

# A Bayesian Approach for the Estimation of AR Coefficients from Noisy Biomedical Data

Vangelis P. Oikonomou and Dimitrios I. Fotiadis, *Senior Member IEEE*

**Abstract**—In this paper we study the identification of AR parameters in a biomedical signal corrupted by additive white gaussian noise. The identification of AR parameter is treated as a signal estimation problem, whose aim is to obtain an estimate of the clean signal, given the noisy observations, and after that to obtain the noise free AR parameters. The novelty of our approach is the simultaneous estimation of AR parameter and the model order of the AR process. This is done adopting a Bayesian framework and using a special form for the prior of AR parameters. To obtain the solution we use the Variational Bayesian (VB) Framework. Simulation results have shown that the proposed approach correctly identifies the model order of AR process while at the same time produces an estimate for the AR parameters.

## I. INTRODUCTION

The autoregressive (AR) model can be used in a number of applications such as linear predictive coding (LPC) of speech, spectral estimation and biomedical signal processing [1], [2]. When the signal is observed in noise the AR parameters are biased and produce a smoother spectral estimator. Several methods have been proposed to overcome this difficulty. They include:

- Application of general ARMA estimators
- Filtering of data to reduce the observation noise
- Compensation of the AR parameters or the reflection coefficients estimates for the biasing effect of noise
- Use a higher order for the AR model

Details on the above approaches can be found in [2] and the references therein. The proposed method belongs to the filtering approaches.

The first approach to estimate the AR parameters in a noisy environment was proposed by Lim [3] for speech enhancement. In this approach the noisy signal is modelled as AR process corrupted by additive white Gaussian noise. An iterative scheme is employed for the successive estimation of the AR coefficients and the clean speech. A slightly different approach for the estimation of the AR parameters is to use the Kalman filter/smoothing [4], where the noisy AR model is written in a state-space form and after that the Kalman Filter is used, in an EM algorithm, to estimate the noise-free signal and the AR parameters [5]. The Kalman Filter replaces the noncausal Wiener Filter of Lim approach with a Kalman Smoother or a Kalman Filter in the case of a recursive algorithm. However, all the above algorithms assumed that

V.P. Oikonomou and D.I. Fotiadis are with the Unit of Medical Technology and Intelligent Informations Systems, Department of Computer Science, University of Ioannina, Ioannina, Greece {voikonom, fotiadis}@cs.uoi.gr

the model order of the AR process is known or estimated separately from the AR parameters.

The model order selection is a difficult problem. To find the model order of an AR process the MDL [2] or the BIC criterion [6] can be used. However, this is a separate process from the estimation of AR parameters and applied after the estimation of AR coefficients for each model order. In this work we propose a method where the AR parameters and the model order are inferred simultaneously. In [7] a Bayesian approach for parameter estimation and model order selection in an AR process is introduced based on the Variational Bayesian Methodology. For model order selection the negative free energy was used for each model order and the maximum of negative free energy provides us the correct model order. The above algorithm was applied in the AR process without the presence of additive disturbing noise. In [8] a Bayesian approach for time-varying AR process based on Monte Carlo Markov Chain (MCMC) is studied. However, the problem with the MCMC methods is that they do not provide with closed form solutions and there is not a global accepted measure to guarantee convergence of the algorithm [9].

In this work we propose a Bayesian model for the estimation of the AR parameters and the model order in the presence of additive white gaussian noise. For the identification of the model parameters the Variational Bayesian Framework is used. This results in an EM-like algorithm, in the same spirit as the algorithms which uses a Wiener or Kalman Filter i.e. first estimation of the noise - free signal from noisy observations and after that identification of AR parameters. To identify the correct model order we introduce a special form of priors over the AR parameters. This form of prior is called Automatic Relevance Determination (ARD) [10] and it has been used extensively in machine learning. In what follows we present the proposed algorithm. Next we present simulation results which are compared with the Lim approach [3]. Finally, conclusions and future directions are discussed.

