OPTIMAL WEIGHTED LS AR ESTIMATION IN PRESENCE OF IMPULSIVE NOISE

E.D. Di Claudio**, G. Orlandi**, F. Piazza, A. Uncini

Dept. of Electronics and Automatics, University of Ancona, Ancona, Italy
(*) Telettra SpA, Chieti Scalo, Italy
(**) Dept. INFOCOM, University - "La Sapienza", Rome, Italy

ABSTRACT

The parameters of an AR model for spectrum estimation are often obtained by the Least Squares (LS) solution of an overdetermined set of linear prediction equations. However, when the signal is contaminated by few strong noise spikes, the standard LS algorithm can easily lead to biased solutions characterized by a strongly reduced dynamic range of the estimated spectra. In order to face this problem, the classical approach is to weight the prediction errors before applying the LS minimization algorithm.

In this paper a new procedure for assigning optimal weights to the prediction equations is presented. The set of weights is computed, by linear programming techniques, in order to reduce the effects of strong impulsive noise onto the AR parameter estimate. The proposed method is particularly effective when the gaussian white noise component is much smaller than both spikes and useful signal. In order to demonstrate the capability of the proposed approach, the results of a simple AR parameter estimation experiment is also reported.

1. Introduction

Least Square (LS) estimation algorithms are well known techniques for the computation of AR model parameters by linear prediction methods, for the adaptive filtering of signals, eventually multidimensional, and for the solution of other related problems (interpolation, deconvolution, etc.).

In these cases, an overdetermined set of linear equations is built. In matrix form, we get, for a system of m equations and n<m unknowns:

\[ A \hat{\chi} = b + e \]  

where A and b contain known signal samples. The elements of the vector e are the equation errors associated to each choice of the unknown vector \( \hat{\chi} \).

The Least Squares solution \( \hat{\chi} \) of the generally inconsistent system (1) is obtained by minimizing the norm of e with respect to the elements of vector \( \hat{\chi} \).

If the rank of the matrix A is full (i.e. equal to n), such solution is unique and can be written as [1]:

\[ \hat{\chi} = (A^* A)^{-1} A^* b \]  

where \( * \) denotes the hermitian conjugation operator. The (n,m) rectangular matrix \( (A^* A)^{-1} A^* \) is said to be the pseudoinverse of A and it will be indicated in the following as \( A^* \). For this choice of \( \hat{\chi} \) the error vector e is found to be orthogonal to the column space of matrix A [2].

If the rank of A is not full (i.e. less than n), then the data matrix \( (A^* A) \) is singular and the minimum of the norm of e is no longer attained for a unique solution vector \( \hat{\chi} \). In fact, if g is a generic vector belonging to the right nullspace of A, i.e.:

\[ A g = 0 \]  

it results, by combining expressions (1) and (3) [1]:

\[ A (\hat{\chi} + g) = b + e(\hat{\chi}) \]  

In harmonic retrieval and direction finding problems with noiseless signals, the matrix A typically has a rank lower than n, as a consequence of the data model and of the overestimation of its true order. A minimum norm solution \( \hat{\chi} \) is often of interest, and is obtained by means of the Moore-Penrose inverse [3], easily computed through an SVD decomposition of A. It can be shown that in all cases only the projection of b onto the column space (often referred to as "signal subspace") of A is of interest for the LS solution.

As pointed out by Cadzow [4] and by Tufts and Kumaresan [3], the presence of noise can worsen the obtainable results. Noise increases the actual rank of the matrix A and perturbs the estimate of the signal subspace basis vectors.

In particular the presence of a small number of short-time strong pulses, can induce large projections of b onto the subspace of A created by noise (often referred to as "noise subspace"). It will increase the difficulty of LS parameter estimation by reducing the average difference between the signal and noise related eigenvalues of the data matrix \( A^* A \).

In this case, if the LS estimate of the AR model parameters by linear prediction is involved, flattened spectra are usually obtained. In fact, since an AR model assumes the stationarity of the process during the observation time, noise pulses of very small duration can easily affect the results of the LS algorithm over the entire data segment. Another way to explain this behaviour is to observe that the norm of e is actually dominated by errors in equations involving severely corrupted data samples.

The traditional way to cope with the problem of the presence of noise, is that of setting an higher prediction order and decomposing data or correlation matrices by SVD or EVD techniques [3,4].

This work was supported in part by the Consiglio Nazionale delle Ricerche of Italy under the project "Materiali e Dispositivi per l'Eletronica allo Stato Solido" and in part by the Ministero dell'Universita e della Ricerca Scientifica e Tecnologica of Italy.
However, when the noise is impulsive, these methods cannot obtain satisfactory results since a large number of prediction equations is severely perturbed even by a single highly corrupted sample and therefore the whole A^T A matrix results biased.

