NEUVIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 16 au 20 MAI 1983

'SISSI' - A silent input selective sequential identifier for AR systems

'SISSI' - Un identificateur sequenciel pour les systèmes autoregressifs avec certaines excitations corrélées

J.P. Laebens and J.R. Deller, Jr.

Illinois Institute of Technology
Dept. of Electrical Engineering: Speech Processing Laboratory
Chicago, IL 60616 USA

RESUME

Les systèmes autorégressifs modélisés par l'équation :

$$X(n) = \sum_{i=1}^{P} a_i \, x(n-i) + v(n) \qquad (1)$$

ne peuvent être identifiés sans biais que dans le cas simple où v(n) est une séquence d'innovations non corrélées. Cette communication montre une méthode adaptée au cas particulier où l'entrée v(n) est telle que :

$$v(n) = u(n) + w(n) \qquad (2)$$

où u(n) est une séquence intermittente quelconque, telle que u(n) = 0, pour $n \in \mathcal{N}$, et w(n) est un bruit blanc de variance σ. Cette situation correspond à plusieurs applications dont la déconvolution des signaux sismiques et l'estimation de l'excitation laryngienne dans le processus de la production de la parole.

On montre d'abord que si la fonction de coût utilisée pour la détermination des coefficients a_i est définie comme la somme restreinte sur des erreurs de prédiction au carré, alors l'identification converge sans biais. Une forme récursive est obtenue qui ne nécessite pas la connaissance préalable de l'ensemble mais uniquement d'un court segment de où l'algorithme est initié. Lors de cette initiation, la valeur de σ est estimée puis les observations restantes sont incluses dans la fonction coût ou ignorées selon comparaison de l'erreur de prédiction avec σ. La détection de la région d'initiation, faisant l'objet d'une analyse préliminaire, est rendue aisée par d'éventuelles caractéristiques temporelles de l'entrée u(n). Par exemple, dans le cas de l'excitation laryngienne, la fermeture du larynx provoquant une brusque tombée de u(n), peut être détectée à partir d'une simple analyse préliminaire du signal x(n).

Enfin, simplicité de programmation et faible nombre d'opérations sont obtenus grâce à l'utilisation de transformations de Givens, et rendent le procédé aisément implantable sur petits systèmes fonctionnant en temps réel.

SUMMARY

The identification of systems following the model

$$(1) \qquad x(n) = \sum_{i=1}^{P} a_i x(n-i) + v(n)$$

yields a biased estimate of the a_i parameters if the excitation signal v(n) is non white. This paper presents a method that produces exact results when v(n) is of the form:

$$(2) \qquad v(n) = u(n) + w(n)$$

where u(n) is a general sequence with the property that u(n) = 0 for some large set $\mathcal{N}$ and w(n) is white noise with variance σ^2. Such inputs occur in several practical problems such as seismic deconvolution or glottal waveform estimation in speech processing.

It is first shown that if the cost function used in the determination of the parameters a_i is redefined as the squared errors summed over n's restricted to those in $\mathcal{N}$, then the identification is unbiased. The a priori knowledge of $\mathcal{N}$ is shown unnecessary if a selective recursive least squares algorithm is used. The observations in $\mathcal{N}$ are selected on-line, using a comparison between the least squares residual and an estimate for σ. obtained during an initiation period. The detection of this period employs some additional time domain features of u(n).

Finally, the method is made computationally efficient by employing Givens' Reduction in the recursion.

Acknowledgement

This work was supported in part by the National Science Foundation of the United States under Grant No. ECS-8006866.

1. INTRODUCTION

It is well known that attempts to identify autoregressive (AR) systems of the form

$$(1)\quad x(n) = \sum_{i=1}^{P} a_i x(n-i) + v(n) \triangleq \underline{x}^T(n)\underline{a} + v(n)$$

using conventional methods [1] yield biased estimates of the a_i parameters if the driving signal, v(n), is correlated [2]. In this paper, we present a modified AR (or linear predictive coding (LPC)) identification approach which produces asymptotically exact results for a certain class of inputs:

$$(2)\quad v(n) = u(n) + w(n)$$

where u(n) is a general sequence (often deterministic) with the property that u(n)=0 for some large set, say $\mathcal{N}$, of, generally nonsequential, n's, and w(n) is a zero mean, variance σ^2, second order ergodic, uncorrelated sequence (often included to account for model errors). Random impulse trains used in seismic models, and periodic impulses or glottal waveform pulses used in the modelling of voiced speech are example inputs from this class which occur in practical problems. The results derived for the general case in this work are readily applicable to such specific applications.

