EFFICIENT IMPLEMENTATION OF THE HMARM MODEL IDENTIFICATION AND ITS APPLICATION IN SPECTRAL ANALYSIS

Chunjian Li and Søren Vang Andersen

Department of Electronic System, Aalborg University
DK-9220 Aalborg Øst, Denmark
e-mail: cl@es.aau.dk, sva@es.aau.dk

ABSTRACT

The Hidden Markov Auto-Regressive model (HMARM) has recently been proposed to model non-Gaussian Auto-Regressive signals with hidden Markov-type driving noise. This model has been shown to be suitable to many signals, including voiced speech and digitally modulated signals received through ISI channels. The HMARM facilitates a blind system identification algorithm that has a good computational efficiency and data efficiency. In this paper, we solve an implementation issue of the HMARM identification, which can otherwise degrade the efficiency of the model and hinder extensive evaluations of the algorithm. Then we study in more detail the properties associated with the autoregressive (AR) spectral analysis for signals of interest.

1. INTRODUCTION

Exploiting the non-Gaussianity of signals in spectral analysis can often offer significant improvements in estimation accuracy over traditional Gaussianity based methods. In [1] and [2], Li and Andersen show that specially designed non-Gaussian models for specific types of signals can exploit the structures in the signals and achieve higher computational and data efficiency than general purpose non-Gaussian methods such as the higher order statistics methods and Gaussian Mixture Model based methods. The Hidden Markov Auto-Regressive model (HMARM) proposed in [1] is tailored for signals generated by exciting an autoregressive (AR) filter with either a finite-alphabet symbol sequence or a hidden Markov sequence. Due to the non-Gaussian nature of the excitation, this type of signal belongs to the class of non-Gaussian AR signals. Li et. al. proposed an efficient learning algorithm for the HMARM to jointly estimate the AR coefficients and the excitation symbols or the parameters of the hidden Markov sequence. The joint estimation is what distinguishes the method from other identification algorithms of models that have similar source-filter structure: most known methods estimate the source parameters and the filter parameters in a sequential way, resulting in lower efficiencies. The HMARM algorithm is an exact EM algorithm, which solves for a set of linear equations iteratively and converges in a few iterations. It is shown that compared to the classical autocorrelation method of AR spectral analysis, the HMARM has a smaller bias, a smaller variance, and a better shift invariance property. In [2], the HMARM is extended for robust analysis of noisy signals by introducing an observation noise model to the system. At moderate noise levels, the algorithm achieves a high estimation accuracy without *a priori* knowledge of the noise variance. Applications of the model to different signals, including noise robust spectral analysis of speech signals and blind channel estimation, are demonstrated in [1] [2], and promising results are obtained.

One critical issue in the frame based implementation of the HMARM algorithm in [1] is that, if a signal is segmented into frames, the HMARM could have problems estimating the parameters for those frames that do not contain the onset of the signal. This is because when estimating the AR parameters of the current frame, the estimator has no knowledge about the excitation in the previous frame, but the large impulses in the previous excitation can cause large "ripples" in the beginning of the current frame, which then causes the state estimator in the HMARM to make wrong decisions. Since the parameter estimations are based on the state decisions, these estimates become erroneous too. In the previous papers, this problem is solved by pre-processing the frame to remove the "ripples" caused by the previous frame. For simplicity of that approach, all samples before the first impulse in the current frame are set to zero. This solution is somewhat troublesome since it requires an impulse detector in the residual domain, whose accuracy affects the performance of the whole system. This and other ways of subtracting the ripples also lower the computation efficiency and data efficiency, since they add extra complexity and discard data samples. In this paper, we address this problem by exploiting the Markovian property of the AR model in a way analogous to the covariance method for AR spectral analysis. Our proposed solution costs no extra complexity, and is highly reliable.

The rest of the paper is organized as follows. Section 2 describes the covariance implementation, and discusses its benefits. Then, in Section 3, we investigate some interesting properties of the HMARM using the proposed implementation in application to spectral analysis.

