Technical University of Denmark

DTU

Maximum likelihood estimation of the parameters of nonminimum phase and noncausal ARMA models

Rasmussen, Klaus Bolding

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

rather than

$$\begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

3) Equation (24) is missing a $\sqrt{2}$. This equation should read

$$\begin{bmatrix} H^{(m)}(z) \\ (z^{2(N-1)}G)^{(m)}(z) \end{bmatrix} = \\ \sqrt{2}[V^{(m)}(z)h_0(z) + mV^{(m-1)}(z)h_0^{(1)}(z)]. \tag{1}$$

4) The matrix V_0 defined at the beginning of Section III-B should be

$$V_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}$$

rather than

$$V_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

5) Equations (32)–(35) are missing a $\sqrt{2}$. These equations should read

$$c_0 = \sqrt{2} \sin \theta_1 \sin \left(\theta_1 - \frac{\pi}{4} \right) \tag{2}$$

$$c_1 = \sqrt{2} \sin \theta_1 \sin \left(\theta_1 + \frac{\pi}{4} \right) \tag{3}$$

$$c_2 = \sqrt{2} \cos \theta_1 \sin \left(\theta_1 + \frac{\pi}{4} \right) \tag{4}$$

$$c_3 = \sqrt{2} \cos \theta_1 \sin \left(\frac{\pi}{4} - \theta_1 \right). \tag{5}$$

Maximum Likelihood Estimation of the Parameters of Nonminimum Phase and Noncausal ARMA Models

Klaus Bolding Rasmussen

Abstract—The well-known prediction-error-based maximum likelihood (PEML) method can only handle minimum phase ARMA models. This paper presents a new method known as the back-filtering-based maximum likelihood (BFML) method, which can handle nonminimum phase and noncausal ARMA models. The BFML method is identical to the PEML method in the case of a minimum phase ARMA model, and it turns out that the BFML method incorporates a noncausal ARMA filter with poles outside the unit circle for estimation of the parameters of a causal, nonminimum phase ARMA model.

I. INTRODUCTION

Parameter estimation for stochastic signal models has always received much attention in signal and image processing. It has proved usable in low-rate coding of speech signals [1], in deconvolution of seismic signals [2], and in many other applications [3]. The most used stochastic signal model is the Gaussian, minimum phase,

autoregressive (AR) model [3]. The popularity of this model is due to the fact that the maximum likelihood estimators of the parameters can be computed analytically. Maximum likelihood estimation of the parameters of the Gaussian, minimum phase, autoregressive moving average (ARMA) model is also well-known [4], [5], and maximum likelihood estimation has been done for the nonGaussian, minimum phase AR-model as well [6]. This has all been done using the well-known prediction-error-based maximum likelihood (PEML) method and has therefore been limited to minimum phase ARMA models. In this correspondence, a new algorithm known as the back-filtering-based maximum likelihood (BFML) method is presented, and it is able to handle nonminimum phase and noncausal ARMA models. The BFML method is therefore an alternative to the higher order statistics approach [7], [8].

II. THE BFML METHOD

We will assume the following stochastic signal model for the given signal $y(n)$:

$$y(n) = h(n) * e(n) = \sum_{i=-\infty}^{+\infty} h(i)e(n-i) \tag{1}$$

where $e(n)$ denotes the driving stochastic signal, and $n = 1, \ldots, N$ is the discrete-time index. $e(n)$ is independent identically distributed with probability density function $f_{e(n)}(e(n))$. The filter $h(n)$ does not have to be minimum phase or even causal. The first task for deriving a maximum likelihood estimator is to find the likelihood function $f_y(y; \theta)$ as a function of the parameter vector θ. The second task is to derive a fast optimization algorithm. The idea behind the BFML method is to interpret the filtering in (1) as a matrix transformation from $e(n)$ to $y(n)$. This transformation is linear:

$$y = \mathcal{H}e \tag{2}$$

where e and y are $N \times 1$ column vectors. $\mathcal{H}$ is a $N \times N$ nonsymmetric Toeplitz matrix. It is known (see Sec. 8-1 of [9]) that