2. GENERAL THEORY

The "covariance" approach yields the estimates of the a_i parameters given by the familiar set of equations [1,p.14]:

$$(3)\quad \sum_{i=1}^{P} \hat{a}_i \phi_{xx}(N;i,j) = \phi_{xx}(N;0,j) \mid j = 1, \cdots, P$$

in which, for any sequences x and y, the <u>covariance function</u> is defined as

$$(4)\quad \phi_{xy}(N;i,j) \triangleq \frac{1}{N}\sum_{n=0}^{N-1} x(n-i)y(n-j)$$

Eqns. (3) are, in turn, the solution to:

PROBLEM (P): Derive the coefficients of the P-order linear predictor of the sequence x(n) which minimizes

$$(5)\quad J(N) \triangleq \sum_{n=0}^{N-1} e^2(n) = N\,\phi_{ee}(N;0,0) \triangleq \underline{E}^T(N)\underline{E}(N)$$

where $e(n) \triangleq x(n) - \sum_{i=1}^{P} \hat{a}_i x(n-i)$ = the prediction error at time n, and where $\underline{E}(N) \triangleq [e(0),\ldots,e(N-1)]^T$

Although the scalar solution, (3), is widely popular in the speech processing literature, the alternative formulation of the covariance solution is more useful in subsequent developments:

$$(6)\quad \hat{\underline{a}} = [\underline{A}^T(N)\underline{A}(N)]^{-1}\,\underline{A}^T(N)\,\underline{X}(N)$$

where,

$$\underline{X}(N-i) \triangleq [x(-i)x(1), \cdots x(N-1-i)]^T$$
$$\underline{A}(N) = [\underline{x}(N-1) \cdots \underline{x}(N-P)]$$

This solution is obtained by direct manipulation of (3) or by solving the prediction problem from the more classical "batch least squares" approach (see, eg., [4,Ch.6]). NOTE: Throughout this discussion we shall assume that the matrix inverse in (6) exists. Conditions on the verity of this assumption are given in [2].

Now consider solving (P) subject to the minimization of the modified cost function

$$(7)\quad J'(N) = \sum_{n=0}^{N-1} q(n)e^2(n) = \underline{E}^T(N)\,\underline{Q}(N)\,\underline{E}(N)$$

in which $\underline{Q}(N) \triangleq \mathrm{diag}(q(0),q(1),\ldots,q(N-1))$ and q(n) is any weighting sequence. This is the classic weighted least squares prediction problem for which the solution is [4,Ch.6]:

$$(8)\quad \hat{\underline{a}}(N) = [\underline{A}^T(N)\underline{Q}(N)\underline{A}(N)]^{-1}\underline{A}^T(N)\underline{Q}(N)\underline{X}(N)$$
$$= [\underline{\phi}'(N)]^{-1}\{\tfrac{1}{N}\underline{A}^T(N)\underline{Q}(N)\underline{X}(N)\}$$

in which, $\underline{\phi}'(N) \triangleq (1/N)\{\underline{A}^T(N)\underline{Q}(N)\underline{A}(N)\}$.
We now prove the fundamental result.

THEOREM: The identification of the AR system of (1) by (8) is asymptotically exact if the q(n) are chosen such that $q(n) = I_{\mathcal{N}}(n)$, where $I_{\mathcal{N}}$ is the indicator function for the set $\mathcal{N}$.