2. COVARIANCE METHOD FOR THE HMARM

The causality problem associated with the frame based implementation[1] of the HMARM is functionally different from the boundary problem in the least-squares (LS) method. The classical LS solution to the AR spectral analysis assumes the excitation to the AR filter to be a stationary white Gaussian sequence. With this assumption, the only parameter of the excitation statistics, the variance, is decoupled from the estimation of the AR filter coefficients. Therefore, the excitation has no effect on the AR filter estimates. However, the HMARM has a more sophisticated model for the excitation, and the estimations of the excitation parameters and the

This work was supported by The Danish National Centre for IT Research, Grant No. 329, and Microsound A/S.

[1] In this context, the frames have no overlap.

AR parameters affect each other. Specifically, the HMARM models the excitation as a hidden Markov sequence. During the estimation, the states of the excitation sequence at each time instant are first estimated by calculating the state probabilities. Based on the state decisions, the AR filter coefficients and the parameters of the hidden Markov model are estimated by a set of coupled linear equations, c.f. [1] and [2] for derivations. For convenience, we list below the signal model and the final equations of the estimator.

For a signal generated by the following model,

$$x(t) = \sum_{k=1}^{p} g(k)x(t-k) + r(t) \tag{1}$$

$$r(t) = v(t) + u(t), \tag{2}$$

where $x(t)$ is the signal, $g(k)$ is the kth AR coefficient, and $r(t)$ is the excitation sequence consisting of a Markovian sequence $v(t)$ and additive white Gaussian noise $u(t)$, the estimates of the parameters are obtained from solving the following $p+m$ equations, where p is the order of the AR model, and m is the number of states of the HMM. For $k = 1, \cdots, p$, and $j = 1, \cdots, m$:

$$\sum_{j}^{m} \sum_{t=1}^{T-1} \gamma(j,t)\Big(x(t) - m_x(j,t)\Big)x(t-k) = 0, \tag{3}$$

$$\sum_{t}^{T-1} \gamma(j,t)\Big(x(t) - m_x(j,t)\Big) = 0, \tag{4}$$

where $\gamma(j,t)$ is the posterior probability of the states, and

$$m_x(j,t) = \sum_{k=1}^{p} g(k)x(t-k) + m_r(j), \tag{5}$$

where $m_r(j)$ is the mean of state j.

The state posterior $\gamma(j,t)$ is estimated by a forward-backward induction, based on an initial estimate of the AR coefficients. The LS estimates of the AR coefficients are used as the initialization. With the voiced speech signal as an example, the voiced speech can be modeled as a noisy impulse train filtered by a vocal tract filter, and a two-state HMM is sufficient for representing the impulse train: a state with a mean equal to the magnitude of the impulses, and a state with a zero mean. For a frame that does not contain the onset of the impulse train, there must be ripples, or ringing, at the beginning of the frame, which is originated from an impulse in the previous frame. If the ringing is large enough, it will be erroneously interpreted by the algorithm as having a non-zero-mean state at the beginning of the frame although the true state is a zero-mean state. The wrong decision on the state certainly has a negative impact on the subsequent estimation of parameters. To illustrate the problem, in Fig. 1, we plot the log-spectral distance (LSD) between an estimated spectrum and the true spectrum for frames of signal beginning at different time instants. The signal is a synthetic speech signal, generated by filtering a noisy impulse train with a 10th order AR filter (the first 200 samples of the signal and its excitation are shown in Fig. 2). The first impulse, i.e. the onset, is located at the 50th sample. A hundred frames with length of 320 samples are taken from the signal by shifting the frame one sample each time. The figure shows that for the first 50 frames, i.e. all the frames that contain the onset, the spectral distortions of the HMARM spectra are low

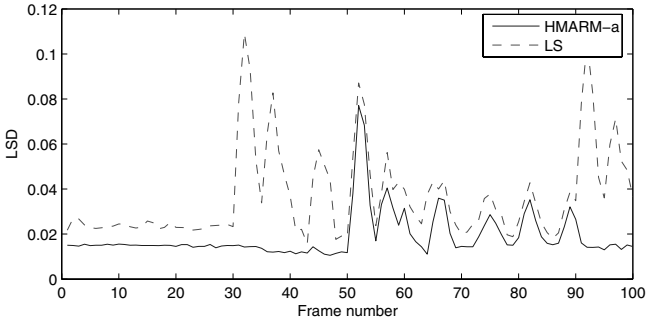

Figure 1: The log-spectral distances between the true AR spectrum and the estimates.

