AUTOREGRESSIVE MODELS FOR NOISY SPEECH SIGNALS

Y. GRENIER, K. BRY, J. LE ROUX, M. SULPIS

DEPARTEMENT SYSTEMES ET COMMUNICATIONS
ENST 46 Rue BARRAULT
75634 PARIS CEDEX 13 FRANCE

ABSTRACT

Linear prediction is a well extended technique for transmission, synthesis and recognition. However when the signal is corrupted by noise, the estimation of the autoregressive model is known to be biaised.

This paper is devoted to methods allowing a reduction of this bias. We will consider first a global method, in which the Yule Walker equations are modified to take into account the variance of an additive white noise. The problem becomes non-linear and is solved recursively. In a second approach, we will examine a time - recursive method based on Kalman filtering.

1. INTRODUCTION

Estimation of autoregressive models leads to performing techniques in signal processing (maximum entropy spectral estimation, low rate speech coding, speech synthesis ...).
Among the reasons for the success of such techniques, one can find the easy way through which the model is obtained : solution of a linear system of equations, and moreover with a Toëplitz matrix leading to fast (Levinson) algorithms. Another reason is the fact that these models give an estimate of the spectrum under a "resonant" form, (generated through an all-pole system) and this is a kind of spectrum encountered in most physical phenomena.

Unfornately, when the autoregressive process to be estimated is disturbed by an additive noise, it has been shown that the bias of the estimator is no longer zero and increases with a decreasing signal to noise ratio. The existence of an additive noise is a commonly encountered situation : ambient noise added to a speech signal when recorded through a microphone, quantization noise ... . Due to this great practical importance, this problem has received a lot of attention in the last years, which focused in several directions : evaluation of the quality of the estimators in noisy conditions (2), (3), algorithms for the reduction of bias (6), (9), use of more complex models to estimate simultaneously AR model and noise (4), (5). We give in Section 2 of this paper a short review of the problems arising in this context, and methods to solve them. In section 3 we investigated a new solution to the problem of estimating the coefficients of the autoregressive model, from the autocorrelation function of the noisy process. In Section 4 we formulate a recursive method in which the adjustment of parameters is performed after the acquisition of each new sample through a single step of a Kalman filter. In Section 5 we will discuss the properties of these two methods.

2. EFFECT OF NOISE ON THE ESTIMATES

Let us suppose that the process we are working with, is scalar, centered, gaussian, (This hypothesis is not necessary but may simplify some of the calculus).
The signal y_t ($t \varepsilon Z$) is modelized as an autoregressive process if there exist an order p, p parameters $a_1 \dots a_p$, and a white noise ε_t with variance σ_ε^2 such that

$$\forall t \; : \; y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = \varepsilon_t \qquad (1)$$

The identification of the model is the determination of p, $a_1 \dots a_p, \sigma_\varepsilon^2$ minimizing some criterion. The most commonly used criterions are : the quadratic mean error, the maximum likelihood of the observations $y_0 \dots y_T$ and the maximum entropy of the estimated spectrum. They lead in the gaussian case (ε_t and y_t gaussian) to the same solution : the Yule-Walker equations.

$$\begin{bmatrix} R_{yy}(0) & & R_{yy}(p) \\ & \ddots & \\ R_{yy}(p) & & R_{yy}(0) \end{bmatrix} \times \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (2)$$

$R_{yy}(k)$ being the k-th correlation coefficient of the signal y_t :

$$R_{yy}(k) = E\,(y_t \, y_{t+k}) \qquad (3)$$

The spectrum estimate associated to this model is the following one :

$$\hat{S}(\omega) = \left[\frac{\sigma^2_\varepsilon}{A(z)\;A(z^{-1})} \right]_{z = e^{j\omega}} \qquad \omega \varepsilon \left[-\pi, +\pi\right] \qquad (4)$$

where A(z) is the z-transform of the sequence of

CH1610-5/81/0000-1093 $00.75

the model coefficients.