In order to face this problem, the corresponding equations are weighted according to a suitable criterion of noise identification and suppression. Two approaches have been historically introduced:

1) Iteratively Reweighted L.S. In this approach the LS algorithm is first applied to the data. A suitable influence function [2,5] is applied to equation errors in order to obtain a set of weights. These weights are used in a further LS minimization stage and the whole procedure is repeated until the results stabilize. This algorithm is often computationally expensive and the convergence check is still left to a subjective judgement.

2) A Priori Weighted L.S. The prediction errors are weighted before applying the minimization procedure. These weights are chosen in order to reduce the influence of the equations corrupted by the impulsive noise. This requests an "a-priori" knowledge of signal and noise statistics.

Following the second approach, in this paper a new procedure for assigning optimal weights to the prediction equations is presented. The set of weights is computed by linear programming techniques, in order to reduce the effects of strong impulsive noise onto the AR parameter estimate.

2. The Proposed Algorithm

For linear prediction problems, we can build the classical forward-backward equation matrix with N available signal samples \{x_i, i = 1, ..., N\} and p unknown regression parameters \{a_j, i = 1, ..., p\} in the following way:

\[
x_n = \sum_{k=1}^{p} a_k x_{n-k} \quad n = (p+1), ..., N
\] (forward)

\[
x^*_n = \sum_{k=1}^{p} a_k x^*_{n+k} \quad n = 1, ..., (N-p)
\] (backward)

or, in matrix notation:

\[
X a = x
\] (m x p) \( (p x 1) \) (m x 1)

with obvious meaning of symbols. Then 2(N - p) = m > p equations are obtained, resulting in an overdetermined linear system in the unknown \(a\), similar to (1).

If a diagonal \((m x m)\) weight matrix \(W\) is introduced, with nonnegative diagonal elements \(w_i\) \((i=1, ..., m)\), the overdetermined system of equations becomes:

\[
(WX) a = -(WX) \quad W = \text{diag}[w_i \geq 0, \quad i = 1, ..., m]
\]

The proposed algorithm allows to compute the matrix \(W\) in order to reduce the influence of impulsive noise onto the estimate of \(a\). The algorithm is based on the following issues:

1) Better AR parameters are often obtained by weighting each equation proportionally to the square root of the energy of samples contained in the prediction window [6].

2) It is well known that the use of the L1 norm let better cope with the presence of spikes [7].

3) Predictable (correlated) signals show generally smaller envelope fluctuations than wide-band noise pulses.

In order to find the optimal weights matrix \(W\), the following quantities are computed:

\[
E_i = \sqrt{\frac{1}{n} \sum_{j=1}^{n} |X_{ij}|^2} \quad P_i = |x_i|
\]

where \(i=1, ..., m\), \(X_{ij}\) is the \((i-th, j-th)\) element of \(X\) and \(x_i\) is the \(i-th\) element of \(x\). These values represent the square root of the average signal energy in each prediction window. Their computation can be simplified using the symmetry relationship reported in [6].

A strong difference between the \(E_i\) and \(P_i\) values indicates a possible large prediction error due to a sudden change of the signal statistics (for example the presence of a noise impulse). Such an error can also impair the AR estimate by moving the poles outside the unit circle.

The scope of the algorithm is to compute \(w_i\) weights which minimize the influence of the prediction equations that correspond to windows with small energy or/and with a sudden change in the signal statistics. The problem can be formalized as a linear programming one by choosing a cost function as:

\[
C = \sum_{i=1}^{m} w_i |E_i - P_i|
\]

and set of suitable linear constraints:

\[
w_i \leq E_i \quad i = 1, ..., m
\]

\[
\sum_{i=1}^{m} w_i \geq \alpha \sum_{i=1}^{m} E_i \quad 0 < \alpha < 1
\]

The first group of inequalities imposes a set of constraints similar to that in [6]. It minimizes the influence of equations relating consecutive samples with a small average energy. It is expected that these signal segments have an higher stationary noise content.

The last equation simply avoid the convergence to the trivial solution with all zero weights.

The parameter \(\alpha\) is selected through the following simple formula:

\[
\alpha = \frac{\sum_{i=1}^{m} E_i}{\sum_{i=1}^{m} E_i}
\]

where \(\text{Int}()\) indicates the floor operator and the \(E_i\) values are ordered in a descending manner.

If the observed process is wide sense stationary, the highest \(E_i\) values should indicate spike-contaminate prediction
windows, while the remaining ones should exhibit a small variance with respect to the mean value of signal energy. Therefore, a good estimate for the $\alpha$ parameter should be a number proportional to the ratio between the average of the energy of uncontaminated windows and the total average energy. More sophisticated formulas can be used in the case of transient signal estimation, where both raising and falling edges of energy must be excluded from the averages used in (13). Typically the values for the $K$ parameter can be chosen around 0.5.

