AN IMPROVED INVERSE FILTERING METHOD FOR PARAMETRIC SPECTRAL ESTIMATION

Chong-Yung Chi and Diing Wang

Department of Electrical Engineering
National Tsing Hua University
Hsinchu, Taiwan, Republic of China

Abstract:

For a wide-sense stationary process x(k), it is well-known [1] that its power spectral 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 paper, we propose an improved inverse filtering method which simultaneously estimates the coefficients of V(z) as well as the initial conditions. By simulation, the resultant power spectral estimator with the initial conditions being estimated is demonstrated to have much smaller variance as well as much smaller bias than that with the initial conditions wrongly set to be zero as the time constant of V(z) is comparable to the number of data.

I. INTRODUCTION

Spectral estimation of a wide-sense stationary process x(k) can be found in various science and engineering areas. Various parametric spectral estimators were developed [2,3] which have less variance (i.e. less statistical fluctuations) and less bias (i.e. higher resolution) than conventional Fourier based spectral estimators [2,3] as the number of data is finite and limited. 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, ..., N-1 were generated from a minimum-phase recursive model as follows:

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

where u(k) is a white noise with variance σ^2. This model has the following rational transfer function H(z)

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

where A(z) and B(z) denote the denominator polynomial and the numerator polynomial of H(z), respectively. The estimated power spectrum, $\hat{P}_{xx}(f)$, of x(k) is obtained by

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

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

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

$$\underline{\theta}_a = (a_1, a_2, ..., a_p)' \qquad (4)$$

and

$$\underline{\theta}_b = (b_1, b_2, ..., b_q)'. \qquad (5)$$

It is well-known [1] that $\underline{\theta}_a$, $\underline{\theta}_b$ and σ^2 can be estimated by minimizing the following objective function:

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

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

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

The inverse filter V(z) is nothing but a linear prediction error 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≥p. One can judiciously select J_n with

$$n = \max\{p,q\} \qquad (8)$$

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

$$\underline{\theta}_0 = (e(n-1), e(n-2), ..., e(n-q))'. \qquad (9)$$

Due to lack of the value of $\underline{\theta}_0$, it is often wrongly set to be zero in parametric spectral estimators involving inverse filtering of data. It is well-known, from linear systems theory, that e(k) for k≥n is the superposition of the zero-input response due to $\underline{\theta}_0$ 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≤k≤N-1. Surely, the estimation accuracy of $\underline{\theta}_a$ and $\underline{\theta}_b$ heavily depends on the unknown initial value of $\underline{\theta}_0$ for this case.

In this paper, 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 $\underline{\theta}_a$, $\underline{\theta}_b$ and $\underline{\theta}_0$ by minimizing J_n.

II. AN IMPROVED INVERSE FILTERING METHOD

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

$$B_1\, \underline{e}_0 = -B_2\, \underline{\theta}_0 + A_1\, \underline{x}_{n-p} \qquad (10)$$

where

$$\underline{x}_j = (x(j), x(j+1), ..., x(N-1))', \qquad (11)$$

$$\underline{e}_j = (e(n-j), e(n-j+1), ..., e(N-1-j))', \qquad (12)$$

B_1 is an (N-n)×(N-n) matrix and B_2 is an (N-n)×q matrix

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ b_1 & 1 & 0 & & \\ \vdots & b_1 & 1 & & \\ b_q & \vdots & & \ddots & \\ 0 & b_q & & & \\ \vdots & & & & \\ 0 & \cdots & b_q & \cdots & b_1\ 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} b_1 & b_2 & \cdots & b_q \\ b_2 & b_3 & \cdots & b_q\ 0 \\ \vdots & \vdots & & \\ b_q & & & \\ 0 & & & \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

and A_1 is an (N-n)×(N-n+p) matrix

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

Thus,

$$\underline{e}_0 = -B_1^{-1} B_2\, \underline{\theta}_0 + B_1^{-1} A_1\, \underline{x}_{n-p}. \qquad (13)$$

Notice, from (13), 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:

$$J_n = \underline{e}_0'\, \underline{e}_0. \qquad (14)$$

We, now, present an iterative block component method (BCM) [4] to search for a local minimum of J_n. This iterative BCM consists of the following 5-step procedure:

(S1) find an initial guess for $\hat{\underline{\theta}}_b$;

(S2) update $\hat{\underline{\theta}}_a$ using (18) below;

(S3) update $\hat{\underline{\theta}}_0$ using (15) below;

(S4) update $\hat{\underline{\theta}}_b$ using (19) below;

(S5) if J_n converges, then stop; otherwise, go to (S2).