$$A(z) = 1 + a_1 z^{-1} + \dots + a_p z^{-p} \tag{5}$$

Let us consider now the case where the observed signal y_t is the summation of an autoregressive process x_t and a white noise v_t

$$y_t = x_t + v_t, \quad E(v_t) = 0, \quad E(v_t^2) = \sigma_v^2$$

We will assume that the noise v_t and the "useful" signal x_t are statisticaly independent. It is easy to compute the autocorrelation $R_{yy}(k)$ of y_t, from that of x_t.

$$R_{yy}(k) = E(y_t y_{t+k}) = E(x_t x_{t+k}) + E(x_t v_{t+k}) + E(v_t x_{t+k}) + E(v_t v_{t+k})$$

which leads to

$$R_{yy}(0) = R_{xx}(0) + \sigma_v^2$$
$$R_{yy}(k) = R_{xx}(k) \qquad \forall k \neq 0 \tag{6}$$

These equations show immediately that the estimation of the autoregressive model of the signal x_t is biased, for the equations of Y.W use the matrix with R_{yy} instead of R_{xx}. The correct matrix should be

$$\mathbb{R}_x = \begin{bmatrix} R_{xx}(0) & \ddots & R_{xx}(p) \\ R_{xx}(p) & & R_{xx}(0) \end{bmatrix}$$

but we use $\mathbb{R}_y$:

$$\mathbb{R}_y = \begin{bmatrix} R_{yy}(0) & \ddots & R_{yy}(P) \\ R_{yy}(P) & & R_{yy}(0) \end{bmatrix} = \mathbb{R}_x + \sigma_v^2 I \tag{7}$$

We can also intuitively deduce from (7) that the bias will be increasing with σ_v^2. This was shown formaly by Sakai and Arase (6).

The relations (4) and (6) allow us to retrieve the effect of noise on the estimated spectrum. The power spectrum of y_t is :

$$S_{yy}(\omega) = S_{xx}(\omega) + \sigma_v^2 \tag{8}$$

also :

$$S_{yy}(\omega) = \left[\frac{\sigma_\varepsilon^2}{A(z)\, A(z^{-1})} \right]_{z=e^{j\omega}} + \sigma_v^2 \tag{9}$$

The effect of noise is therefore the addition of a constant to the spectrum, which acts as a flattening effect on the log spectrum, an increase of the bandwidth for each formant, these phenomena being sharper with low signal to noise ratios. Another interpretation of formula (9) appears when writing S_{yy} as :

$$S_{yy}(z) = \left[\frac{\sigma_\varepsilon^2}{A(z)\, A(z^{-1})} \right] + \sigma_v^2 = \frac{B(z)\, B(z^{-1})}{A(z)\, A(z^{-1})} \tag{10}$$

where

$$B(z)\, B(z^{-1}) = \sigma_\varepsilon^2 + \sigma_v^2\, A(z)\, A(z^{-1}).$$

Under this formulation, we see that the observed spectrum is the same as that of an ARMA process $\frac{B(z)}{A(z)}$. Pagano (4) and Kay (3) have shown the validity of this interpretation : the zeros of the ARMA model are responsible for the flattening of the spectrum, they are positionned inside the unit cercle of the z-plane, and when the signal-to-noise ratio varies from 0 to $+\infty$, the zeros describe continuous curves joining the poles (SNR = 0) to the origin (SNR = $+\infty$).

Two class of methods allow an unbiased estimation of the model :

- bias compensation by estimation of $(\underline{a}-\hat{\underline{a}})$ where $\underline{a}$ (resp. $\hat{\underline{a}}$) is the true (resp. estimated) vector of parameters.

- identification of the AR part of an ARMA (or other) model including the additive noise.

The rest of the paper is devoted to this second case.