<u>Discussion</u>. Apparently, the approach suggested by the theorem requires minimization of the error only on the data samples, x(n), for which $n\epsilon\mathcal{N}$, ie., the criterion becomes

$$(9)\quad \text{minimize}\{J'(N) = \sum_{n=0}^{N-1} q(n)e^2(n) = \sum_{\substack{n=0\\ n\epsilon\mathcal{N}}}^{N-1} e^2(n)$$

It is not difficult to show that, in the scalar formulation of the problem, this modified criterion results in a solution of form (3) where the covariance functions are replaced by modified functions:

$$(10)\quad \phi'_{xx}(N;i,j) = \frac{1}{N}\sum_{\substack{n=0\\ n\epsilon\mathcal{N}}}^{N-1} x(n-i)x(n-j)$$

PROOF: Using (1), (2), and the definitions in (6)

$$(11)\quad \underline{X}(N) = \underline{A}(N)\underline{a} + \underline{U}(N) + \underline{W}(N)$$

where, $\underline{U}(N)=[u(0),\ldots,u(N-1)]^T$, and $\underline{W}(N)=[w(0),\ldots,w(N-1)]^T$. Now using (8), we have that

$$(12)\quad \hat{\underline{a}}(N) = \underline{a}+[\underline{\phi}'(N)]^{-1}\{\tfrac{1}{N}\underline{A}^T(N)\underline{Q}(N)\underline{U}(N) + \tfrac{1}{N}\underline{A}^T(N)\underline{Q}(N)\underline{W}(N)\}$$

Using theorems in [2] it is possible to argue that $\underline{\phi}^{-1}(N)$ remains bounded as $N\to\infty$ for all reasonable data sequences, x(n). The last two terms in the brackets are N-vectors with jth-elements

$$(13)\quad \phi'_{xu}(N;j,0) = \frac{1}{N}\sum_{n=0}^{N-1} x(n-j)q(n)u(n) = \frac{1}{N}\sum_{\substack{n=0\\ n\epsilon\mathcal{N}}}^{N-1} x(n-j)u(n)$$

$$\phi'_{xw}(N;j,0) = \frac{1}{N}\sum_{n=0}^{N-1} x(n-j)q(n)w(n) = \frac{1}{N}\sum_{\substack{n=0\\ n\epsilon\mathcal{N}}}^{N-1} x(n-j)w(n)$$

Clearly, both of these expressions are asymptotically zero and, hence, the rightmost term vanishes in the limit. QED

Finally we note that $\hat{\underline{a}}(N)$ can be computed recursively in a sequential weighted least squares formulation based on (8) [4]. The reader is referred to [5] for specific equations. We omit these here since a more efficient recursive method is developed in the following section.

3. SISSI

<u>Sequential detection of $\mathcal{N}$</u>. From (5) and (1), the error in prediction at time n based on $\hat{\underline{a}}(N)$ is given by

(14) $e_N(n) = \underline{x}^T(n)\{\underline{a}-\hat{\underline{a}}(N)\} + v(n)$

In particular, consider the decision as to whether $N \in \mathcal{N}$ (recall that $\hat{\underline{a}}(N)$ is based on the interval $n \in [0,N-1]$). If $\hat{\underline{a}}(N)$ is a good estimate of $\underline{a}$, then from (14)

(15) $e_N(N) \simeq v(N) = u(N) + w(N)$

The decision process is based on the fact that $e_N(N)$ is approximately the model noise at any $n \in \mathcal{N}$ and should therefore reflect the stochastic properties of the noise alone. Assume, for example, that w(n) is mean zero, variance σ^2, and that, for $n \notin \mathcal{N}$, the average value of u(n) >> σ. Then a typical decision scheme is

(16) SELECT $\{n \in \mathcal{N}$ if $e^2(n) \leq k\sigma^2$; $n \notin \mathcal{N}$ otherwise$\}$

The value of k is chosen in accordance with the properties of u(n). Such decision processes are highly problem dependent and can be much more sophisticated than (16). Some specific examples from the class of inputs (2) are discussed in [5]. Any such process is dependent upon the solution of a more difficult problem, the detection of an appropriate temporal region in which to initiate the recursion. The initial time values must be in $\mathcal{N}$ in order to assure unbiased convergence.