## II. METHODOLOGY

Throughout this paper, the desired signal  $s(n)$  is represented by a  $p$ -th order process:

$$\begin{aligned} s(n) &= \sum_{k=1}^p a(k)s(n-k) + e(n) & (1) \\ &= \mathbf{a}^T \mathbf{s}(n) + e(n), & (2) \end{aligned}$$

where  $\mathbf{a} = [a(1), a(2), \dots, a(p)]$  denotes the AR parameter vector and  $\mathbf{s}(n) = [s(n-1), s(n-2), \dots, s(n-p)]$  is the

signal vector. The driving noise (or innovation noise)  $e(n)$  of the AR model is assumed to be zero mean white Gaussian noise with variance  $\sigma_e^2$ . Observations  $y(n)$  of the desired signal  $s(n)$  may be contaminated according to:

$$y(n) = s(n) + v(n), \quad (3)$$

where the disturbing noise  $v(n)$  is uncorrelated to  $s(n)$  and is zero mean white Gaussian noise with variance  $\sigma_v^2$  and  $n = 1, 2, \dots, N$ , where  $N$  is the number of samples. Collecting all the samples together in vector form we have for the AR model:

$$\mathbf{s} = \mathbf{S}\mathbf{a} + \mathbf{e}, \quad (4)$$

where  $\mathbf{S}$  is an  $N \times p$  matrix containing the past observations,  $\mathbf{s}$  is an  $N \times 1$  vector containing the current observations and  $\mathbf{a} = [a(1), a(2), \dots, a(p)]$  is the vector of AR coefficients. Also, the above equation can be written as:

$$\mathbf{A}\mathbf{s} = \mathbf{e}, \quad (5)$$

where the matrix  $\mathbf{A}$  is a  $N \times (p+1)$  matrix of AR coefficients and  $\mathbf{s}$  is the vector of the observations.

Similarly, for the observations we have:

$$\mathbf{y} = \mathbf{s} + \mathbf{v}. \quad (6)$$

where  $\mathbf{y}$  is the noisy observations,  $\mathbf{s}$  the AR model observations and  $\mathbf{v}$  the additive disturbing noise. If we know the variance of the disturbing noise  $\mathbf{v}$ , then we can obtain an estimate of the AR coefficients using an EM-like algorithm. In this algorithm, first an estimate of the desired signal  $\mathbf{s}$  is obtained from the noisy observations,  $\mathbf{y}$ , with a Wiener filter. Second, this estimate is fed to the Yule-Walker equations to obtain an estimate for the AR coefficients and the variance of the innovation noise,  $\mathbf{e}$ . However, this approach presents two difficulties: first, we must know the order of the AR coefficients and second we must also know the variance of the disturbing noise. One approach presented in [11] solves the problem of knowing the variance of the disturbing noise using a Bayesian approach based on the Bayesian Evidence framework [10]. In this work, we present an algorithm where both problems mentioned above are solved. The framework we use is the Variational Bayesian Methodology [9].

The parameters of interest are  $\theta = [\mathbf{a}, \sigma_e^2, \sigma_v^2]$ . From now, we will use the inverse variances as parameters so the vector of parameters is  $\theta = [\mathbf{a}, \lambda_e, \lambda_v]$ , where  $\lambda_v = \frac{1}{\sigma_v^2}$  and  $\lambda_e = \frac{1}{\sigma_e^2}$ . To set up a Bayesian framework for our problem, we must define prior distributions over the parameters  $\theta$ . The disturbing noise,  $\mathbf{v}$ , follows zero mean white gaussian distribution with inverse variance  $\lambda_v$ :

$$p(\mathbf{v}|\lambda_v) = N(0, \lambda_v^{-1}\mathbf{I}). \quad (7)$$

The innovation noise  $\mathbf{e}$  follows zero mean gaussian distribution with inverse variance  $\lambda_e$ :

$$p(\mathbf{e}|\lambda_e) = N(0, \lambda_e^{-1}\mathbf{I}). \quad (8)$$

The crucial choice about prior is related to the AR parameters. We encode the preference for less complex model by

making the choice of a zero mean Gaussian distribution over the AR parameters  $\mathbf{a}$ :