Next, we present how to update $\hat{\underline{\theta}}_0$, $\hat{\underline{\theta}}_a$ and $\hat{\underline{\theta}}_b$, respectively.

A. Updating $\hat{\underline{\theta}}_0$:

From (13), one can see that $\underline{e}_0$ is linearly related to $\underline{\theta}_0$. Thus, we have the least-squares estimate $\hat{\underline{\theta}}_0$ as follows [4]:

$$\hat{\underline{\theta}}_0 = \{ B_2'\, B_1'^{-1} B_1^{-1} B_2 \}^{-1} \{ B_2'\, B_1'^{-1} \}\, B_1^{-1} A_1\, \underline{x}_{n-p}. \qquad (15)$$

B. Updating $\hat{\underline{\theta}}_a$:

The vector equation (13) can also be expressed as

$$\underline{e}_0 = B_1^{-1}(-B_2\, \underline{\theta}_0 + X\, \underline{\theta}_a + \underline{x}_n) \qquad (16)$$

whose $\underline{x}_n$ is given by (11) and X is an (N-n)×p matrix

$$X = \begin{bmatrix} x(n-1) & x(n-2) & \cdots & x(n-p) \\ x(n) & x(n-1) & \cdots & x(n-p+1) \\ x(n+1) & x(n) & & x(n-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-1-p) \end{bmatrix}. \quad (17)$$

Again, from (16), we see that $\underline{e}_0$ is also linearly related to $\underline{\theta}_a$. Thus, we have the least-squares estimate $\hat{\underline{\theta}}_a$ as follows [4]:

$$\hat{\underline{\theta}}_a = \{ X' B_1^{-r-1} B_1^{-1} X \}^{-1} X' B_1^{-r-1} B_1^{-1} (B_2 \underline{\theta}_0 - \underline{x}_n). \quad (18)$$

C. Updating $\hat{\underline{\theta}}_b$:

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

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

$$\hat{\underline{\theta}}_b(i+1) = \hat{\underline{\theta}}_b(i) - \rho H_i^{-1} \underline{g}_i \quad (19)$$

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

$$[\underline{g}_i]_j = 2 \underline{e}_0' \frac{\partial \underline{e}_0}{\partial b_j}\bigg|_{\underline{\theta}_b=\hat{\underline{\theta}}_b(i)} = -2 \underline{e}_0' B_1^{-1} \underline{e}_j \bigg|_{\underline{\theta}_b=\hat{\underline{\theta}}_b(i)} \quad (20)$$

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

$$[H_i]_{jk} \cong 2 \left[\frac{\partial \underline{e}_0}{\partial b_j}\right]' \left[\frac{\partial \underline{e}_0}{\partial b_k}\right] \bigg|_{\underline{\theta}_b=\hat{\underline{\theta}}_b(i)} = 2 \underline{e}_j' B_1^{-1} B_1^{-1} \underline{e}_k \bigg|_{\underline{\theta}_b=\hat{\underline{\theta}}_b(i)}. \quad (21)$$

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

$$\hat{\sigma}^2 = \frac{1}{N-n} J_n(\hat{\underline{\theta}}_a, \hat{\underline{\theta}}_b, \hat{\underline{\theta}}_0). \quad (22)$$

III. SIMULATION EXAMPLE

We now present a simulation example to support that the performance of the proposed iterative BCM with $\underline{\theta}_0$ to be estimated is superior to that as $\underline{\theta}_0$ is wrongly set to be zero (i.e. (S3) removed). 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 our spectral estimator. The average (shown as a dot-dashed curve) of 30 estimates is also shown together with the true power spectral density (shown as a solid curve). The initial guess for $\hat{\underline{\theta}}_b$ was obtained using the Durbin's algorithm [2] from the residual series $y(k) = x(k) * \hat{a}_k$ where $\hat{a}_k$, k=1, 2, ..., p, were obtained using the least squares modified Yule-Walker equations method [2].

