TWO-KALMAN FILTERS APPROACH FOR UNBIASED AR PARAMETER ESTIMATION FROM NOISY OBSERVATIONS, APPLICATION TO SPEECH ENHANCEMENT

David Labarre(1), Eric Grivel(1), Mohamed Najim(1) and Ezio Todini(2)

[1]Equipe Signal & Image, UMR LAP 5131, ENSEIRB BP 99, F-33402 Talence, France (Europe)
phone: +33 5 40 00 61 85, fax: +33 5 40 00 84 06, email: {labarre ; grivel ; najim}@tsi.u-bordeaux1.fr
[2]Dept. of Earth and Geo-Environmental Sciences, University of Bologna, Via Zamboni, 67, 40127 Italy (Europe)
phone: +39 51 354 537, fax: +39 51 354 522, email: todini@tin.it

ABSTRACT

The AR model is used in many applications such as speech processing. However, when the observations are contaminated by additive noise, standard methods produce biased AR parameter estimates. To avoid a non linear approach to estimate the signal and the parameters at the same time, such as the EKF, a sequential method using two conditionally linked Kalman filters running in parallel is here presented. Following the instrumental variables concept, at each step in time, one of the filters uses the latest estimated AR parameters to estimate the signal, while the second one uses the estimated signal to estimate the AR parameters, to be used in the next step. This approach, already applied in hydrological applications, has the advantage of providing an unbiased estimation of the parameters and an estimation of the signal in the steady state. This method is then derived in the framework of speech enhancement.

1. INTRODUCTION

The autoregressive (AR) modelling technique is used in a wide range of applications such as speech processing and biomedical engineering. It consists in modelling a signal by a p^{th} order AR process $s(k)$ defined as follows:

$$s(k) = -\sum_{i=1}^{p} a_i s(k-i) + u(k) \tag{1}$$

where $\underline{\theta} = [a_1 \quad \cdots \quad a_p]^T$ is the vector of the AR parameters and $u(k)$ is the so-called driving process, assumed to be zero-mean Gaussian white noise with variance $\sigma_u^2(k)$. Many approaches have been developed to obtain an estimation $\hat{\underline{\theta}}$ of $\underline{\theta}$, based on $s(k)$. A widely used solution consists in solving the Yule-Walker (YW) equations. On-line methods have been also developed which are based on adaptive filters such as LMS, RLS or Kalman filter.

In real cases, the signal $s(k)$ is often corrupted by an additive noise $v(k)$, usually assumed to be zero-mean Gaussian with variance σ_v^2:

$$y(k) = s(k) + v(k) \tag{2}$$

In this case, the above approaches provide biased parameter estimates and produce a "flatter" spectrum because the corresponding poles are closer to the centre of the unit circle.

The Modified Yule Walker (MYW) equations [10] make it possible to obtain a consistent estimation of AR parameters. However, only a reduced number of noisy observations is available in real cases. Zheng et al. [15] have developed an iterative bias correction based algorithm. In [1], Davila et al. propose to view the AR parameter estimation issue as a quadratic eigenvalue problem. On-line methods using for instance the γ-LMS [12] or a derived version of the so-called ρ-LMS [14] can also be considered.

When developing a Kalman filter-based method, signal estimates are necessary to retrieve the AR parameters. This is a special case of the joint parameter and signal estimation problem, also referred to as the dual estimation problem [6]. A non linear solution consists in using an Extended Kalman Filter (EKF). However, the convergence properties of the EKF are not guaranteed due to the non linearity of the problem. For this reason, Nelson et al. [8] have used in the framework of control two separate Kalman filters. Once the parameter estimation convergence has been reached, the signal is retrieved by deriving a Kalman filter with the innovation model and the steady-state Kalman gain.