3. ESTIMATION BY A GLOBAL METHOD

The signal x_t being autoregressive, its autocorrelation satisfies the equation :

$$R_{xx}(k) + a_1 R_{xx}(k-1) + \dots + a_p R_{xx}(k-p) = 0 \tag{11}$$

for every $k \neq 0$. For $k = 0$ the equation becomes

$$R_{xx}(0) + a_1 R_{xx}(1) + \dots + a_p R_{xx}(p) = \sigma_\varepsilon^2 \tag{12}$$

where σ_ε^2 is the variance of the innovation ε_t and using the fact that R_{xx} is symetrical. The autocorrelation is estimated as

$$R_{xx}(k) = \frac{1}{T+1} \sum_{t=0}^{T-k} x_t x_{t+k} \tag{13}$$

Using the equation (11) for values of k between 1 and p leads to the Yule-Walker equations, and to Levinson algorithm. However in our case (noisy signals), R_{xx} is unknown, we take R_{yy} in place of R_{xx}, for it differs from R_{xx} only by the coefficient

$R_{yy}(0) = R_{xx}(0) + \sigma_v^2$. Needing q equations to solve for $a_1 \dots a_p$, we can take the p first equation (11) which do not include $R_{yy}(0)$.

The system of equations becomes

$$\begin{bmatrix} R_{xx}(p+1) & \ddots & R_{xx}(P) & \ddots & R_{xx}(1) \\ R_{xx}(2p) & & & & R_{xx}(P) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{14}$$

But this is nothing else than the equations for the determination of the AR part of the ARMA model (8) ; we are implicitely refering to the AR + noise process as an ARMA process. The main difficulty with this algorithm is the use of badly estimated correlation coefficients $R_{xx}(k)$ with k large : the variance of the (unbiased, see (8)) estimator is larger (9).

We suggest here the use of a method which uses equation (11) for values k = 1 ... k = p, adding an equation which provides an estimate for σ_v^2. In order to avoid an increase in the variance of the estimation, we chose equation (11) with k = p + 1, obtaining the following system :

$$\begin{bmatrix} R_{yy}(1) & R_{yy}(0)-\sigma^2 & \cdots & R_{yy}(p-1) \\ & \ddots & \ddots & \vdots \\ & & \ddots & R_{yy}(0)-\sigma^2 \\ R_{yy}(p+1) & & & R_{yy}(1) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \quad (15)$$

Let $\underline{\theta}$ be the vector of parameters:

$$\underline{\theta}^T = [\sigma_v^2 \; a_1 \; \ldots \; a_p]$$

$\underline{\theta}$ will be computed recursively through the algorithm

$$\underline{\theta}_{n+1} = \underline{\theta}_n + G \cdot f(\theta n) \quad (16)$$

where f (θ) is the function of θ we want to set to zero, that is the left part of equation (15) which may be rewritten as :

$$f(\theta) = \begin{bmatrix} (R - \sigma_v^2 . I)\,\underline{a} + \underline{r} \\ \underline{r}^T J \,\underline{a} + R_{yy}(p+1) \end{bmatrix}$$

$$R = \begin{bmatrix} R_{yy}(0) & \cdots & R_{yy}(p-1) \\ \vdots & \ddots & \vdots \\ R_{yy}(p-1) & \cdots & R_{yy}(0) \end{bmatrix} \quad \underline{r} = \begin{bmatrix} R_{yy}(1) \\ \vdots \\ R_{yy}(p) \end{bmatrix} \quad \underline{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}$$

$$J = \begin{bmatrix} 0 & & 1 \\ & ⋰ & \\ 1 & & 0 \end{bmatrix}$$

The gain matrix can be taken as $(\frac{1}{x} I)$, leading to a gradient algorithm, but it is more efficient to choose :

$$G = \left[\frac{\partial f}{\partial \sigma^2} \; \vdots \; \frac{\partial f}{\partial a_1} \; \vdots \; \cdots \; \vdots \; \frac{\partial f}{\partial a_p} \right]^{-1} \quad (18)$$

We obtain thus a second-order algorithm of the Newton-type. It is easy to compute the matrix G in terms of R ,$\underline{r}$, $\underline{a}$, σ_v^2 :

$$G = \left[\begin{array}{c|c} -\underline{a} & R - \sigma_v^2 . I \\ \hline 0 & \underline{r}^T J \end{array} \right]^{-1}$$

The inversion formule for a block (2x2) matrix applied to matrix G permits to derive the following algorithm.