The transfer function H(z) was selected to be an ARMA model with (p,q)=(2,2) as follows:

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

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

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(z) significantly affect the estimation accuracy of $\hat{V}(z)$. In this paper, the order (p,q) of H(z) was assumed to be known a priori although it needs to be determined from data [2,3]. We have presented a new iterative block component method which can simultaneously estimate $\hat{V}(z)$ as well as the initial conditions. By simulation, we demonstrated that our power spectral estimator has much smaller variance and smaller bias than that with the initial conditions wrongly set to be zero as the time constant of V(z) is comparable to N. The study reported in this paper also indicates that some performance degradation to any spectral estimator 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.

REFERENCES

[1] S. A. Tretter and K. Steiglitz, "Power-Spectrum identification in terms of rational models," IEEE Trans. on Automatic Control, vol. AC-12, pp. 185-188, April 1967.

[2] S. M. Kay, Modern spectral estimation, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[3] S. L. Marple, Digital spectral analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[4] J. M. Mendel, Lessons in digital estimation theory, Prentice-Hall Englewood Cliffs, New Jersey, 1987.

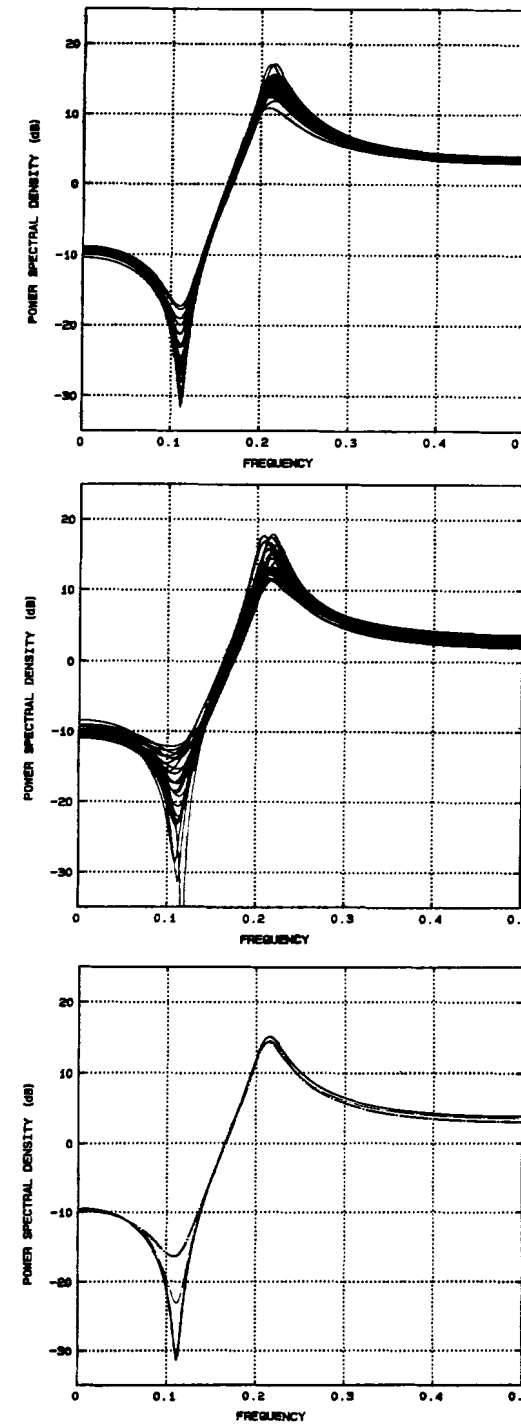

Figure 1. (a) Overlaid realizations associated with $\underline{\theta}_0$ being estimated; (b) overlaid realizations associated with $\underline{\theta}_0=0$ (c) Average (thin dot-dashed curve) of realizations associated with (a); average (thick dot-dashed curve) of realizations associated with (b); and the true power spectral density (solid curve).