$$f_y(y; \theta) = \frac{\prod_{n=1}^{N} f_{e(n)}(\epsilon(n; \theta); \theta)}{|\det(\mathcal{H})|} \tag{3}$$

where the back-filtered sequence $\epsilon(n; \theta)$ as an $N \times 1$ column vector is given by

$$\epsilon(\theta) = \mathcal{H}(\theta)^{-1} y. \tag{4}$$

Note that the back-filtered sequence depends on θ; see Fig. 1. The BFML estimate θ_{BFMLE} of the true parameter vector θ_{TRUE} is

$$\theta_{\text{BFMLE}} = \arg \max \frac{\prod_{n=1}^{N} f_{e(n)}(\epsilon(n; \theta); \theta)}{|\det(\mathcal{H})|}. \tag{5}$$

For the purpose of finding a convenient expression for the likelihood function, we will introduce the z-transform $H(z)$ of $h(n)$:

$$H(z) = \sum_{n=-\infty}^{+\infty} h(n)z^{-n}. \tag{6}$$

Now, let $H(z)$ be of ARMA form

$$H(z) = H_1(z)H_2(z) \tag{7}$$

$$H_1(z) = \frac{B_1(z)}{A_1(z)} \tag{8}$$
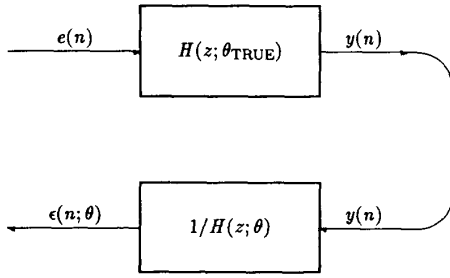
$$H_2(z) = \frac{B_2(z)}{A_2(z)} \tag{9}$$

Fig. 1.   Difference between $e(n)$ and $\epsilon(n;\theta)$.

$$B_1(z) = 1 + \sum_{n=1}^{N_{b_1}} b_1(n) z^{-n} \tag{10}$$

$$A_1(z) = 1 + \sum_{n=1}^{N_{a_1}} a_1(n) z^{-n} \tag{11}$$

$$B_2(z) = 1 + \sum_{n=1}^{N_{b_2}} b_2(n) z^{n} \tag{12}$$

$$A_2(z) = 1 + \sum_{n=1}^{N_{a_2}} a_2(n) z^{n} \tag{13}$$

where $H_1(z)$ is minimum phase, i.e., all poles and zeros are inside the unit circle. $H_2(z)$ is maximum phase, i.e., all poles and zeros are outside the unit circle. All z-transforms with no poles or zeros on the unit circle can be written on the form $A z^r H(z)$, where A is a real number, and r is an integer. By convention, we will include the gain A and delay z^r in the description of $f_{e(n)}(e(n))$ rather than in $h(n)$. A z-transform defined by (6) with poles outside the unit circle does not describe an unstable filter. It describes a stable, noncausal filter (see Ch. 4 of [10]). It should be noticed, however, that if the filter corresponding to a z-transform with poles outside the unit circle is implemented in the straightforward causal way, then the implementation is unstable. The filtering or transformation from $e(n)$ to $y(n)$ is implemented as a causal recursive filtering with $H_1(z) = (B_1(z))/(A_1(z))$ followed by an anticausal recursive filtering with $H_2(z) = (B_2(z))/(A_2(z))$. The back-filtering or transformation from $y(n)$ to $\epsilon(n;\theta)$ is implemented as an anticausal recursive filtering with $1/(H_2(z;\theta)) = (A_2(z;\theta))/(B_2(z;\theta))$ followed by a causal recursive filtering with $1/(H_1(z;\theta)) = (A_1(z;\theta))/(B_1(z;\theta))$. The transformation matrix corresponding to the causal filter $H_1(z)$ is lower triangular with ones in the diagonal, and the transformation matrix corresponding to the anticausal filter $H_2(z)$ is upper triangular with ones in the diagonal. Consequently, $|\det(\mathcal{H})| = 1$, and the BFML estimator θ_{BFMLE} is therefore simplified to