In this paper, we present an alternative solution involving two interacting standard Kalman filters and following the instrumental variables concept [13] [7]. At each step in time, one of the Kalman filters estimates the signal conditionally to the latest estimated value of the parameters. Conversely, the second filter estimates the parameters conditionally to the latest a posteriori signal estimate. It can be shown that the optimality conditions keep the two filters mutually orthogonal, which avoids the necessity of specifying the covariance between parameters and signal as in the EKF. Thus, this algorithm has the advantage of providing unbiased AR parameter estimates from noisy observations.

The remainder of the paper is organized as follows: In part 2, we present the proposed algorithm. In part 3, a comparative study is completed on synthesised data with existing methods. In part 4, a derived version of the algorithm is tested in the framework of speech enhancement and is compared with various Kalman filter-based speech enhancement algorithms.

It should be noted that this paper extends a previous result proposed in the framework of water resources by one of the authors [11].

2. SEQUENTIAL ESTIMATION OF THE AR PARAMETERS FROM NOISY OBSERVATIONS

2.1 Estimation of the signal from noisy observations

Here, the purpose is to generate signal estimates $\hat{s}(k)$ from noisy observations. Since the Kalman filter produces a recursive estimate of the state vector $\underline{x}(k)$, we define it by:

$$\underline{x}(k) = \left[s(k) \quad \cdots \quad s(k-p+1)\right]^T \tag{3}$$

The resulting state space representation of the system (1)-(2) is given by:

$$\begin{cases} \underline{x}(k) = \Phi(k)\underline{x}(k-1) + \Gamma u(k) \\ y(k) = H\underline{x}(k) + v(k) \end{cases} \tag{4}$$

where the transition matrix $\Phi(k)$, the input vector Γ and the observation vector H respectively satisfy:

$$\Phi(k) = \begin{bmatrix} -\hat{a}_1(k-1) & \cdots & \cdots & -\hat{a}_p(k-1) \\ 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and
$$H = \Gamma^T = \left[1 \quad 0 \quad \cdots \quad 0\right] \tag{5}$$

In the following, $\hat{\underline{x}}(k/l)$ and $P(k/l)$ will respectively denote the estimation of the state vector $\underline{x}(k)$ given *l* observations and its corresponding error covariance matrix.

Once a new observation $y(k)$ is available, the Kalman filter (denoted by KF1 in Fig. 1) provides a recursive estimation of $\underline{x}(k)$ and an estimation of the speech signal, as follows:

$$\hat{s}(k/k) = H\hat{\underline{x}}(k/k) \tag{6}$$

where
$$\hat{\underline{x}}(k/k) = \Phi(k)\hat{\underline{x}}(k-1/k-1) + K(k)\upsilon(k) \tag{7}$$

$K(k)$ and $\upsilon(k) = y(k) - H\Phi(k)\hat{\underline{x}}(k-1/k-1)$ are respectively the Kalman gain and the innovation process.

The filter is optimal if the innovation $\upsilon(k)$ is a zero-mean Gaussian white noise. Its covariance matrix then satisfies:

$$C(k) = H^T P(k/k-1)H^T + \sigma_v^2(k) \tag{8}$$

However, the filtering can be completed providing an estimation of the linear prediction coefficients $\hat{\underline{\theta}}(k)$ and the noise variances $\sigma_u^2(k)$ and σ_v^2 are known. For this purpose, we propose to run another Kalman filter in parallel.

2.2 Estimation of the AR parameters from the filtered signal

In this section, we aim at estimating $\underline{\theta}(k)$ from the filtered version of the observations $\hat{s}(k/k)$. Indeed, from equations (7) and (6), $\hat{s}(k/k)$ and $\underline{\theta}(k)$ satisfy:

$$\begin{aligned} \hat{s}(k/k) &= H\left[\Phi(k)\hat{\underline{x}}(k-1/k-1) + K(k)\upsilon(k)\right] \\ &= -\hat{\underline{x}}(k-1/k-1)^T\underline{\theta}(k) + HK(k)\upsilon(k) \\ &= -\hat{\underline{x}}(k-1/k-1)^T\underline{\theta}(k) + \upsilon^*(k) \end{aligned} \tag{9}$$

Besides, the AR parameters can be assumed time-varying and hence be modelled as a random walk:

$$\underline{\theta}(k) = \underline{\theta}(k-1) + \underline{u}^*(k) \tag{10}$$

where $\underline{u}^*(k)$ is a zero-mean Gaussian random vector with covariance matrix Q^*. If Q^* is the null matrix, the signal is assumed stationary.