$$p(\mathbf{a}|\mathbf{w}) = \prod_{i=1}^p N(0, w_i^{-1}), \quad (9)$$

where  $\mathbf{w}$  is a vector of  $p$  hyperparameters and  $w_i$  is the  $i$ -th component of the  $\mathbf{w}$ . More specifically, there is an individual hyperparameter associated independently with every coefficient. To complete the specification of this hierarchical prior, we must specify hyperpriors over the hyperparameters  $\mathbf{w}$ , as well as the inverse variances,  $\lambda_e$  and  $\lambda_v$ . Suitable priors over these quantities are Gamma distributions

$$p(\mathbf{w}) = \prod_{i=1}^p \Gamma(w_i; b, c), \quad (10)$$

$$p(\lambda_e) = \Gamma(\lambda_e; b_{\lambda_e}, c_{\lambda_e}), \quad (11)$$

$$p(\lambda_v) = \Gamma(\lambda_v; b_{\lambda_v}, c_{\lambda_v}), \quad (12)$$

where

$$\Gamma(x; b, c) = \frac{1}{\Gamma(c)} \frac{x^{c-1}}{b^c} \exp\left(-\frac{x}{b}\right). \quad (13)$$

where  $b$  and  $c$  are parameters entering the Gamma distribution. This formulation of prior of the AR coefficients is known as Automatic Relevance Determination (ARD) [10]. In the Variational Bayesian (VB) Framework we are interested to maximize the marginal likelihood  $p(\mathbf{y})$ . This marginal likelihood can be bounded from below using the Jensen's inequality. In our formulation the noise-free signal  $\mathbf{s}$  are the hidden data and the parameters are  $\theta = [\mathbf{a}, \lambda_e, \lambda_v, \mathbf{w}]$ . The use of prior distribution for the AR coefficients introduces the vector  $\mathbf{w}$  of hyperparameters, which control the order of the AR model. In the VB Framework we want to approximate the true posterior  $p(\mathbf{x}, \theta|\mathbf{y})$  with an approximation distribution  $q(\mathbf{s}, \theta)$ , which is constrained to have a separable form  $q(\mathbf{s}, \theta) = q(\theta)q(\mathbf{s})$ . The application of the VB framework results in an iterative algorithm, under the same spirit with the EM algorithm [12]. The solution for  $q(\theta)$  and  $q(\mathbf{s})$  is:

$$q^{(t+1)}(\mathbf{s}) \propto \exp \langle \ln p(\mathbf{y}, \mathbf{s}|\theta) \rangle_{q^{(t)}(\theta)}, \quad (14)$$

$$q^{(t+1)}(\theta) \propto p(\theta) \exp \langle \ln p(\mathbf{y}, \mathbf{s}|\theta) \rangle_{q^{(t+1)}(\mathbf{s})}. \quad (15)$$

Details on those solutions can be found in [9]. The above equations give a general solution of our problem. Next we obtain the specific solution for the hidden variables  $\mathbf{s}$  and the parameters  $\theta$ . We assume that the prior distribution of parameters has the separable form  $p(\theta) = p(\mathbf{a}|\mathbf{w})p(\mathbf{w})p(\lambda_v)p(\lambda_e)$ . Now applying the VB Framework we obtain the following:

$$q(\mathbf{s}) = N(\hat{\mathbf{s}}, C_{\hat{\mathbf{s}}}), \quad (16)$$

$$q(\mathbf{a}|\mathbf{w}) = N(\hat{\mathbf{a}}, C_{\hat{\mathbf{a}}}), \quad (17)$$

$$q(\mathbf{w}) = \prod_{i=1}^p \Gamma(\hat{w}_i; b_i, c_i), \quad (18)$$

$$q(\lambda_e) = \Gamma(\hat{\lambda}_e; \hat{b}_{\lambda_e}, \hat{c}_{\lambda_e}), \quad (19)$$

$$q(\lambda_v) = \Gamma(\hat{\lambda}_v; \hat{b}_{\lambda_v}, \hat{c}_{\lambda_v}), \quad (20)$$

where

$$C_{\hat{s}} = (\hat{\lambda}_v \mathbf{I} + \hat{\lambda}_e \langle \mathbf{A}^T \mathbf{A} \rangle_{q(\alpha)})^{-1}, \quad (21)$$

