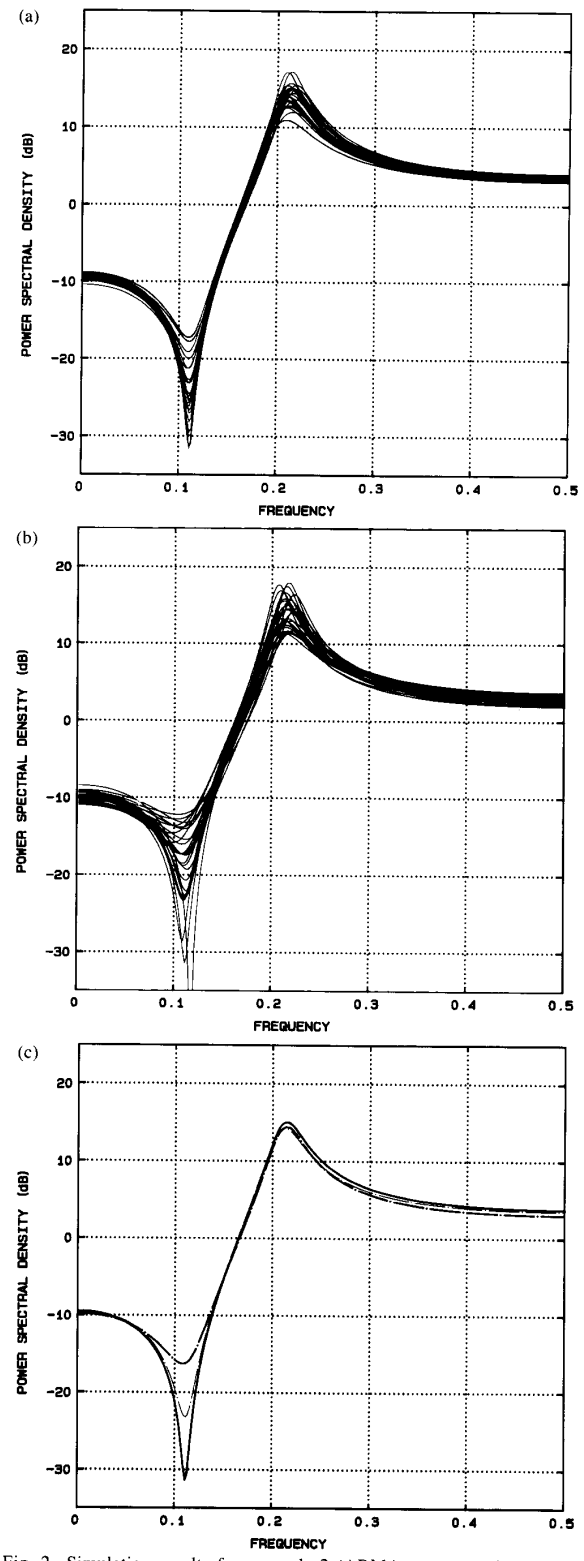

Fig. 2. Simulation results for example 2 (ARMA case). (a) Overlaid realizations associated with θ_0 being estimated; (b) overlaid realizations associated with $\theta_0 = 0$. (c) Average (thin dot-dashed curve) of realizations associated with (a); average (thick dot-dashed curve) of realizations associated with (b); and the true power spectral density (solid curve).

which has a pair of strong zeros located at $z = 0.98e^{\pm j2\pi(0.11)}$ and a pair of poles located at $z = 0.9e^{\pm j2\pi(0.21)}$. The simulation results are shown in Fig. 2(a) where $\hat{\theta}_0$ was estimated and Fig. 2(b) where $\hat{\theta}_0 = 0$. From these two figures, one can observe a spectral null and a spectral peak. As we predicted for this case where the time constant τ_2 of $V_2(z) = 1/H_2(z)$ is about the same as $\tau_1 \approx N/2$, the variance associated with the results shown in Fig. 2(a) is much smaller than that shown in Fig. 2(b). From Fig. 2(c), one can observe that the bias associated with the results shown in Fig. 2(a) at the proximity of the spectral null is also much smaller than the bias associated with the results shown in Fig. 2(b) with a maximum of 7-dB bias improvement at the spectral null.