Detection of the initiation set. In the absence of useful temporal knowledge, it is possible to extend the argument above as follows: Let $\hat{\underline{a}}^s(N)$ be the AR solution beginning at time n=s. (Obtained using (3) or (6) with a readjusted time origin.) Let

(17) $J^s(N) \triangleq \sum_{n=s}^{s+N-1} e^2(n)$

Then it is plausible, and has been theoretically justified in [5], that

(18) $\mathcal{S} \triangleq [s, s+N+1] \subset \mathcal{N} \Rightarrow J^s(N) \simeq N\sigma^2$

The discovery of a temporal region $\mathcal{S}$ for which $J^s(N)$ has this property yields an appropriate initial estimate (assuming that $J^s(N)$ is known to have a much different value for $\mathcal{S} \not\subset \mathcal{N}$. This idea is similar to that employed by Wong in the determination of the "closed phase" of the glottal excitation in speech processing [6].

The analysis suggested above is generally complex and expensive. In practice, one must take advantage of available time domain features of u(n) to derive an efficient procedure. As an example, consider the case in which u(n) contains rapid changes in time. Such waveforms occur in the modelling of speech in which discrete time impulse trains or lowpass "glottal" pulse trains with sharp falling edges [7] are used to excite a resonant vocal tract model with impulse response h(n). Laebens has shown in this case that [5]

(19) $e(n) = h(n) * (a_n-\hat{a}_n) * u(n) + u(n)$

Since $h(n)*(a_n-\hat{a}_n)$ represents an averaging process in cascade with the AR system, the first term in the sum cannot contain high frequencies and jumps in e(n) must therefore correspond temporally with those in u(n), the second term. Such jumps in e(n) in this example can serve as indicators of a silent excitation period to follow.

Having achieved the means to obtain a good initial estimate for $\underline{a}$, and for discerning points in $\mathcal{N}$, we wish to turn to the derivation of an efficient recursive algorithm for the solution of (8).

Sequential procedure using Givens' reduction. Consider rewriting (8) as

(20) $\underline{Q}(N)\underline{A}(N)\hat{\underline{a}}(N) = \underline{Q}(N)\underline{X}(N)$

which represents a system of N equations and P unknowns. (Note, however, that there are only $\sum_{i=1}^{N-1} q(n)$ equations not of the form $\underline{0}=\underline{0}$.) We now solve (20) employing Givens' reduction [8]. The significance of the Givens' transformation for this work, in addition to its computational efficiency, is that it can be performed row-wise, i.e., each equation can be considered independently of the others. A second advantage over the classical recursive least squares algorithm is that it provides the prediction residuals without requiring the back substitution for the coefficients, a_i, where ill-conditioning often occurs.

The Givens operations are performed within a (P+1)X(P+1) working array, $\underline{W}$ where equations corresponding to <u>selected</u> times, n, are entered on the last row. Two "Givens' rotations" [8] are then performed and the error signal at time n, e(n), is extracted from the (P+1,P+1) element of $\underline{W}$. The temporal selection process uses a "backup" array, $\underline{BW}$, which is updated only when the residual meets the criterion of (16), e.g. Fig. 1 shows the flowchart for this procedure. The recursion is initiated with $\underline{W}$ and $\underline{BW}$ full of zeros and the coefficients are obtained from the final back substitution at the end of the processing window.

The computational costs per sample using the modified Givens' method of Gentleman [9] is shown in Table 1. In this case 50% of the samples were rejected. The costs compared with those for a recursive least squares algorithm [4,10] demonstrate a clear advantage for the new method. The absence of ill-conditioning and the simplicity of the array processing also represent significant improvements.

SISSI. The new algorithm which combines the temporal selection procedure with the efficient simultaneous equation solution has been named SISSI: a <u>S</u>ilent <u>I</u>nput <u>S</u>elective <u>S</u>equential <u>I</u>dentifier for AR systems.