$$\hat{\mathbf{s}} = C_{\hat{s}} \hat{\lambda}_v \mathbf{y}, \quad (22)$$

$$b_i = \frac{1}{(\frac{1}{2}a^2(i) + \frac{1}{b})}, i = 1, 2, \dots, p, \quad (23)$$

$$c_i = \frac{1}{2} + c, i = 1, 2, \dots, p, \quad (24)$$

$$\hat{w}_i = b_i c_i, i = 1, 2, \dots, p, \quad (25)$$

$$tb = \frac{1}{2}(\mathbf{a}^T \mathbf{a} + Tr(C_{\hat{a}} \hat{\mathbf{S}}^T \hat{\mathbf{S}}) + \mathbf{a}^T \hat{\mathbf{S}}^T \hat{\mathbf{S}} \mathbf{a} - 2\mathbf{a}^T \hat{\mathbf{S}} \hat{\mathbf{s}}) + \frac{1}{b_{\lambda_e}}, \quad (26)$$

$$\hat{b}_{\lambda_e} = \frac{1}{tb}, \quad (27)$$

$$\hat{c}_{\lambda_e} = \frac{N}{2} + c_{\lambda_e}, \quad (28)$$

$$\hat{\lambda}_e = \hat{b}_{\lambda_e} \hat{c}_{\lambda_e}, \quad (29)$$

$$\mathbf{C}_a = (\hat{\lambda}_e \hat{\mathbf{S}}^T \hat{\mathbf{S}} + \mathbf{w} \mathbf{I})^{-1}, \quad (30)$$

$$\mathbf{a} = \hat{\lambda}_e \mathbf{C}_a \hat{\mathbf{S}}^T \hat{\mathbf{s}}, \quad (31)$$

$$\hat{b}_{\lambda_v} = \frac{1}{\frac{1}{2}(\mathbf{y} - \hat{\mathbf{s}})^T (\mathbf{y} - \hat{\mathbf{s}}) + \frac{1}{b_{\lambda_v}}}, \quad (32)$$

$$\hat{c}_{\lambda_v} = \frac{N}{2} + c_{\lambda_v}, \quad (33)$$

$$\hat{\lambda}_v = \hat{b}_{\lambda_v} \hat{c}_{\lambda_v}. \quad (34)$$

In the above the matrix  $\mathbf{I}$  is the identity matrix having dimension  $pxp$ . The matrix  $\hat{\mathbf{S}}$  is the matrix,  $\mathbf{S}$ , of Eq. (4) but the estimated signal  $\hat{\mathbf{s}}$  is used for its construction. Also, in the simulation results which follow we approximate the matrix  $\langle \mathbf{A}^T \mathbf{A} \rangle_{q(\alpha)}$  with the matrix  $\langle \mathbf{A}^T \rangle_{q(\alpha)} \langle \mathbf{A} \rangle_{q(\alpha)}$ . The matrix  $\langle \mathbf{A} \rangle_{q(\alpha)}$  is the matrix of Eq. (5), where for the construction we use the estimated AR coefficients. The learning algorithm consists from equations (21)-(34). These equations are applied iteratively.

### III. RESULTS

The proposed methodology is evaluated and compared with the method proposed in [3] using simulated data. Our goal is twofold: first we want to obtain a good estimate for the noise free signal  $\mathbf{s}$  and second to obtain the true order of the AR model. To measure how good is the estimate of the signal  $\mathbf{s}$  we use the total SNR output measure which is defined as:

$$SNR = 10 \log_{10} \frac{\sum_{n=0}^{N-1} s^2(n)}{\sum_{n=0}^{N-1} [s(n) - \hat{s}(n)]^2}, \quad (35)$$