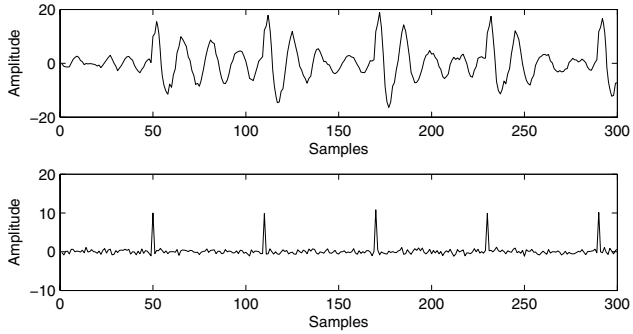

Figure 2: The synthetic signal waveform (upper panel) and its excitation (lower panel).

and constant, indicating accurate estimates and a good shift-invariance property of the HMARM analysis. In the same period, the LSD for the LS estimates of the signal spectra fluctuates a lot, indicating a poor shift-invariance property. In the rest 50 frames, where the onset impulse is absent, the HMARM method loses its nice property and the distortion of the HMARM method is much higher and fluctuating almost as much as the LS method. Note that here, the problem with the LS method and the one with the HMARM method are different: the LS estimates have a large variance because it fails to represent the non-Gaussianity of the impulse train structure in the excitation; the HMARM has a good model for the impulse train structure, so it succeded to bring down the distortion and estimation variance in the first 50 estimates, but it failed to do so in the last 50 estimates due to the causality problem discussed above.

The results of the HMARM shown in Fig. 1 are without any preprocessing. To avoid the problem, in [1] and [2], a preprocessor detects the position of the first impulse of the excitation in the current frame, and sets all samples before this position to zero, such that large ripples trailing from the previous frame are removed. The problem with this solution is that removing samples reduces data efficiency of the algorithm. The reliability of the impulse detector is also a concern. Another solution is to calculate the ripples from the previous frame, using the estimated AR filter and the impulses of the previous frame, and subtracting it from the current frame. This solution also reduces data efficiency, since a certain part of the signal energy is discarded, which could have been used by the estimator. Furthermore, the ringing will be subtracted using an inaccurate estimate of the AR co-

efficients. Moreover, these solutions add extra complexity to the algorithm.

The solution we propose in this paper is based on the observations that the HMARM has a built in linear predictor, i.e. (5), and that an AR(p) process is a Markovian process with vector states of p-dimension. So, instead of calculating the long trailing ripples from the previous frame using estimated parameters and subtract it from the following frames, it is better to initialize the predictor of the current frame with the p samples in the end of the previous frame, which gives the state estimator all the information about the past. thereby the causality problem is avoided.

To implement this solution, we only have to change the way the data matrix and the p covariance vectors are populated. They are used in the matrix form of the predictor (5) and the equations system (3) in the following forms:

$$\begin{bmatrix} x_0 & x_{-1} & x_{-2} & \cdots & x_{-p+1} \\ x_1 & x_0 & x_{-1} & \cdots & x_{-p+2} \\ x_2 & x_1 & x_0 & \cdots & x_{-p+3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{T-1} & x_{T-2} & x_{T-3} & \cdots & x_{T-p} \end{bmatrix}, \quad (6)$$

where T is the frame length, and

$$[x_1 x_{1-k}, \quad x_2 x_{2-k}, \quad \cdots, \quad x_T x_{T-k}]^t, k = 1, \cdots, p. \quad (7)$$