The previous two examples support that the proposed method works well and that when $H(z)$ has strong zeros such that the time constant of the associated inverse filter is comparable to N, the initial conditions for computing the output residual of the inverse filter should be taken into account in estimating the power spectral density.

Although the previous simulation results indicate that the inverse filtering method can improve its performance as the initial conditions are taken into account, two general issues of iterative optimization algorithms including convergence to false local minima and computational load are worth to be discussed in the following.

Durbin's algorithm always provided a good initial guess for $\hat{\theta}_b$. Only very few realizations (which, however, were not included in the previous simulation results) happened for which the objective function J_n converged to a local minimum for both the inverse filtering method and the improved inverse filtering method. The number of iterations spent by the latter typically ranged between 15 and 20 for one realization and it was larger than that spent by the former by about 1 to 5. Moreover, for each iteration the computation expense for the improved inverse filtering method was also larger than that for the inverse filtering method due to extra moderate computations for estimating initial conditions.

IV. Conclusions

For a wide-sense stationary process $x(k)$, when the associated rational transfer function model $H(z)$ has strong zeros such that the time constant of the inverse filter $V(z) = 1/H(z)$ is comparable to number of data N, the initial conditions for computing the output residual $e(k)$ of $V(k)$ significantly affect the estimation accuracy of $\hat{V}(z)$. We have presented a new iterative block component method which can simultaneously estimate $\hat{V}(z)$ as well as the initial conditions. The presented simulation results showed that the resultant spectral estimator with the initial conditions being estimated has much smaller variance and smaller bias than that with the initial conditions forced to zero as the time constant of $V(z)$ is comparable to N. The study reported in this correspondence also indicates that some performance degradation to any spectral estimators involving the inverse filtering of data without estimating the initial conditions is inevitable as the number of data is not much larger than the time constant of the associated inverse filter. As a final remark, when $x(k)$ is Gaussian the proposed spectral estimation method is also a maximum-likelihood method.

Acknowledgment

The authors are grateful for the valuable suggestions by the anonymous reviewers. The research described in this correspondence was performed at the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China.

References

[1] S. A. Tretter and K. Steiglitz, "Power-spectrum identification in terms of rational models," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 185-188, Apr. 1967.

[2] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[3] S. L. Marple, *Digital Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[4] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561-580, Apr. 1975.
[5] J. M. Mendel, *Optimal Seismic Deconvolution: An Estimation-Based Approach*. New York: Academic, 1983.
[6] H. Sorensen, *Parameter Estimation*. New York: Marcel Dekker, 1980.
[7] J. Mendel, *Lessons in Digital Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[8] N. K. Gupta and R. K. Mehra, "Computational aspects of maximum-likelihood estimation and reduction of sensitivity function calculations," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 774-783, 1974.
[9] J. Durbin, "Efficient estimation of parameters in moving-average models," *Biometrika*, vol. 46, pp. 306-316, 1959.
[10] J. Cadzow, "High performance spectral estimation—a new ARMA method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 524-529, Oct. 1980.
[11] J. Durbin, "Estimation of parameters in time series regression models" *Stat. Soc.*, vol. 22, pp. 139-153, 1960.

On the Complexity of IQML Algorithms

Michael P. Clark and Louis L. Scharf

***Abstract*—In this correspondence we study the computational complexity of two methods for solving least squares and maximum likelihood modal analysis problems. In particular, we consider the Steiglitz-McBride and iterative quadratic maximum likelihood (IQML) algorithms. Recently, McClellan and Lee have shown the iterations of the two methods to be equivalent. However, they suggest that the Steiglitz-McBride algorithm may be computationally preferable. We provide in this correspondence a method for reducing the dimension of the matrix inversion required at each iteration of IQML. The resulting reduction in the computation makes the computational complexity of IQML commensurate with that of the Steiglitz-McBride algorithm.**