where  $\hat{s}(n)$  is the estimated signal and  $N$  is the number of samples. While the results for the AR model order are given numerically and visually. In our experiments we use two AR models which have been chosen to represent spectra with narrowband and broadband features. In all experiment, we use uninformative priors by setting  $b = 10^3$  and  $c = 10^{-3}$  for each of the Gamma priors. The initial value for the variance of the noise is obtained as  $\sigma_n^2 = \frac{1}{2}(\mathbf{y} - \mathbf{Y}\hat{\mathbf{a}})^T (\mathbf{y} - \mathbf{Y}\hat{\mathbf{a}})$ , where  $\mathbf{Y}$  is a matrix similar to matrix  $\mathbf{S}$ , but is constructed

using the noisy observations. The vector  $\hat{\mathbf{a}}$  is the vector of AR coefficients when we use the noisy observations as an AR process. Both algorithms are stopped in 5 iterations, while for the Lim's approach we use the true variance of the noise in the Wiener Filter.

#### A. Experiment 1

A 8192 point sequence was created using an AR model with coefficients  $\mathbf{a} = [1, -1.352, 1.338, -0.662, 0.240]$  (broadband process). In order to allow initial transients to die out, the first 7936 points have been discarded, and the last 256 samples have been considered as an example of a stationary AR process. In this process additive noise for various SNR level has been added to obtain a noisy AR process. For each method 100 realizations of the AR process were obtained by computer simulation. The AR model order used in Lim method [3] and the proposed approach was a wrong model order  $p=20$ . The true model order is  $p=4$ . To show the ability of our method to select the appropriate model order we show the average of the 100 estimates of the Hinton diagram [6] of AR coefficients for the case of having noise 10dB. However, similar results are obtained in the cases of other SNR levels. The Hinton diagram is shown in Fig. 1a. In this diagram white indicates positive, black negative and the area is proportional to magnitude. We can observe that the use of ARD prior reduce the magnitude of the coefficients larger than 4. From this visualization we can conclude that the proposed approach suggests the model order that must be used for the observations is equal to 4, which is the true model order.

#### B. Experiment 2

A second experiment is performed using an AR model with coefficients  $\mathbf{a} = [1, -2.760, 3.809, -2.654, 0.924]$  (narrowband process). We use the same conditions as in the first experiment. The Hinton diagram for the AR coefficients is shown in Fig. 1b. Again, we can observe that the AR coefficients which correspond to an AR model with order larger than 4 have small magnitude. Our method indicates again model order equal to 4.

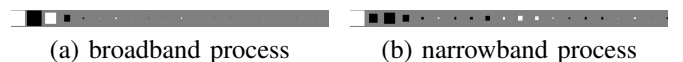
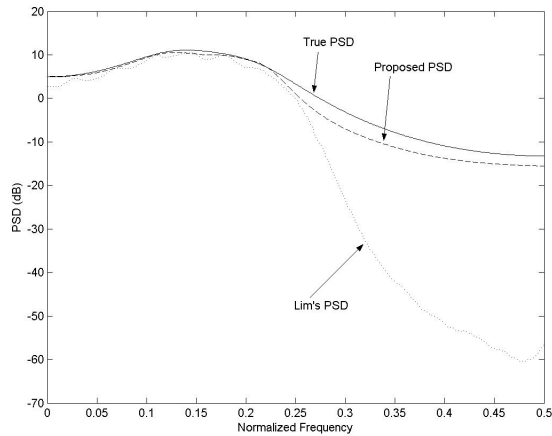
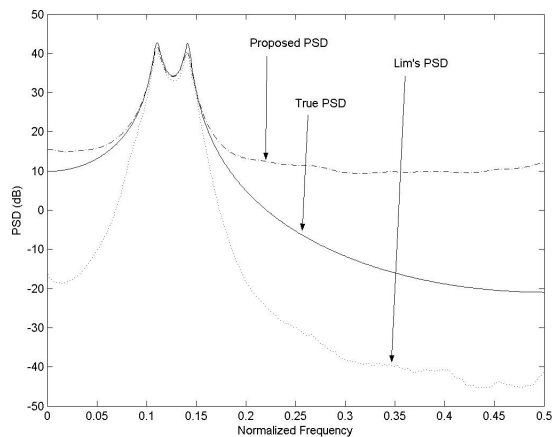


Fig. 1. Hinton diagrams for the broadband and narrowband processes

The average spectral estimator over 100 realizations obtained for each method is shown in Fig. (2) for SNR=10dB. We observe that for the narrowband process the two methods are able to identify the two spectral peaks with high accuracy. However, for the broadband process the Lim's method [3] present some oscillations due to the wrong model order used. These oscillations were not observed in the proposed method because our method is able to estimate the correct model order. The results for SNR improvement are shown in Table I for broadband and narrowband processes and for SNR values 0, 5, 10, 15 and 20 dB. We can see that the two methods



(a) broadband process



(b) narrowband process

Fig. 2. PSD for broadband and narrowband processes.