Algorithm

$$\gamma(n) = -\underline{r}^T J(R-\sigma^2(n)I)^{-1}\,\underline{r} + R_{yy}(p+1)$$

$$\alpha(n) = \underline{r}^T J(R-\sigma^2(n)I)^{-1}\,\underline{a}(n)$$

$$\underline{a}(n+1) = -(R-\sigma^2(n)I)^{-1}\left(\underline{r} + \frac{\gamma(n)}{\alpha(n)}\,\underline{a}(n)\right) \quad (20)$$

$$\sigma^2(n+1) = \sigma^2(n) - \frac{\gamma(n)}{\alpha(n)}$$

No explicit matrix inversion is necessary, for $(R - \sigma^2(n)I)$ is a Toeplitz matrix and the Levinson algorithm gives the solution of the linear equations appearing in (20).

4. RECURSIVE METHOD

The preceeding approach which was derived for batch processing can be duplicated for on-line processing of the signal. We could for instance estimate the a_i coefficients as the AR part of an ARMA model estimated through any recursive ARMA method (10), (11), but we prefer to identify a more specialized model, using a Kalman filter to identify the parameters (12), (13). The state equations linking the observations to the parameters are augmented by an equation incorporating the additive noise. The model is the following one :

Model :

$$\xi_t = [a_1 \; \ldots \; a_p \; v_t]^T$$

$$\begin{cases} \xi_{t+1} = \Phi \xi_t + \underline{V}_{t+1} \\ y_t = C_t \xi_t + \varepsilon_t \end{cases} \qquad \Phi = \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix} \quad \underline{v}_t = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ v_t \end{bmatrix}$$

$$C_t = [-x_{t-1} \; \ldots \; -x_{t-p} \; 1]$$

The estimation of the state ξ_t (parameters + noise) by Kalman filtering is not easy, because the values of x are not known, so that C_t has to be replaced by $\hat{C}_t$ where $\hat{x}_{t-i}$ is $y_{t-i} - \hat{v}_{t-i}$

$$C_t = [\hat{v}_{t-1} - y_{t-1} \;,\; \ldots \;,\; \hat{v}_{t-p} - y_{t-p} \;,\; 1]$$

With this approximation, the derivation of a Kalman filter for the estimation of the state ξ_t (parameters + noise) of this model, is trivial, and will not be given here.

Conclusion

The two estimators described in Sections 3 and 4 have been evaluated on synthetic signals and also on speech signals. The first experiment consisted in the building of a synthetic AR process from a known model. A white noise was added to this signal, its variance was choosen to obtain a prescribed signal-to-noise ratio

$\left[+\infty, \text{10dB, 3dB, 0dB, -3dB}\right]$

The modelisation was performed through classical AR method (Levinson) and through each of the two new methods. In the second experiment, high quality speech signals were modelized through classical AR method, and the result was taken as a reference to which were compared the results of our two methods, after addition of the white noise (fig. 1).

In both cases, it appears that no method is entirely satisfactory for signal-to-noise ratios less than 3dB, if the length of the analysis window is limited to 1000 samples. The estimation is of course better with larger windows. For signal to noise ratios greater than 3dB, the method described in section 3 gives good results but there may be some problems with the initial value of the noise variance $\sigma_v^2(0)$: in some cases the estimation converged towards absurd values (noise variance negative), but a second use of the algorithm with another value of $\sigma_v^2(0)$ provided in all cases a good convergence. The use of this method gave better results than the use of the method estimating the AR part of an ARMA model.

The recursive method also gave good results but to the cost of an enormous increase in computation time. The first method converged in less than 10 iterations. However both methods do not explicit a criterion which would be minimized, the theoretical properties of these algorithms are not easily derived. This remains to be done.

References :

(1) J. MAKHOUL. Linear Prediction : a Tutorial Review. Proceedings of IEEE. Vol. 63, n° 4, pp 561-580 (April 1975).