In the frame based implementation the samples with negative indices are of value zero. To provide the estimator a correct starting state, the samples in the previous frame must be put into the appropriate positions of the matrices. In the case that the previous frame is missing, the first p rows of the matrices in (6) and (7) must be removed, so that there is no un-populated elements (the zeros) in the matrices. This is formally similar to the covariance method of the LS analysis of AR models [3]. Therefore, we term it the covariance method HMARM (HMARM-c), and the original implementation the autocorrelation method HMARM (HMARM-a). The LSD of the two implementations are plotted in Fig. 3 for comparison. It is clear from this figure that the covariance method HMARM maintains its good performance for all frames. Notice that for frames that contain the onset impulse, the performance of the covariance method HMARM is similar to the autocorrelation method HMARM. This is in contrast to the LS, whose covariance method implementation always outperforms its autocorrelation method implementation, given that the signal length is small.

3. HMARM FOR SPECTRAL ANALYSIS

Now, we discuss some properties of the HMARM that can be beneficial in the AR spectral analysis. The HMARM hereafter refers to the covariance method implementation.

3.1 Window design and covariance methods

As shown in [1] and [2], the HMARM estimate of the AR spectrum has significantly lower bias and variance than the LPC analysis, which is an autocorrelation LS method. The variance studied therein is the shift variance, where the set of realizations of an AR process is generated by shifting a time window many times with one sample as the shift step length. Other known methods for reducing the shift variance of the LS analysis are the window design and the covariance

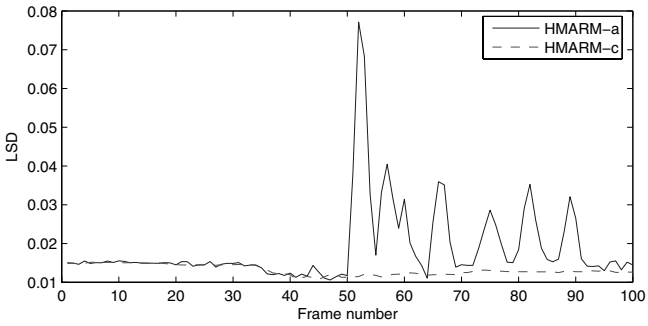

Figure 3: The log-spectral distances between the true AR spectrum and the estimates. HMARM-a: the autocorrelation method of the HMARM; HMARM-c: the covariance method of the HMARM.

method LS. In [1], it has been shown that applying a Hamming window reduces the shift variance of the LPC analysis, but the reduced variance is still significantly larger than that of the HMARM. Besides, any window other than the rectangular window has the side effect of reduced spectral resolutions. Here, we discuss the covariance method LS analysis, and compare the three methods under a more general variance analysis.

The covariance method LS reduces the shift variance by avoiding the boundary effect. This is done by feeding a number of samples preceding the current frame to the data matrix. In this way, the covariance matrix of the signal becomes non-Toeplitz. Nevertheless, the optimality of the method is still based on the assumption that the excitation is white stationary Gaussian. Therefore, for the signals of interest in this work, the large variance caused by the mismatch between the assumption and the signal is still there. To reveal a more general statistics than only the shift variance, we let the sliding window shift so many times that the beginning frames and the ending frames contain entirely different samples. In this way, it is possible to show a variance consisting of both the shift variance and the variance due to different realizations. We investigate the statistical properties of the three estimators, with a synthetic speech signal and a bipolar signal received through an AR channel. The synthetic speech signal is the one used in the previous example (Fig. 2), and the received bipolar signal is generated by filtering a random [-1,1] sequence with an AR filter. They are the two typical non-Gaussian AR signals with different characteristics: the excitation of the speech signal is spectrally colored due to the periodic impulses, and has a Gaussian component due to the noise; while the transmitted bipolar sequence is spectrally white, and very non-Gaussian since there is no Gaussian noise in it. Tab. 1 shows the biases and variances of the three methods. The statistics are obtained from estimating 600 frames of an AR process, and the frames are obtained by moving a 320-sample window 600 times by one sample each time.

The results show that: 1) the HMARM has a consistently smaller variance than the autocorrelation method LS, especially for a signal that has no Gaussian componets, and 2) generally, the Hamming windowing and the covariance method do not reduce the variance of an LS AR analysis.

