EXPECTATION-MAXIMIZATION (EM) ALGORITHM FOR INSTANTANEOUS FREQUENCY ESTIMATION WITH KALMAN SMOOTHER

Md. Emtiyaz Khan, D. Narayana Dutt

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore, India
phone: +91 80 22932742, email: emtiyaz@protocol.ece.iisc.ernet.in

ABSTRACT

In this paper, we propose an expectation-maximization (EM) algorithm based approach for instantaneous frequency (IF) estimation in a Kalman smoother framework. We formulate time-varying AR (TVAR) model as a state-space model and describe EM algorithm for model parameter estimation. This is used with Kalman smoother for IF estimation. We show that our scheme EMIF shows best performance among the other existing adaptive algorithms like RLS and LMS. Performance analysis is reported on a class of FM signals.

1. INTRODUCTION

Estimation of instantaneous frequency (IF) is of primary interest in many fields like wireless communications, speech, radar and underwater acoustics etc. Time-varying autoregressive (TVAR) model based IF estimation was first proposed in [6]. Later, Shan and Beex [1] applied basis function approach to a TVAR model. Various IF estimation algorithms, including the adaptive algorithms (RLS and LMS) were reviewed and compared by Boashash [7]. It was observed that performance of these algorithms is low because of suboptimal smoothing.

We propose an expectation-maximization (EM) algorithm based IF estimation (EMIF) in a Kalman smoother framework. We first formulate the TVAR model as a state-space model. Then we estimate the model parameters with EM algorithm. Once we have the model parameters, IF estimates are computed with a Kalman smoother. As EMIF falls in the category of adaptive algorithms, we compare it with RLS and LMS algorithms. We show that EMIF performs significantly better than the other adaptive algorithms.

2. TVAR AND STATE SPACE MODEL

Consider the time series $\{y_t\}_{t=1}^T$, following a time-varying AR (TVAR) model,

$$y_t = \sum_{k=1}^{p} a_k^t y_{t-k} + v_t \tag{1}$$

where, $\{a_k^t\}_{k=1}^p$ are the TVAR parameters, p is the model order and v_t is observation noise, $v_t \sim \mathcal{N}(0,\sigma_v^2)$. Define $\mathbf{x}_t \equiv [a_1^t a_2^t \ldots a_p^t]'$ and $\mathbf{h}_t \equiv [y_{t-1} y_{t-2} \ldots y_{t-p}]'$ ($'$ denotes transpose). The TVAR parameters are modeled as a multivariate AR(1) process, leading to the following state-space model,

$$y_t = \mathbf{h}_t' \mathbf{x}_t + v_t \tag{2}$$

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{w}_t \tag{3}$$

where $\mathbf{w}_t \sim \mathcal{N}(0,Q)$ is the state noise and A is $p \times p$ state transition matrix. Eq.(2)(3) is a class of linear dynamical system. We have,

$$p(y_t|\mathbf{x}_t,\mathbf{h}_t) = \mathcal{N}(\mathbf{h}_t'\mathbf{x}_t,\sigma_v^2) \tag{4}$$

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(A\mathbf{x}_{t-1},Q) \tag{5}$$

$$p(\mathbf{x}_1) = \mathcal{N}(\pi_1,\mathbf{V}_1) \tag{6}$$

Denote the model parameters as $\Theta \equiv \{A,\sigma_v^2,Q,\pi_1,V_1\}$. This kind of model has been previously studied in [2] and [5]. If Θ is known, $\mathbf{x}_t$ (the TVAR parameters) can be inferred using Kalman filter. The advantages of Kalman filter are well known [2]. As the usual practice in IF estimation is to process a block of data, we can use both the preceding and succeeding time series samples. This leads to the notion of a Kalman smoother.

However, in practice, we do not know Θ, so usually these parameters are set to arbitrary values or various *ad-hoc* /suboptimal procedures are used for their estimation. We use expectation-maximization (EM) algorithm for parameter estimation of our state-space model. Similar approach has been used in [3] and [4]. But our model differs from those in the sense that the observation vector $\mathbf{h}_t$ is not constant, but depends on the lagged output $\{y_{t-1},y_{t-2}\ldots,y_{t-p}\}$. We first derive the joint log-likelihood for our model and then describe EMIF.