(2) B. YEGNANARAYANA, T.K. RAJA. Performances of LP Analysis of Speech with Additive Noise. IEEE Conf on ASSP, pp 20-23 (1977).

(3) S. KAY. The Effect of Noise on the AR Spectral Estimator. IEEE Trans. on ASSP. Vol 27, n° 5, pp 478-485 (October 1979).

(4) M. PAGANO. Estimations of Models of AR signals plus white noise. The Annals of Statistics. Vol. 2, n° 1, pp 98-108 (January 1974).

(5) W.J. DONE, C.K. RUSHFORTH. Estimating the Parameters of a Noisy All-Pole Process using Pole-Zero Modelling. IEEE Conf on ASSP. pp 228-231 (1979).

(6) M. SAKAI, M. ARASE. Recursive Parameter Estimation of an AR Process disturbed by White Noise. Int. J. of Control. Vol 30, n° 6, pp 949-966 (1979).

(7) J.S. Lim, A.V. Oppenheim All-Pole Modeling of Degraded Speech. IEEE Trans. on ASSP, Vol. 26, n° 3 pp 197-210 (June 1978).

(8) W. GERSH. Estimation of the AR part of a mixed ARMA time series. IEEE Trans on AC, Vol.15 pp 583-588. (October 1970).

(9) S. KAY. Noise Compensation for AR Spectral Estimates. IEEE Trans on ASSP, Vol. 28, n° 3, pp 292-303 (June 1980).

(10) T. SODESTROM, L. LJUNG, I. GUSTAVSSON. A Theoretical Analysis of Recursive Identification Methods. Automatica, Vol-14, pp 231-244 (1978).

(11) M. MORF, D.T. LEE, J.R. NICKOLLS, A. VIEIRA. A Classification of Algorithms for ARMA models and Ladder Realisation. IEEE Conf on ASSP. pp 13-19 (1977).

(12) C. GUEGUEN, G. CARAYANNIS. Analyse de la Parole par filtrage optimal de Kalman. Automatisme. Vol 18. n° 3. pp 99-105 (Mars 1973).

(13) J.D. GIBSON, J.L. MELSA, S.K. JONES. Digital Speech Analysis using Sequential Estimation Techniques. IEEE Trans on ASSP, Vol. 23, n° 4, pp 362-369 (August 1975).

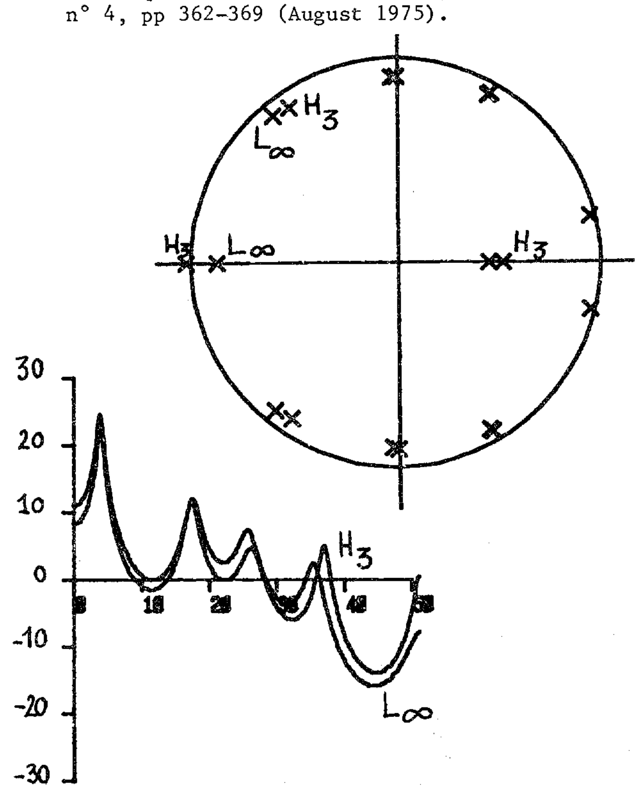

fig. 1.

Method of section 3 (SNR = + 3dB) H_3 compared to Levinson (SNR = $+\infty$) L_∞