	Speech		Bipolar	
	bias	variance	bias	variance
HMARM-c	0.0861	27.68	8.8×10^{-15}	4.7×10^{-24}
LS-c	0.1524	169.39	0.1595	190.41
LS-a-w	0.1276	185.90	0.1862	560.95
LS-a	0.1879	179.22	0.3100	160.46

Table 1: Comparison of biases and variances. HMARM-c: the covariance method HMARM, LS-c: the covariance method LS, LS-a-w: the autocorrelation method LS with Hamming window, LS-a: the autocorrelation method LS.

3.2 Avoiding spectral sampling effect

Having a more sophisticated model for the excitation makes the estimation accuracy of the HMARM superior to the traditional Gaussian AR model when applied to spectral analysis of certain non-Gaussian signals. This is because the excitation to an AR filter is often not spectrally white and/or non-Gaussian. With the HMARM, correlation in the excitation can be separated from that caused by the AR filter. Thus the estimates of the AR spectral envelope are not affected by the excitation. An example of related problems for the Gaussian AR model is the spectral sampling effect due to the impulse train structure in voiced speech.

A voiced speech signal is commonly modeled by AR filtering of an impulse train. The impulse train has a comb-shape spectrum. Although the LPC analysis is intended for estimating the spectral envelope of the signal, which models the vocal tract resonance property, the comb-shape excitation spectrum has a spectral sampling effect on the estimated spectral envelope. This causes the following problems. Firstly, when a formant peak happens to locate at one of the harmonic frequencies of the impulse train, the estimated spectral envelope will have an abnormally sharp peak. This is a well known problem for the LPC analysis in speech coding, especially for high pitch speech [4][5]. Secondly, in the case that the formant peaks do not locate at a harmonic frequency, the peaks of the estimated spectral envelope tend to drift to the neighboring harmonic frequencies. This effect is undesired in applications such as speech synthesis and prosody manipulation. We compare the spectral envelopes estimated by the LPC and the HMARM, using two synthetic speech signals with pitch frequencies of 133Hz and 200Hz. Fig. 4 shows that the LPC spectral envelope has an abnormally sharp peak, while the HMARM estimate does not have the problem. Fig. 5 shows that the spectral peaks of the LPC estimate drift towards the harmonic frequencies, while the HMARM estimate has the peaks in correct positions.

3.3 Avoiding over training

Another problem associated with parametric modeling is known as over training, or over fitting. In the specific case of AR spectral analysis, over training is referred to the phenomena that when modeling the signal with a model order larger than the true order, the AR spectrum tends to fit to the FFT spectrum instead of the spectral envelope. Here we take the bipolar signal as an example. The transmitted signal is a randomly generated bipolar signal with a white spectrum. The signal is convolved with an AR channel before it is received. The receiver seeks to de-convolve the channel distortion by first estimating the channel. In general, the model order is unknown, and using a too large order may result in over training. In this example, the channel is AR(10), but we try to model it using an AR(40) model since we have no access to the true order of the channel. In Fig. 6 we show that the HMARM largely avoids the effect of over fitting, while the LPC spectral envelope starts representing the random peaks due to the spectrum of the transmitted signal.

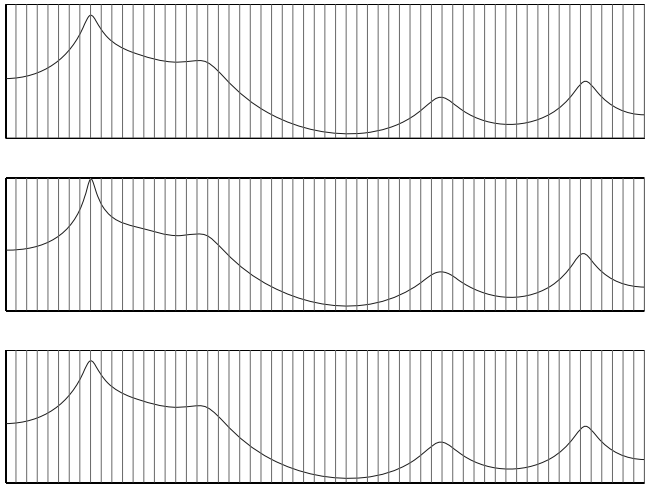

Figure 4: The AR spectra estimated by the HMARM (upper) and the LPC (middle), and the true spectrum (lower). The vertical bars show the harmonic frequencies. The pitch frequency is 133Hz.