$$\theta_{\mathrm{BFMLE}} = \arg\max \prod_{n=1}^{N} f_{e(n)}(\epsilon(n;\theta);\theta) \tag{14}$$

where

$$\epsilon(n;\theta) = \frac{y(n)}{H(z;\theta)}. \tag{15}$$

It is further possible to extend the BFML method to the case of an ARMA filter followed by an invertible, memoryless nonlinearity [11]. The maximization of the likelihood function is done by an iterative Gauss–Newton-type algorithm; see [11] for details. In [11], it is also shown how this can be done recursively.

III. Model Order Selection

In practice, the numbers of minimum/maximum phase poles/zeros are not known *a priori*, and they also have to be estimated. These numbers are fixed during the iterative Gauss–Newton optimization, and they have to be determined by some other means. The Gaussian probability distribution plays a central role in this respect. To begin with, if the input $e(n)$ is Gaussian distributed, then it is not possible to differentiate minimum phase poles from maximum phase poles by using conventional estimation techniques, and the same applies to zeros. This is because the Gaussian likelihood function is a function of the autocovariance only. On the other hand, it can be proved that it is possible to estimate the ARMA filter except for a pure delay and gain factor for all nonGaussian distributions of the input $e(n)$ [12]. Furthermore, in [6], it is proved that the variance of the maximum likelihood estimator of the parameters of a causal AR model is largest in the Gaussian case. This result is true also in the general case of a noncausal ARMA model as the proof in [6] can be extended to this case easily. The model order selection problem for the BFML approach does not differentiate much from the usual PEML case. In fact, almost any one of the traditional approaches, e.g., Akaike's information theoretic criterion, can be used. An overview of these approaches is given in Ch. 16 of [5]. One possible approach is to select several promising models and then optimize for each model. The resulting minima should then be tested for statistical significance. If the situation in this model order selection approach is that the number of zeros (or poles) is given, say N_b, then the number of possibilities, i.e., combinations of minimum/maximum phase zeros, that must be tested is $N_b + 1$.

IV. Example

In this section, an example of how the BFML method works is presented. It has not been the intention that the example should represent any specific application. It is only meant to illustrate the procedure of the new method. Let the input $e(n)$ be zero mean Laplacian distributed:

$$f_{e(n)}(e(n)) = \frac{\exp\left(-\frac{|e(n)|}{\beta}\right)}{2\beta}; \quad -\infty < e(n) < +\infty; \beta > 0. \tag{16}$$

The variance of $e(n)$ is $V\{e(n)\} = 2\beta^2$. Now, let the true model be

$$y(n) = \frac{1 + 2z^{-1}}{1 + 0.5z^{-1}} e_1(n) \tag{17}$$

where $e_1(n)$ is zero mean Laplacian distributed with $\beta = 1$. The variance of $y(n)$ is $V\{y(n)\} = 8$. The filter is of the all-pass type with a pole at -0.5 and a zero at -2. The filter is transformed to the form appropriate for the BFML method

$$y(n) = \frac{1 + 0.5z}{1 + 0.5z^{-1}} e(n) \tag{18}$$

where $e(n) = 2z^{-1} e_1(n)$ is zero mean Laplacian distributed with $\beta = 2$. The negative log-likelihood function is

$$V(\theta) = \sum_{n=1}^{N} \left(\frac{|\epsilon(n;\theta)|}{\beta} + \log(\beta) + \log(2) \right). \tag{19}$$

A simulation with 1024 samples was performed using (17) for generation of data $y(n)$ and (18) for the identification. The initial guess

$$a_1(1) = 0.25$$
$$b_2(1) = 0.25$$

TABLE I
THE NEGATIVE LOG-LIKELIHOOD FUNCTION $V(\theta)$ FOR DIFFERENT
LAPLACIAN ARMA MODELS AND DIFFERENT SIGNAL-TO-NOISE RATIOS SNR