TABLE I  
THE OBTAINED SNR MEASURE

SNRin	Broadband		Narrowband	
	Our Method	Lim	Our Method	Lim
20 dB	19.5	20.1	22.6	21.9
15 dB	15.9	15.5	18.5	17.5
10 dB	11.6	11.1	13.6	13.8
5 dB	7.4	6.9	8.2	10.3
0 dB	2.1	2.6	2.6	6.6

present similar results except for SNR equals to 5 and 0 dB for the narrowband process. This inconsistency of our method can be justified by the fact that the proposed method estimates the variance of the disturbing noise while we use the true variance in the Lim's method [3].

#### IV. CONCLUSIONS

The AR model is a well known model with many applications in signal processing, and especially in the biomedical signal processing [1]. In this work we present an iterative algorithm for the estimation of the AR coefficients in noisy observations. The algorithm consists of two general steps in

a similar way with other approaches described in [3], [11]. The first obtains an estimate of the noise-free signal, and the second step, from this estimated signal obtains the AR coefficients. Compared to [3] and [11], the novelty of our method is the use of prior over the AR coefficients. We use a special form of prior called ARD. This type of prior tends to assign small values to parameters which are irrelevant to our problem. The use of this prior lead us to a Bayesian model for our problem. To obtain meaningful solutions in closed form we use the VB Framework [9]. The application of the VB Framework provide us with an iterative learning algorithm. The results presented here using simulated data indicate that the proposed approach gives the correct model order while at the same time produces an estimate of the noise free signal comparable to Lim's approach [3]. The estimation of the correct model order happens because the ARD prior assigns small values to AR coefficients which are irrelevant to the true AR model. In the future, we intend to use a different model for the noise such as colored gaussian noise and gaussian mixture distribution as well as a different prior model for the AR coefficients.

#### REFERENCES

- [1] M. Akay, *Biomedical Signal Processing*, Academic Press, San Diego, California, 1994.
- [2] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, Englewood Cliffs, NJ 07632, 1998.
- [3] J.S. Lim and A.V. Openheim, "All-pole modeling of degraded speech," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 6, pp. 197–210, 1978.
- [4] R.H. Shumway and D.S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, vol. 3, pp. 253–264, 1982.
- [5] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and Sequential Kalman Filter - Based Speech Enhancement Algorithms," *IEEE Transactions on Speech and Audio Processing*, vol. 6, pp. 373–385, 1998.
- [6] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, 1995.
- [7] W. Penny and S.J. Roberts, "Bayesian methods for autoregressive models," in *Int. Workshop on Neural Networks for Signal processing*, Sidney, Australia, 2000.
- [8] J. Vermaak, C. Adrieu, A. Doucet, and S.J. Godsill, "Particle methods for bayesian modeling and enhancement of speech signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 173–185, 2002.
- [9] M. Beal, *Variational Algorithms for Approximate Bayesian Inference*, Ph.D. thesis, Gatsby Computational Neuroscience Unit, Univ. College London, London, U.K., 2003.
- [10] D. MacKay, "Probable networks and plausible predictions - a review of practical bayesians methods for supervised neural networks," *Network: Computation in Neural Systems*, vol. 6, pp. 469–505, 1995.
- [11] G.M.K. Saleh and M. Niranjan, "Speech enhancement using a bayesian evidence approach," *Computer Speech and Language*, vol. 15, pp. 101–125, 2001.
- [12] A. Dempster, Laird A., and Rubin D., "Maximum likelihood from incomplete data via the em algorithm," *Journal of Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.