2.1 Log-Likelihood

We define, $\mathcal{Y}_1^T = \{y_1,y_2,\ldots,y_T\}$ and $\mathcal{X}_1^T = \{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_T\}$. The joint likelihood is given by,

$$p(\mathcal{X}_1^T,\mathcal{Y}_1^T|\Theta) = p(\mathbf{x}_1)\prod_{t=2}^{T} p(\mathbf{x}_t|\mathbf{x}_{t-1})\prod_{t=1}^{T} p(y_t|\mathbf{x}_t,\mathbf{h}_t) \tag{7}$$

Using eq.(4)-(6), we get,

$$\begin{aligned}\log p(\mathcal{X}_1^T,\mathcal{Y}_1^T|\Theta) = & -\frac{T}{2}\ln\sigma_v^2 - \frac{1}{2\sigma_v^2}\sum_{t=1}^{T}(y_t-\mathbf{h}_t'\mathbf{x}_t)^2 \\ & -\frac{1}{2}\sum_{t=2}^{T}(\mathbf{x}_t-A\mathbf{x}_{t-1})'Q^{-1}(\mathbf{x}_t-A\mathbf{x}_{t-1}) \\ & -\frac{1}{2}(\mathbf{x}_1-\pi_1)'V_1^{-1}(\mathbf{x}_1-\pi_1) - \frac{1}{2}\ln|V_1| \\ & -\frac{T-1}{2}\ln|Q| - \frac{(p+1)T}{2}\ln 2\pi \end{aligned} \tag{8}$$

Clearly, the joint log-likelihood is function of $\mathbf{x}_t$, which is a hidden variable. The joint log-likelihood cannot be maximized directly. We use EM algorithm for maximization.

3. EM ALGORITHM FOR IF ESTIMATION (EMIF)

In E-step, we compute the expected log-likelihood, given observations $\mathscr{Y}_1^T$ and parameter estimate Θ^{i-1} from the previous iteration,

$$\mathscr{Q}(\Theta,\Theta^{i-1}) = E\left[\ln p(\mathscr{X}_1^T,\mathscr{Y}_1^T|\Theta)|\mathscr{Y}_1^T,\Theta^{i-1}\right] \quad (9)$$

We then maximize $\mathscr{Q}(\Theta,\Theta^{i-1})$ with respect to Θ, to obtain the parameter estimate Θ^i for the next iteration. The EM algorithm, as applied to the present problem is discussed below.

3.1 E-step

Computation of $\mathscr{Q}(\Theta,\Theta^{i-1})$ requires $E[\mathbf{x}_t|\mathscr{Y}_1^T]$, $E[\mathbf{x}_t\mathbf{x}_t'|\mathscr{Y}_1^T]$ and $E[\mathbf{x}_t\mathbf{x}_{t-1}'|\mathscr{Y}_1^T]$, which we denote by $\widehat{\mathbf{x}}_t, P_t$ and $P_{t,t-1}$ respectively. These are estimated with a Kalman smoother. Define, $\mathbf{x}_t^\tau \equiv E(\mathbf{x}_t|\mathscr{Y}_1^\tau)$ and $V_t^\tau \equiv Var(\mathbf{x}_t|\mathscr{Y}_1^\tau)$. We obtain the following Kalman filter forward recursions:

$$\begin{aligned}
\mathbf{x}_t^{t-1} &= A\mathbf{x}_{t-1}^{t-1} && (10)\\
V_t^{t-1} &= AV_{t-1}^{t-1}A' + Q && (11)\\
\mathbf{K}_t &= \frac{V_t^{t-1}\mathbf{h}_t}{\sigma_v^2 + \mathbf{h}_t' V_t^{t-1}\mathbf{h}_t} && (12)\\
\mathbf{x}_t^t &= \mathbf{x}_t^{t-1} + \mathbf{K}_t(y_t - \mathbf{h}_t'\mathbf{x}_t^{t-1}) && (13)\\
V_t^t &= (I - \mathbf{K}_t\mathbf{h}_t')V_t^{t-1} && (14)
\end{aligned}$$

where $\mathbf{x}_1^0 = \pi_1$ and $V_1^0 = V_1$. To compute $\widehat{\mathbf{x}}_t \equiv \mathbf{x}_t^T$ and $P_t \equiv V_t^T + \mathbf{x}_t^T\mathbf{x}_t^{T\prime}$ one performs a set of backward recursions using,

$$\begin{aligned}
\mathbf{J}_{t-1} &= V_{t-1}^{t-1}A'(V_t^{t-1})^{-1} && (15)\\
\mathbf{x}_{t-1}^T &= \mathbf{J}_{t-1}(\mathbf{x}_t^T - A\mathbf{x}_{t-1}^{t-1}) && (16)\\
V_{t-1}^T &= V_{t-1}^{t-1} + \mathbf{J}_{t-1}(V_t^T - V_t^{t-1})\mathbf{J}_{t-1}' && (17)
\end{aligned}$$