4. SPEECH PROCESSING EXAMPLE

We now consider an example application employing the new method for the estimation of vowel formants. A simulated speech signal was generated using a cascade form [11] implementation of a sixth order AR model with the pole positions corresponding to the formants of the vowel /i/ at a 10 kHz sample rate (Fig. 2). The simulator was driven by either a periodic discrete time impulse train or a periodic Rosenberg pulse [7] train:

(21) $u(n) = \{1 - \cos[(\pi)n/N1],\ 0 \leq n \leq N1$
$\{\cos[\pi(n-N1)/(2.N2)],\ N1 < n \leq N2$
$\{0,\ N2 < n \leq N3$

where N1 was be chosen to be 20 and N2, 50. In each case random noise of variance σ^2 was added to the input. The period of the waveforms N3, and σ were variable across experiments.

The experimental algorithm is diagrammed in Fig. 3. In this figure, "BEA" and "STL" refer to "batch error analysis", and, "short term LPC," methods for detection of the initiation set, $\mathcal{S}$, which are described in [5]. The parameter K in the figure is the length of the initiation set which was selected as 20 in order to provide a reasonable estimate of the noise variance.

The performance of SISSI, compared with the standard covariance method for various conditions of the input, can be evaluated from the data in Tables 2-4. The summary number in each entry in the tables is the quadratic bias,

$$(22)\quad B = \left| \sum_{i=1}^{P} (\hat{a}_i(N) - a_i)^2 \right|^{1/2} \quad \text{where } N=200 \text{ for all experiments}$$

where, of course, a sixth order "formant" identification is attempted in each case. Tables 2 and 3 are concerned with the pulse train case. Table 2 shows the influence of the period of the pulse train. Note that SISSI produced significantly better results for all tested periods. Even in the period = 100 case where the result from conventional methods would be expected to be minimal, the SISSI bias is roughly a factor of ten better. In Table 3, the period of the waveform has been fixed at 30 samples and the noise variance varied. The signal-to-noise ratio (SNR) is computed as:

$$(24)\quad SNR = \frac{\text{uncorrupted signal energy per period}}{\sigma^2}$$

Again SISSI outperforms the covariance approach for all values of σ.

Table 4 shows performances when u(n) is the Rosenburg pulse train described above with period 100. In this example one can see the serious bias effects which can be induced by unmodelled correlated inputs such as the glottal waveform when using conventional approaches. The SISSI method alleviates this problem.

Other examples and further discussion are given in [5].

5. CONCLUSIONS

An efficient sequential algorithm for the identification of AR systems excited by a certain class of correlated inputs has been developed. The method alleviates the solution bias inherent in conventional methods by discarding incoming data points for which the input is not "silent." Theoretical justification for this idea has been provided.

Two related problem dependent operational procedures are inherent in the technique: detection of silent input times and the detection of an appropriate temporal segment upon which to compute an initial coefficient vector estimate. A framework for solving these problems has been discussed.

Finally, an important contribution of this work is the use of the Givens' reduction in the solution which significantly improves the computational efficiency of the identification.

6. REFERENCES

[1] J.D. Markel and A.H. Gray, Linear Prediction of Speech, Springer-Verlag, New York, NY, 1976.

[2] J.R. Deller, Jr., On the identification of autoregressive systems excited by periodic signals of unknown phase, IEEE Trans. Acoust. Speech Signal Process. (in review)

[3] T. Kailath, Linear Systems, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[4] D. Graupe, Identification of Systems (2nd Edition), Krieger, Huntington, NY, 1976.

[5] J.P. Laebens, Unbiased Identification of AR Systems with a Class of Correlated Inputs (MS Dissertation), Illinois Institute of Technology, Chicago, 1982.

[6] D.Y. Wong, J.D. Markel, and A.H. Gray, Least squares glottal inverse filtering of speech, IEEE Trans. Acoust. Speech Signal Process., ASSP-27, pp. 350-355, 1979.

[7] A.E. Rosenberg, Effect of glottal pulse shape on the quality of natural vowels, J. Acoust. Soc. Amer., 49, pp. 583-590, 1971.

[8] J.C. Nash, Compact Numerical Methods for Computers. Linear Algebra. and Function Minimization, John Wiley and Sons, New York, NY, 1977.