As a consequence, the set of equations (9) and (10) defines the state space representation for the recursive Kalman-based estimation (denoted by KF2 in Fig. 1) of $\underline{\theta}(k)$:

$$\begin{cases} \underline{\theta}(k) = \underline{\theta}(k-1) + \underline{u}^*(k) \\ \hat{s}(k/k) = -\hat{\underline{x}}(k-1/k-1)^T\underline{\theta}(k) + \upsilon^*(k) \end{cases} \tag{11}$$

where $-\hat{\underline{x}}(k-1/k-1)^T$ is now the observation vector. From equations (8) and (9), the covariance matrix of $\upsilon^*(k)$ is equal to $R^* = HK(k)C(k)K(k)^T H^T$.

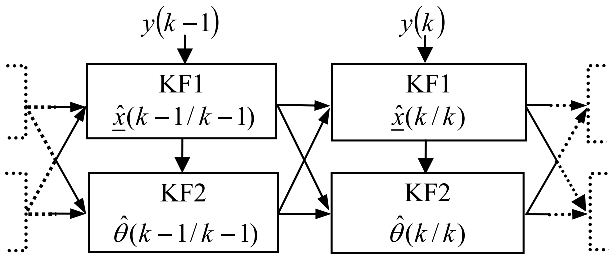

Figure 1: Principal of the proposed algorithm.

In the framework of the instrumental variable techniques [13] [7], $\hat{s}(k/k)$ is considered as an instrument and makes it possible to provide an unbiased estimation of $\underline{\theta}$.

2.3 Estimation of the noise variances

The estimations of the variances $\sigma_u^2(k)$ and $\sigma_v^2(k)$ are necessary to complete this dual Kalman filter based method. An iterative estimation of the variance $\sigma_u^2(k)$ can be derived [11] as follows:

$$\hat{\sigma}_u^2(k) = \frac{k-1}{k}\hat{\sigma}_u^2(k-1) + \frac{1}{k}DL(k)D^T \tag{12}$$

where $D = \left[\Gamma^T\Gamma\right]^{-1}\Gamma^T = \left[1 \quad 0 \quad \cdots \quad 0\right]$

and
$$\begin{aligned} L(k) = {} & P(k/k) - \Phi(k-1)P(k-1/k-1)\Phi(k-1) \\ & + K(k)\upsilon^2(k)K(k)^T \end{aligned}.$$

Besides, in many applications such as speech processing, the variance of the additive noise σ_v^2 can be obtained during non-signal periods, corresponding to the silent frames.

An alternative method consists in using a recursive estimation of $\sigma_v^2(k)$, based on equation (8):

$$\hat{\sigma}_v^2(k) = \frac{k-1}{k}\hat{\sigma}_v^2(k-1) + \frac{1}{k}M(k) \tag{13}$$

where $M(k) = \upsilon^2(k) - HP(k/k-1)H^T$.

3. COMPARATIVE STUDY ON SYNTHETIC DATA

3.1 Protocol

We have carried out a comparative study with five other techniques: the YW and MYW equations [10], Zheng's method [15] and on-line approaches [12] and [14]. A 2000 point 6^{th} order AR process is generated. The corresponding poles are $0.75e^{\pm j0.2\pi}$, $0.8e^{\pm j0.4\pi}$ and $0.85e^{\pm j0.7\pi}$. The sequence is then corrupted by an additive zero-mean white Gaussian noise, whose variance is assumed available. In the results presented here, the Signal-to-Noise Ratio (SNR) is assigned to 5dB.