For calculation of $P_{t,t-1}$,

$$\begin{aligned}
V_{t,t-1} &= \mathbf{J}_{t-1}V_t^T && (18)\\
P_{t,t-1} &= V_{t,t-1}^T + \mathbf{x}_t^T\mathbf{x}_{t-1}^{T'} && (19)
\end{aligned}$$

The conditional likelihood is computed as,

$$P(y_t|\mathscr{Y}_1^{t-1}) = \mathscr{N}(\mathbf{h}_t'\mathbf{x}_t^{t-1}, \mathbf{h}_t'V_t^{t-1}\mathbf{h}_t + \sigma_v^2) \quad (20)$$

With this equation, the progress of the learning algorithm is monitored. This formulation is similar to [4], but there is an important difference. As $\mathbf{h}_t$ is varying with time, the estimate of $\widehat{\mathbf{x}}_t, P_t, P_{t,t-1}$ will also vary with time, unlike the usual formulation of Kalman filters (see [8]), where P_t and $P_{t,t-1}$ are independent of data.

3.2 M-step

Maximizing $\mathscr{Q}(\Theta,\Theta^{i-1})$ with respect to Θ, we get the following estimates in the ith iteration:

$$\begin{aligned}
\widehat{A}^i &= \Big(\sum_{t=2}^T P_{t,t-1}\Big)\Big(\sum_{t=2}^T P_{t-1}\Big)^{-1} && (21)\\
\widehat{\sigma_v^2}^i &= \frac{1}{T}\sum_{t=1}^T\Big(y_t^2 - 2\,\mathbf{h}_t'\widehat{\mathbf{x}}_t y_t + \mathbf{h}_t'P_t\mathbf{h}_t\Big) && (22)\\
\widehat{Q}^i &= \frac{1}{T-1}\Big(\sum_{t=2}^T P_t - \widehat{A}^i\sum_{t=2}^T P_{t-1,t}\Big) && (23)\\
\widehat{\pi}_1^i &= \widehat{\mathbf{x}}_1 && (24)\\
\widehat{V}_1^i &= P_1 - \widehat{\mathbf{x}}_1\widehat{\mathbf{x}}_1' && (25)
\end{aligned}$$

3.3 IF Estimation

Once we have the estimates of Θ, TVAR parameters i.e. $\mathbf{x}_t$ can be inferred from Kalman smoother eq.(10)-(17). Then, we estimate the time-varying frequency spectrum as,

$$\widehat{P}_t(f) = \frac{\widehat{\sigma_v^2}}{|1 - \sum_{k=1}^p \widehat{a}_k^t e^{-i2\pi kf/fs}|^2} \quad (26)$$

The IF estimate is computed as the peak of the spectrum $\widehat{P}_t(f)$, i.e.,

$$\widehat{f}_t = \arg\max_f \widehat{P}_t(f) \quad (27)$$

4. PRACTICAL ISSUES

We observed that initializing the algorithm properly leads to better estimates, and speedens convergence. For initialization, we divide the dataset into overlapping windows, and compute the maximum likelihood estimates of $\mathbf{x}_t$ for each window. By setting A to identity matrix, and assuming state equation (3), we compute maximum likelihood estimates of $\{Q, \mathbf{x}_1, V_1\}$. With these estimates of parameters, we again run the dataset through Kalman smoother to get the estimates of $\mathbf{x}_t$, from which the maximum likelihood estimates of all the parameters are obtained. These are used to initialize the EM algorithm.

After initialization, E and M steps are iterated. The progress of the algorithm is monitored with the likelihood given by eq.(20), and the algorithm is said to have converged, if successive iterations do not improve the likelihood score by more than 0.01%.

