ARMA ORDER SELECTION FOR EEG - AN EMPIRICAL COMPARISON
OF THREE ORDER SELECTION ALGORITHMS

Prasad Gannabathula

I.S.N. Murthy

Department of Electrical Engineering
Indian Institute of Science
Bangalore 560 012, INDIA

ABSTRACT

The performance of three ARMA order estimation algorithms, i) Canonical Correlation analysis b) S-array and c) Franke algorithm on simulated and real EEG signals is presented. It is shown that the S-array always correctly indicates the AR order but makes incorrect estimates of the MA order. The canonical correlation method identifies the model order correctly for simulated data, and overestimates the AR order on real data. The Franke algorithm is shown to perform poorly in comparison to the other two algorithms.

INTRODUCTION

Modelling the EEG signals using ARMA models requires a priori knowledge of the orders of the polynomials and have to be estimated from the data itself. A simple method would be to examine the spectrum of the signal. A normal EEG has at most four component waves, the delta, alpha, beta and theta waves with frequency ranges of 0-10Hz, 5-14Hz, 10-28Hz and 2-7Hz respectively. For an EEG signal with a peak at the origin and two peaks between 1 to 30 Hz in the spectrum, Zetterberg suggested the use of an ARMA (5,4) model. However, in practice the spectrum of real EEG may not show all the component waves. This may happen when a component wave has low energy and wide bandwidth, overlapping with the other waves, or is entirely absent. In such cases the orders of the AR and MA polynomials are not immediately obvious or apparent.

In this paper we consider this problem of ARMA model identification when applied to simulated and real EEG signals and study the performance of three general algorithms to estimate the order. The algorithms allow the signal to be weakly stationary (i.e.) allow the system function to have poles on the unit circle. We are concerned only with order estimation, the actual estimation of the coefficients is not dealt with.

ARMA IDENTIFICATION

All the algorithms that identify the ARMA model use two different approaches. In the first, the order is estimated by examining some function of the residual variance of the signal, estimated by maximizing the likelihood function of a pth order AR and qth order MA model. These algorithms are computationally expensive as a number of optimal ARMA models of different orders have to be examined. In spite of this, there is no assurance that the correct model will be identified. These algorithms were not used in the study.

In the second class of algorithms, two types of algorithms, which do not require optimal estimates of the residual variance or the coefficients are used. In the Franke algorithm [2] the residual variance is recursively estimated by Levinson recursion. The suboptimal estimates thus obtained are used in the AIC or BIC. The other two algorithms used, exploit the correlation properties of the data to estimate the order. In all these algorithms, it is assumed that the signal is weakly stationary (i.e.) has all its poles within the unit circle. This is to take into account the finite precision arithmetic used to compute the correlation coefficients and their functions for signals with poles which are arbitrarily close to but still inside the unit circle. We have used two algorithms, one S-array as proposed by Gray et.al. [3] and the other using canonical correlation analysis [4]. The S-array of Gray also indicates the number of poles which are on the unit circle. Details of these two algorithms are given in [5]. In the next section the performance of these three algorithms when applied to simulated and real EEG data is presented. For simulated data, the effect of data size, the position of poles and zeros is also investigated.

APPLICATION TO SIMULATED EEG

The EEG signals were generated a) by passing Gaussian white noise through a single fifth order AR and fourth order MA filter

and b) as the sum of the outputs of three independent filters each driven by an independent white noise. A sampling rate of 100 was used. The EEG signals were simulated with various centre frequencies and bandwidths and with varying energy contents in the component waves.

The S array identifies the AR order correctly even when the data size is as small as 2.5 seconds and the poles are close to the unit circle. The MA order is however, not always correctly identified. In particular, when the EEG signal is filtered to remove the effect of a pole(s) on or very close to the unit circle, the identified MA order is far too low. Further, when the pole has large bandwidth and low energy the AR order is incorrectly identified. When the bandwidths of alpha and beta wave are between 0.65 and 1.275 and at least 12% energy in each wave, the S array gives a correct estimate of the AR order but underestimates the MA order by one. When bandwidths of both alpha and beta waves are large ($> 2\text{Hz}$), and the energy in either wave is less than half the energy in the other, the two waves are not separately identifiable and the S array underestimates the AR order.

The canonical correlation analysis correctly identifies the AR portion of the model for data of 5 seconds duration or more. The estimated AR coefficients were consistent and stable. For the MA part the order was correctly estimated for data of more than 5 seconds duration. For records of 5 second duration, the MA order was underestimated by one. The MA order is also incorrectly identified when the component waves are very close to each other, or if they have wide bandwidths, and if one wave masks the presence of other. Otherwise the algorithm is insensitive to the percentage energy under the curve and the location of the poles.

An interesting feature of this algorithm is that it is insensitive to the method of simulation of EEG when the half power points of the component waves are small ($< 0.7\text{Hz}$ for the alpha and beta waves and $< 1.25\text{Hz}$ for the delta wave).

Franke's algorithm overestimated the AR order and underestimated the MA order. For all data lengths of 125, 250 and 500 samples, the estimated AR order is either 7 or 8 while the MA order varied from 0 to 3, and is unaffected by the pole location or the energy under the component wave.

4 REAL DATA

The algorithms were applied on a number of normal EEG signals. The maximum length of data which can be considered stationary was determined by using the run test [6].

This length was found to be 11 seconds. The three algorithms were run on records of 11, 10, 7.5, and 5 seconds duration.

The S array works well for records as short as 2.5 seconds. The AR orders were either 3 or 5. An order of 5 implies the presence of a beta wave, which was not visible in the spectra of the signals. The MA orders are highly variable and range from 1 to 4. In most cases the S array identified a pole very close to the unit circle.

The canonical correlation overestimates the AR order. The MA order as 4. In most cases the AR order identified is 6. This is higher than the order identified by the S array. Further, the minimum data size needed to get an estimate of the order is 7.5 seconds, compared to the 2.5 seconds required by the S array. Assuming the presence of a beta component, that could not be identified by the S array, an extra pole has been identified. In all cases the estimated AR coefficients were found to be consistent. Filtering the data to remove the effect of frequencies beyond 25Hz identifies models that are in close agreement with the models identified by the S array. The AR coefficients, thus estimated are however inconsistent.

The Levinson-Durbin recursion in all cases performed poorly. A 7th or 8th order AR model was identified. The MA portion, if identified, had an order of one or two.

CONCLUSIONS

Three algorithms, Franke's algorithm for ARMA models, the S array and the canonical correlation analysis were used for identification of the ARMA model for the EEG signal. The algorithms were tested on a large number of simulated and real EEG data.

REFERENCES

1. L.H.Zetterberg, "Estimation of parameters for linear difference equation with application to EEG", *Math. Biosci.*, 5, pp227-275, 1969
2. J.Franke, "A Levinson-Durbin recursion for Autoregressive-Moving Average process" *Biometrika*, 72, 3, pp573-581, 1985
3. A.Gray et.al, "A new approach to ARMA modelling", *Comm. of Stat.*, B, 7, pp1-77, 1978
4. R.S.Tsay and G.C.Tiao, "Use of Canonical analysis in time series model identification", *Biometrika*, 72, 2, pp299-315, 1985.
5. G.S.S.Durga Prasad and I.S.N.Murthy: Technical report, Indian Institute of Science. 1988
6. J.B.Bendat and A.G.Pearsol, *Random data analysis*, Wiley, New York, 1974.