3.2 Comments

According to Fig. 2, one can notice the biased estimates provided by the YW equations with noisy observations. The MYW equations lead to poor results. In addition, it has been observed that Zheng's method does not necessarily converge when the number of observations is small. This phenomenon may be due to the use of the observation correlation function, whose estimation is poor for large lags.

4. A DERIVED VERSION FOR SPEECH ENHANCEMENT

4.1 AR based speech enhancement issue

Various Kalman filter-based approaches have been proposed to enhance a single sequence of speech contaminated by an additive noise. These methods usually operate as follows:

1. the AR parameters and the variances, σ_u^2 and σ_v^2 are first estimated;
2. the speech signal is then retrieved by means of a Kalman filter.

The approaches essentially differ in the estimation of the AR parameters and the noise variances. Thus, in the pioneering work of Paliwal et al. [9], the AR parameters $\underline{\theta}(k)$, $\sigma_v^2(k)$ and σ_v^2 are respectively estimated from the clean speech and the noise sequence, both assumed to be available. However, this approach cannot be computed in practice. In [4], the AR parameters and the noise variances are estimated from the noisy speech and Kalman filtering is carried out to enhance the observations.

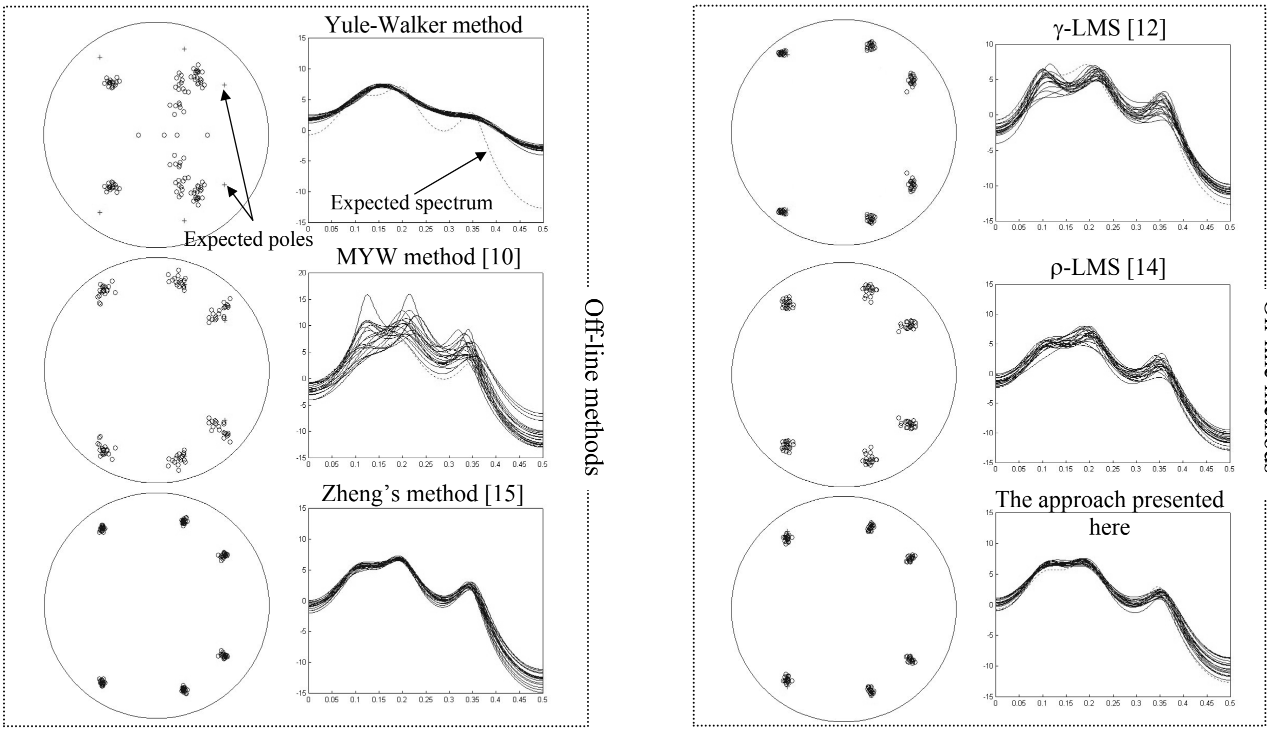

Figure 2: AR parameter estimation for various off-line and on-line approaches; SNR=5dB (20 realisations).

Comparing to other on-line methods, our approach provides estimated spectra closer to the expected one. Indeed, although the LMS-based methods have low computational cost, they converge slower and with higher variance than our technique. The Kalman filter makes it possible to provide minimal variance estimates, contrary to gradient-based filters. In addition, the proposed technique has the advantage to estimate both the parameters and the signal.
Therefore, this drives us to exploit the proposed technique in the framework of speech enhancement.