To generalize the formulation for multiple observations, we need to take care of the fact that $\widehat{\mathbf{x}}_t, P_t, P_{t,t-1}$ are time dependent. For the sake of simplicity, we assume all the observations to be *i.i.d.*, which yields,

$$\mathscr{Q}(\Theta,\Theta^{i-1}) = \sum_{k=1}^N E\left[\left(\ln p(\mathscr{X}^k,\mathscr{Y}^k|\Theta)\right)|\mathscr{Y}^k,\Theta^{i-1}\right] \quad (28)$$

where $\mathscr{X}^k$ and $\mathscr{Y}^k$ are variables associated with the kth observation. Clearly, the estimates of $\widehat{\mathbf{x}}_t, P_t, P_{t,t-1}$ are the sum of the estimates of each observation. Then, $\mathscr{Q}(\Theta,\Theta^{i-1})$ is maximized with respect to Θ.

In this paper, we set model order at $p = 4$. Effect of model order on the performance and computational complexity of the algorithm will be studied separately.

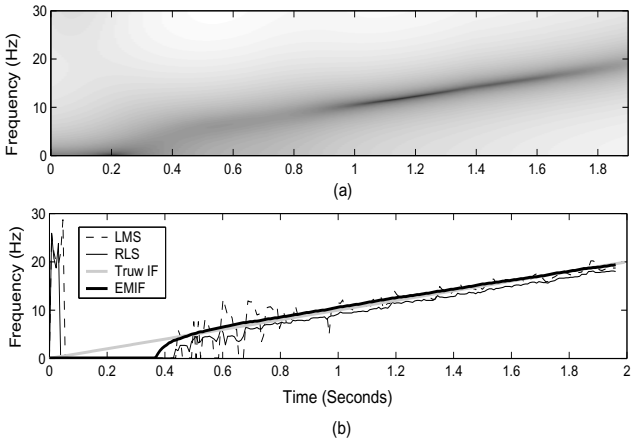

Figure 1: (a) Time-varying spectral estimate of a linear FM signal, estimated with EMIF (b) comparison of EMIF, RLS and LMS, for IF estimation of linear FM (one realization at 10.98 dB).

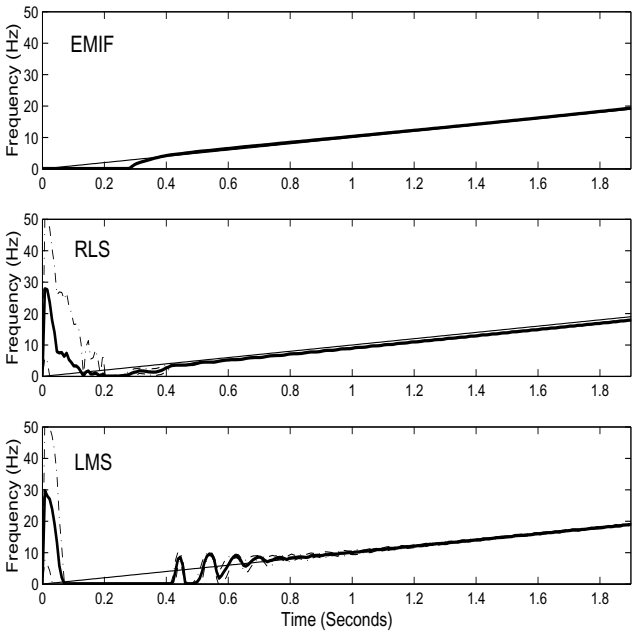

Figure 2: $\widehat{f_t}$(solid line), $\widehat{f_t} \pm \sigma_f$ (dashed line) for IF estimates of 100 realizations of linear FM signal, with SNR=17.9 dB, estimated with EMIF, RLS and LMS

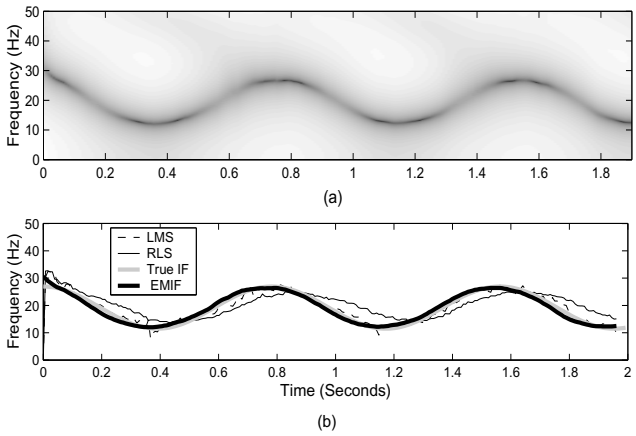

Figure 3: (a) Time-varying spectral estimate of a sinusoidal FM signal, estimated with EMIF (b) Comparison of EMIF, RLS and LMS, for IF estimation of sinusoidal FM (one realization at 10.98 dB).