I. Introduction

Iterative methods for solving the least squares, or maximum likelihood, modal analysis problem have been developed by Kumaresan *et al.* [1], and Bresler and Macovski [2]. These methods are collectively known as iterative quadratic maximum likelihood (IQML). Predating these schemes is the filter design method of Evans and Fischl [3]. All of these methods employ the same iteration to solve a multidimensional optimization problem. Recently, McClellan and Lee [4] have shown that all of these methods are exactly equivalent to the Steiglitz-McBride [5] iteration. However, they indicate that the Steiglitz-McBride technique might be preferred because it is computationally less expensive. Kumaresan *et al.* [1] proposed a method for reducing the computation of the expensive matrix inversion which is required at each iteration of IQML. However, Behrens [6] has shown this method to be unstable under certain conditions. In fact, using the notation in [1], if $\boldsymbol{a}$ (the vector of amplitudes) is real and symmetric, and both M (the number of modes) and N (the number of samples) are odd, the circulant matrix constructed by the method is singular. The problem arises because the inverse of this circulant matrix is required. We present in the next section a stable method which further simplifies the computation of IQML, bringing it to a level commensurate with that of the Steiglitz-McBride algorithm.

Manuscript received July 10, 1991; revised February 4, 1992. This work was supported by the Office of Naval Research, Statistics and Probability Branch, under Contract N00014-89-J-1070.

The authors are with the University of Colorado, Boulder, CO 80309-0425.

IEEE Log Number 9200247.

II. The Deterministic Maximum Likelihood Problem

In this section we give a brief overview of the IQML method presented in [1] and [2]. Consider a signal composed of p damped, sinusoidal modes, $a_i z_i^n$, in additive, Gaussian white noise. An N-dimensional vector of samples of this signal is

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{a} + \boldsymbol{n} \tag{1}$$

where

$$H = [\Psi(z_0) \quad \Psi(z_1) \quad \cdots \quad \Psi(z_{p-1})] \tag{2}$$

$$\Psi = [1 \quad z \quad z^2 \quad \cdots \quad z^{p-1}]^T \tag{3}$$

$$\boldsymbol{a} = [a_0 \quad a_1 \quad \cdots \quad a_{p-1}]^T \tag{4}$$

and $\boldsymbol{n}$ is a normally distributed random variable with mean zero and variance $\sigma^2 I$. Let $B(z)$ be a degree-p polynomial

$$B(z) = \sum_{k=0}^{p} b_k z^{-k}. \tag{5}$$

The maximum likelihood mode estimates are the roots of the polynomial, $B(z)$, whose coefficients solve the optimization problem:

$$\min \boldsymbol{b}^* Y^* (BB^*)^{-1} Y \boldsymbol{b} \tag{6}$$

where B^* denotes the Hermitian transpose of the matrix B. Here the Toeplitz matrix

$$B = \begin{bmatrix} b_p & \cdots & b_1 & b_0 & & 0 \\ & \ddots & & \ddots & \ddots & \\ 0 & & b_p & \cdots & b_1 & b_0 \end{bmatrix} \tag{7}$$

is $(N - p) \times N$-dimensional, and the Toeplitz data matrix

$$Y = \begin{bmatrix} y_p & y_{p-1} & \cdots & y_1 & y_0 \\ y_{p+1} & y_p & \cdots & y_2 & y_1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ y_{N-1} & y_{N-2} & \cdots & y_{N-p} & y_{N-p-1} \end{bmatrix} \tag{8}$$

is $(N - p) \times (p + 1)$-dimensional. The vector of coefficients of the prediction polynomial $B(z)$,

$$\boldsymbol{b} = [b_0 \quad b_1 \quad \cdots \quad b_p]^T \tag{9}$$

is $(p + 1)$-dimensional. The parameter b_0 is normally restricted to be unity. This is the *nontriviality constraint*. The minimization problem requires the computation of both $(BB^*)^{-1}$ and $(Y^*(BB^*)^{-1}Y)^{-1}$ (neglecting, for now, the fact that the coefficients of $B(z)$ are usually constrained). The first matrix to be inverted is of size $(N - p) \times (N - p)$ while the second is $(p + 1) \times (p + 1)$. As the size of the data sample N is generally much larger than the number of modes to be estimated, p, the majority of the algorithmic complexity lies in the first inversion. However, the matrix BB^*, which is a moving average correlation matrix, has a wealth of structure. This structure led Kumaresan *et al.* [1] to an algorithm which reduces the computation required to find $(BB^*)^{-1}$. The

1053-587X/92$03.00 © 1992 IEEE