Then, an iterative process aims at estimating the parameters from the enhanced signal and using them to perform anew Kalman filtering with the noisy signal. However, the authors in [4], [9] do not explicitly address the way the speech parameters can be obtained. Besides, many other Kalman filter-based algorithms have since been proposed [2] [3] [5]. For each method - except in [9] - the enhanced speech is usually contaminated by a residual noise. This noise, responsible of the so-called "musical phenomenon", can be weakened by using a Kalman smoothing [2].

4.2 Proposed algorithm

Based on the results obtained in section 3, we now propose to investigate a derived version of the presented dual Kalman filter-based method for speech enhancement.

In this algorithm, the signal, the parameters and the variance of the driving process are simultaneously estimated. Therefore, the algorithm convergence rate may be quite low. However, when processing speech (frame by frame), only a limited number of samples are available (256 samples at 8 KHz). To obtain a significant number of observations, we propose an iterative approach which processes the noisy frame, as follows:

1. the dual Kalman filter-based estimator is carried out to provide a first estimation of the model parameters, with no specific initial conditions;
2. the same frame is processed again, but the parameter initial conditions are here the previously estimated values.

Besides, the variance of the additive noise is estimated during non signal periods.

4.3 Protocol and results

The real French sentence /*Le tribunal va bientôt rendre son jugement*/, sampled at 8 kHz is contaminated by an additive white noise, with various SNR ranked from 5 to 15 dB. Results are based on 100 realisations.

SNR input (in dB)	Paliwal[1] [9]	Gibson[1] [4]			Proposed algorithm		
		iteration			iteration		
		#1	#2	#3	#1	#2	#3
5	4.9	3,2	4.3	4.5	2.9	4.3	4.4
10	3.4	2,4	3,0	3,1	1.9	3.0	3.0
15	2,2	1,7	2,0	2,1	1.7	2.0	2.0

Table 1: Average SNR improvements (in dB).

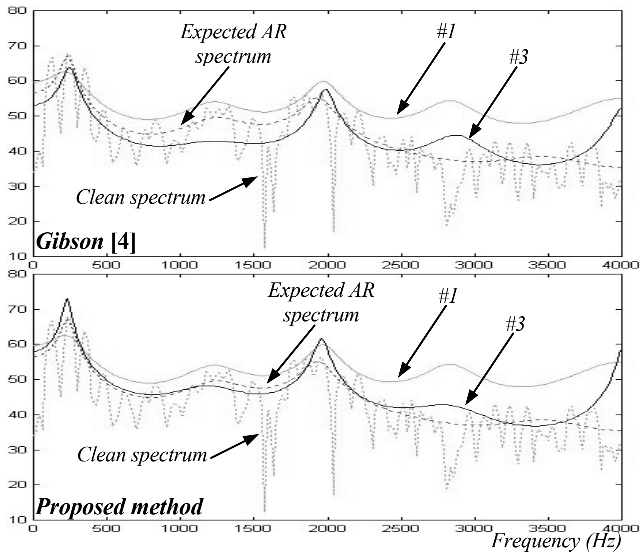

Figure 3: Clean and estimated AR spectra for one processed frame (phoneme /JU/ in French); SNRinput=10 dB.