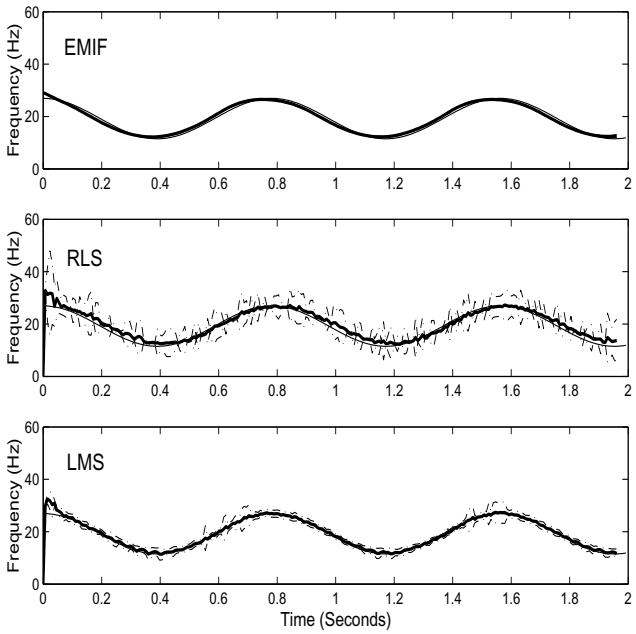

Figure 4: $\widehat{f_t}$(solid line), $\widehat{f_t} \pm \sigma_f$ (dashed line) for IF estimates of 100 realizations of sinusoidal FM signal, with SNR=17.9 dB, estimated with EMIF, RLS and LMS

5. RESULTS

5.1 Experiment I

First, we compare the performance, for linear IF estimate. Consider,

$$y_t = 5\sin(2\pi f_t t) + u_t \tag{29}$$

where f_t is the instantaneous frequency given by $f_t = 10t$, and u_t is a white Gaussian noise with variance σ_u^2. We generated one realization with $\sigma_u^2 = 1$ (SNR=10.98 dB), of 2 seconds duration, and sampled at 128 Hz. Time-varying spectral estimate is shown in Fig.1(a). IF estimates were obtained with EMIF, RLS and LMS. Forgetting factor λ, for RLS and step size μ for LMS, were set to 0.95 and 0.04, respectively. One can clearly see in Fig.1(b), that the steady state error, as well as settling time, are very less in EMIF, compared to RLS and LMS. Hence, we see that EMIF gives the best overall performance.

To analyze the average performance over realizations, we generated 100 realizations with $\sigma_u^2 = 0.2$. λ and μ were set to 0.95 and 0.01, respectively. Fig.2 shows $\widehat{f_t}$ and $\widehat{f_t} \pm \sigma_f$, where $\widehat{f_t}$ and σ_f denote mean and standard deviation, respectively, of the IF estimates. Here again, EMIF is observed to have very less steady state error and settling time, compared to RLS and LMS. Also, EMIF has the closest mean to the true IF, and least variance of all. Thus, EMIF has the best average performance too.

5.2 Experiment II

Now we compare for non-linear FM signal (sinusoidal) given as,

$$y_t = 5\sin(2\pi f(t + 0.05\sin(2\pi f_{mod} t))) + u_t. \quad (30)$$

where $u_t \sim \mathcal{N}(0, \sigma_u^2)$. The instantaneous frequency, underlying the signal, is given by,

$$f^i(t) = f + 0.1\pi f f_{mod} \cos(2\pi f_{mod} t) \quad (31)$$

We set $f = 19.2$ Hz and $f_{mod} = 1.28$ Hz. We produced one realization of 2 seconds duration with $\sigma_u^2 = 1$ (SNR=10.98 dB), and sampled at 128 Hz. Time-varying spectral estimate with EMIF is shown in Fig.3(a). λ and μ were set to 0.85 and 0.01 respectively. IF estimates are shown in Fig.3(b). We see that EMIF tracks the true IF better than RLS and LMS.

To compare the average performance, we compute IF estimates of 100 realizations, with $\sigma_e^2 = 1.25$ (SNR=10dB). λ and μ were set to 0.85 and 0.01 respectively. Fig.4 shows $\widehat{f_t}$ and $\widehat{f_t} \pm \sigma_f$. Clearly, EMIF has the closest mean and least spread of all.