SNR		$+\infty$	10dB	0dB
$H(z;\theta)$	β	$V(\theta)$		
$\frac{1+0.5z^{-1}}{1+0.5z^{-1}}$	2	2408	2470	2976
$\frac{1+0.5z}{1+0.5z^{-1}}$	2	2331	2401	2917
$\frac{1+0.5z^{-1}}{1+0.5z}$	2	2426	2477	2944
$\frac{1+0.5z}{1+0.5z}$	2	2408	2470	2976

TABLE II
THE BIAS INCREASES AS THE SIGNAL-TO-NOISE RATIO SNR DECREASES
FOR THE LAPLACIAN, NONMINIMUM PHASE ARMA MODEL

SNR	$a_1(1)$	$b_2(1)$	β
θ_{TRUE}	0.5000	0.5000	2.0000
$+\infty$	0.5067	0.5404	1.9600
10dB	0.4849	0.5031	2.1067
0dB	0.4948	0.4971	3.1567

resulted in $V = 2328$ for

$$a_1(1) = 0.5070 \quad \pm \quad 0.0189$$
$$b_2(1) = 0.5461 \quad \pm \quad 0.0185$$
$$\beta = 1.9602 \quad \pm \quad 0.0625$$

after ten iterations. The values after $\pm$ are the estimated standard deviations.

Now, white Gaussian noise is added to the measurements $y(n)$. Table I shows the negative log-likelihood function $V(\theta)$ for the true ARMA model and the corresponding ARMA models with mirrored poles and zeros. For all ARMA models, the $V(\theta)$'s are shown for different signal-to-noise ratios. We have that only differences between the $V(\theta)$'s greater than $\chi^2_{95\%}(3)/2 \approx 4$ are statistical significant at level 5% [11]. It is seen that the correct ARMA model can be identified even for very low signal-to-noise ratios. As noticed in Section III, this would not be possible using the usual Gaussian likelihood function even in the noise-free case. Table II shows the BFML estimates using the true values as the initial guess. It is seen that the BFML estimates become a little biased as the signal-to-noise ratio decreases.

V. CONCLUSION

A parameter estimation method known as the BFML method, which can handle nonminimum phase and noncausal ARMA models, has been derived. The likelihood function is found interpreting filtering as a matrix transformation. The BFML method is identical to the usual PEML method in the case of a minimum phase ARMA filter, and the BFML method can therefore be seen as an extension of the PEML method. Although the BFML method is derived assuming

noiseless measurements, it performed well even for low signal-to-noise ratios.

REFERENCES

[1] P. E. Papamichalis, *Practical Approaches to Speech Coding*.Englewood Cliffs, NJ: Prentice-Hall, 1987.
[2] D. G. Stone, "Wavelet estimation," *Proc. IEEE*, vol. 72, pp. 1394–1402, 1984.
[3] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[4] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
[5] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[6] D. Sengupta and S. Kay, "Efficient estimation of parameters for non-Gaussian autoregressive processes," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 785–794, 1989.
[7] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, pp. 278–305, 1991.
[8] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, pp. 869–891, 1987.
[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 1984.
[10] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[11] K. B. Rasmussen, "Two-dimensional deconvolution of ultrasound images," Ph.D. dissertation, Electronics Institute, Technical Univ. of Denmark, 1992.
[12] K. -S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Annals Stat.*, vol. 10, pp. 1195–1208, 1982.

On the Estimation of Bispectral Density Function in the Case of Randomly Missing Observations

M. M. Gabr and T. Subba Rao

Abstract—The estimation of the bispectral density function of a stationary stochastic process when some observations are missing according to a point binomial distribution is considered in the paper. The existence of the bispectrum, and its estimation are considered. The methods are illustrated with simulated examples.

I. INTRODUCTION AND DEFINITIONS

Second order spectra plays an important role in signal processing and the need for power spectral analysis arises in a variety of contexts such as the design of optimal filters, detection of narrow band signals,