4. CONCLUSION

In this paper, we propose a covariance-method like implementation of the HMARM system identification algorithm. The method solves the causality problem that can cause the state estimator to fail in a frame based HMARM analysis. The proposed method costs no additional complexity to the system, and shows in experiments to be highly reliable. Based on the results of the new implementation, a few interesting issues concerning the AR spectral analysis are addressed. Examples are given for speech and digitally modulated signals with promising results.

REFERENCES

[1] C. Li and S. V. Andersen, "Blind identification of non-Gaussian Autoregressive models for efficient analysis of speech signals," *Proc. of ICASSP*, Apr. 2006.

[2] C. Li and S. V. Andersen, "Efficient blind identification of non-Gaussian Autoregressive models with HMM modeling of the excitation," *IEEE Trans. on Signal Processing*, 2006, Accepted for publication.

[3] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, Prentice Hall, 2005.

[4] L. Anders Ekman, W. Bastiaan Kleijn, and M. N. Murthi, "Spectral envelope estimation and regularization," *Proc. ICASSP*, pp. I245–I248, 2006.

[5] M. N. Murthi, "Regularized linear prediction all-pole models," *Proc. IEEE Workshop on Speech Coding*, pp. 96–98, 2000.

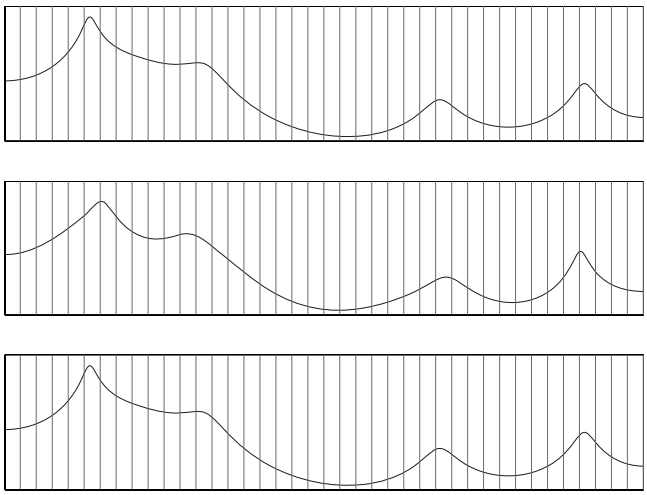

Figure 5: The AR spectra estimated by the HMARM (upper) and the LPC (middle), and the true spectrum (lower). The vertical bars show the harmonic frequencies. The pitch frequency is 200Hz.

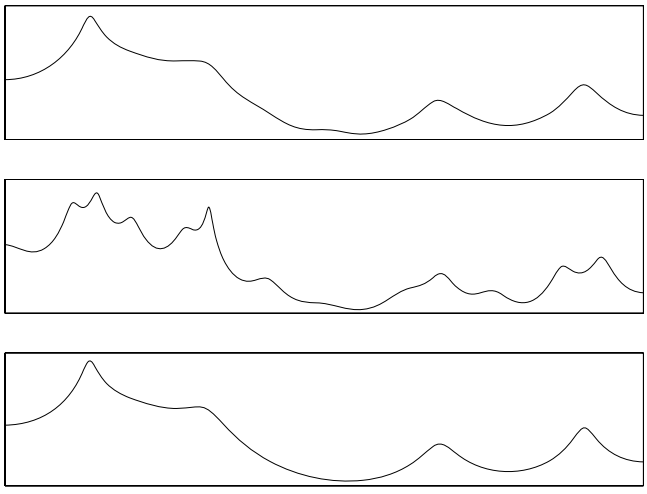

Figure 6: The AR spectra estimated by the HMARM (upper) and the Least Squares method (middle) with order 40, and the true spectrum of order 10 (lower).