[9] W.M. Gentleman, Least squares computations by Givens transformations without square roots, J. Inst. Math. Applics., 12, pp. 329-336, 1973.

[10] G.F. Franklin and J.D. Powell, Digital Control of Dynamic Systems, Addison-Wesley, Boston, MA, 1981.

[11] A.V. Oppenheim and R.W. Schafer, Digital Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1975.

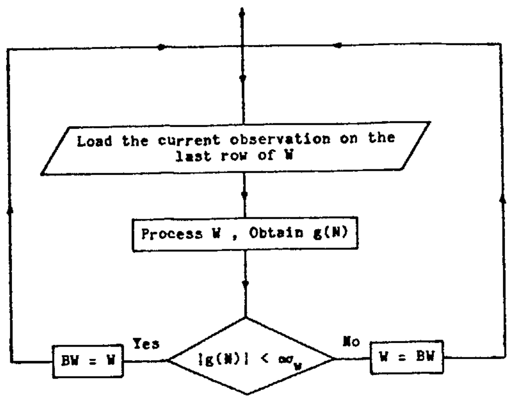

Figure 1. Selection Process with Givens' Reduction

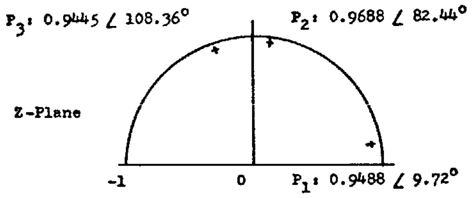

Figure 2. Experimental Poles Loci

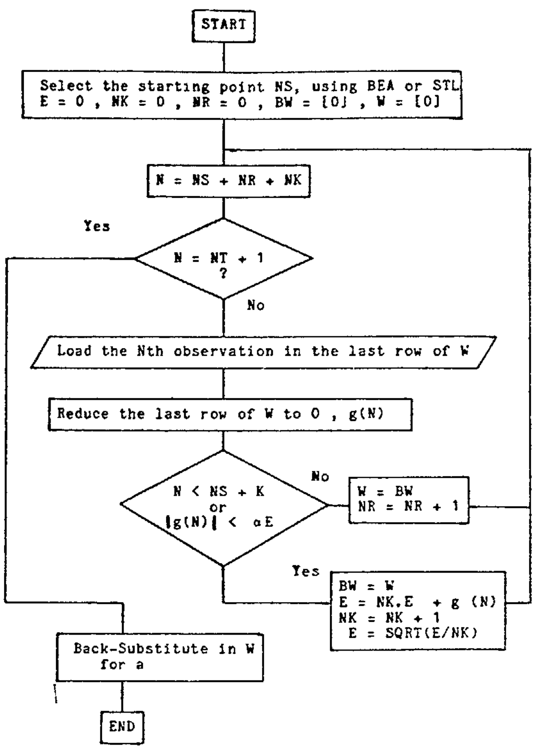

Figure 3. The Experimental Algorigthm

Table 1. Computational Costs

	Multiplications	Additions
Classical Recursive Least Squares	$2P^2 + 2P$	$2.5P^2 + P$
Givens' Reduction w/o Square Roots	$1.5P^2 + 6P$	4P

Table 2. u(n): Impulse Train: Influence of the Period

Period	200	100	70	30
Cov. Meth.	0.931 E-3	0.701 E-2	0.618 E-1	0.136 E-0
New Algor.	0.536 E-3	0.614 E-3	0.641 E-3	0.411 E-2

Table 3. u(n): Impulse Train: Influence of Input Noise

SNR (db)	200	40	20	10
Cov. Meth.	0.136 E-0	0.136 E-0	0.142 E-0	0.138 E-0
New Algor.	0.411 E-2	0.417 E-2	0.402 E-2	0.138 E-1

Table 4. u(n): Rosenberg Pulse Train: Influence of Input Noise

SNR (db)	200	40	20	10
Cov. Meth.	0.855 E+1	0.554 E+1	0.202 E+1	0.930 E-0
New Algor.	0.590 E-3	0.147 E-1	0.173 E-0	0.250 E-0