5.3 Experiment III

To complete the analysis, we compare the error performance at different noise levels, for both linear and sinusoidal FM signal. We vary σ_u^2, and estimate the IF for 100 realizations. The mean square error (MSE) was calculated as,

$$MSE = \sum_{t=\frac{T}{2}}^{T} \sum_{k=1}^{N} (\widehat{f_t^k} - f_t)^2 \quad (32)$$

where, $\widehat{f_t^k}$ is the IF estimate from k^{th} realization, f_t is true IF, N is the total number of realizations, T is observation length. Note that rather than computing MSE at a point, we averaged MSE for all the points, from middle of the block to end. This is because of our interest in tracking the IF and not in estimating it at a point. Also, first half of the estimate is not used for MSE computation, because the RLS and LMS algorithms have large initial error. Including them will bias the error measure and hence, they were not considered. Only $[T/2, T]$ was used.

The average error against SNR for linear FM and sinusoidal FM, is shown in Fig.5 (a) and (b). One can clearly see that error is least for EMIF. At higher SNR too, EMIF performs satisfactorily, whereas the performance of LMS and RLS is very bad.

6. CONCLUSION

In this paper, we proposed a new algorithm EMIF for instantaneous frequency estimation, using EM algorithm approach in a Kalman smoother framework. We compared EMIF with RLS and LMS algorithms for IF estimation. EMIF was shown to have best IF tracking. We compared MSE at different noise levels, and EMIF was shown to have least error for all. Even at higher SNR, where the performance of LMS and RLS was very bad, EMIF performed reasonably well. Hence, EMIF is a significant improvement over other available adaptive algorithms.

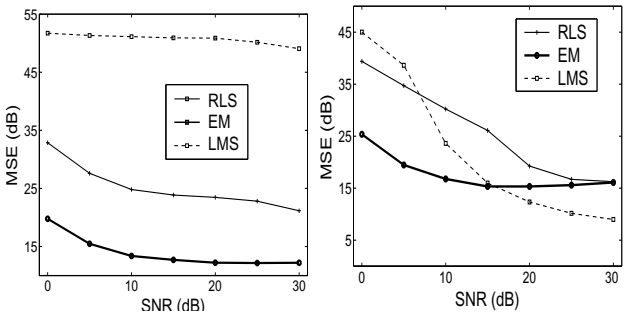

Figure 5: MSE vs. SNR for (a) linear FM signal (b) sinusoidal FM signal

7. ACKNOWLEDGMENT

The authors would like to thank S. Chandra Shekhar, Dept. of Electrical Communication Engg., Indian Institute of Science, Bangalore for his advice and support during this work.

REFERENCES

[1] P. Shan and A. A. Beex, "High-resolution instantaneous frequency estimation based on time-varying AR modeling", *IEEE 4th International Symposium on Time-Frequency and Time-Scale Analysis (TFTS'98)*, Pittsburgh, PA, Oct. 1998.

[2] M. Arnold, W. H. R. Miltner, H. Witte, R. Bauer, and C. Braun, "Adaptive AR modelling of nonstationary time series by means of Kalman filtering", *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 553-562, May 1998.

[3] R.H. Shumway and D.S. Stoffer, "An approach to Time Series smoothing and forecasting using the EM algorithm", *J. Time Series Anal.*, vol. 3, No. 4, pp. 253-264, 1982.

[4] Z. Ghahramani and G. E. Hinton (1996), "Parameter Estimation for Linear Dynamical Systems", Technical report, Univ. Toronto, Dept. Comput. Sci, Toronto, ON, Canada.

[5] M. J. Cassidy and W. D. Penny, "Bayesian nonstationary autoregressive models for biomedical signal analysis", *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1142-1152, Oct.2002.

[6] K. C. Sharman and B. Friedlander, "Time-varying autoregressive modeling of a class of nonstationary signals", *Proc. ICASSP'84*, vol. 2, pp. 22.2.1-22.2.4, San Diaego, CA, March 1984.

[7] B. Boashash, "Estimation and interpreting the instantaneous frequency of a signal - part 2: algorithm and applications", *Proceedings of IEEE*, vol. 80, no. 4, pp. 540-568, 1992.

[8] Kay, S.M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall International, 1993.