When comparing the SNR improvements in Table 1, Gibson's approach and our algorithm have similar performances. Informal subjective listening tests confirm that the differences between the enhanced signals are not really audible. Nevertheless, the AR spectrum estimated with the proposed approach exhibits resonances that are comparatively sharper than Gibson's one (Cf. Fig. 3).

[1] The AR parameters are estimated with the YW equations.

5. CONCLUSION

In this paper, the AR parameter estimation issue from noisy observations is addressed. The proposed solution is based on two Kalman filters running in parallel and avoids a non linear approach, for instance based on the EKF. This method has the advantage of providing sequential unbiased estimates of the AR parameters. This is confirmed by a comparative study based on synthetic data. Then, a derived version of the algorithm is proposed in the framework of speech enhancement.

REFERENCES

[1] C. E. Davila, "A subspace approach to estimation of autoregressive parameters from noisy measurements", IEEE Trans. on SP, vol. 46, n°2, pp. 531-534, February 1998.

[2] S. Gannot, D. Burchtein, E. Weinstein, "Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms", IEEE Trans. on SAP, vol. 6, n°4, pp. 373-385, July 1998.

[3] E. Grivel, M. Gabrea, M. Najim, "Speech Enhancement as a realization issue", Signal Processing, vol. 82, pp. 963-1978, December 2002.

[4] J. D. Gibson, B. Koo, S. D. Gray, "Filtering of colored noise for speech enhancement and coding", IEEE Trans. on SP, vol. 39, n° 8, pp. 1732-1742, August 1991.

[5] Z. Goh, K. C. Tan, B. T. G. Tan, "Kalman-Speech Enhancement Method Based on a Voiced-Unvoiced Speech Model", IEEE Trans. on SAP, vol. 7, n°5, pp. 510-524, September 1999.

[6] S. Haykin (Ed), Kalman Filtering and Neural Networks, Wiley-Interscience, 2001, chap. 5.

[7] M. Kendall, A. Stuart, The advanced theory of statistics, Griffin London, 1979.

[8] L. W. Nelson, E. Stear, "The simultaneous On-Line Estimation of Parameters and States in Linear Systems", IEEE Trans. on AC, vol. 21, n°2, pp. 94-98, February 1976.

[9] K. K. Paliwal, A. Basu, "A Speech Enhancement Method Based on Kalman Filtering", Proc. ICASSP'87, vol. 1, pp.177-180.

[10] C. W. Therrien, Discrete Random Signals and Statistical Signal Processing, Prentice Hall, 1992.

[11] E. Todini, "Mutually Interactive State/Parameter Estimation (MISP) – Application of Kalman Filter to Hydrology, Hydraulics and Water Resources", Proc. AGU Chapman Conf., Univ. of Pittsburg, May 1978.

[12] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer", IEEE Trans. on ASSP, vol. 27, n°1, pp- 53-62, February 1979.

[13] K. W. Wong, E. Polak, "Identification of Linear Discrete Time Systems Using the Instrumental Variable Method", IEEE Trans. on AC, vol. AC-12, n°6, pp. 707-718, December 1967.

[14] W-R. Wu and P-C Chen, "Adaptive AR modeling in White Gaussian Noise", IEEE Trans. on SP, vol. 45, n°5, pp. 1184-1191, May 1997.

[15] W. X. Zheng, "Autoregressive Parameter Estimation from Noisy Data", IEEE Trans. on Circuits and Systems II: Analog and Digital signal processing, vol. 47, n°1, January 2000.