The optimal weights $w_i$ ($i=1, \ldots, m$) are then obtained by a standard linear programming technique, for example with the simplex method. The final weighted LS solution $\hat{\theta}$ for the prediction vector is computed through the pseudoinverse of the matrix $(WX)$, resulting in:

$$\hat{\theta} = - (WX)^\dagger (WX) x.$$ \hspace{1cm} (14)

3. Some Experimental Results

In order to demonstrate the capability of the proposed method, the results of a simple AR parameter estimation experiment is reported. The signal is composed by 40 samples of two complex exponentials:

$$x_n = e^{j(2\pi f_1 n)} + e^{j(2\pi f_2 n + \pi/4)}$$ \hspace{1cm} (15)

with $n=1, \ldots, 40$, $f_1=0.15$ and $f_2=0.2$; corrupted by adding 3 spikes $S_k$, $k=1,2,3$

$S_1 = 10 + j$ \hspace{1cm} for $n = 5$

$S_2 = -10 + j$ \hspace{1cm} for $n = 9$

$S_3 = 10 - j$ \hspace{1cm} for $n = 22$

Two linear prediction vectors, respectively of order $p=2$ and $p=4$, have been computed by means of either the standard Forward Backward Least Squares Algorithm (FBLS) [8], or by the proposed weighted LS approach. In both cases, the pseudoinverse of the data matrix was computed by the SVD algorithm reported in [9]. For uncontaminated data the rank of $X$ remains fixed at two for any greater prediction order [3]. In our case, however, the effective rank of the matrix $WX$ is determined by the spread between the largest singular values, (belonging to the signal subspace) and the smallest ones (introduced by the noise).

After adding the three spikes, the FBLS algorithm is unable to detect and resolve the two complex exponentials for a selected order $p=4$. The singular values (Table 1) show no clear distinction between signal and noise subspaces. The resulting AR spectrum is therefore severely perturbed (see Fig. 1). Several other trials with higher prediction orders do not exhibit any improvement in detectability of signals, because several spurious peaks appear in the spectra which mask those generated by the noisless signal.

Using the proposed method, the weight matrix $W$ has been computed after selecting a value of $K=0.5$ (this choice is shown to be not critical) and prediction orders $p=4$ and $p=2$. As shown in Fig. 2, in both cases the algorithm assigns very small or zero weights to the equations contaminated by the impulsive noise. The resulting AR spectra are almost ideal, as shown in Fig. 3 and 4 respectively for order 2 and order 4. The corresponding singular values are reported in Table 1.

In the case of $p=4$, two singular values of the matrix $WX$ are very small, indicating that the impulsive noise has been perfectly suppressed and the ideal rank of the noiseless matrix has been restored. The AR spectrum was computed after suppressing these small singular values and projecting the vector $x$ onto the estimated signal subspace, spanned by the two dominant singular vectors [3].

4. Limits on applicability

This approach is more effective when the gaussian white noise component is much smaller than both spikes and useful signal. If this hypothesis is satisfied, the algorithm will work properly, since in most cases the number of severely underweighted equation is small and it performs actually as the basic Covariance Least Squares algorithm of [6].

With the suppression of noise pulses and localized non stationarities, the effective rank of the matrix $(WX)$ is often reduced. This fact allows to reduce also the order of the AR model, thus increasing the number of "useful" equations since a lower number of prediction windows are now contaminated by spikes. However, although an optimized simplex algorithm exploiting the quasi-banded nature of the involved matrix is used, the computational cost remains several times higher than that required by the classical LS minimization.

Experimental results indicate that the LS matrix $(WX)^*$ $(WX)$ can be nearly singular for the selected order $p$, requiring the use of truncated pseudo-inverses through SVD or EVD decompositions.

More advantages are expected by applying the proposed weighting approach to data matrices used in multichannel filtering or in direction finding problems. In fact, spatially correlated pulsed noise can distort the measured correlation of sources of interest, causing sudden changes of the average energy received by the sensors at the same instant [10]. These effects can be revealed and corrected by a slight modification of the proposed approach.

REFERENCES


<table>
<thead>
<tr>
<th>AR coeff.</th>
<th>noiseless signal system order = 2</th>
<th>standard FBLS system order = 2</th>
<th>proposed method system order = 2</th>
<th>noiseless signal system order = 4</th>
<th>standard FBLS system order = 4</th>
<th>proposed method system order = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.8968 -j1.7601</td>
<td>0.0584 -j0.0915</td>
<td>-0.8968 -j1.7600</td>
<td>-0.8395 -j0.8625</td>
<td>-0.02493 -j0.0841</td>
<td>-0.4395 -j0.8626</td>
</tr>
<tr>
<td></td>
<td>-0.5878+j0.8090</td>
<td>0.0438 -j0.0959</td>
<td>-0.5878+j0.8090</td>
<td>0.2851 -j0.3923</td>
<td>0.0561 -j0.0629</td>
<td>0.2850 -j0.3923</td>
</tr>
<tr>
<td></td>
<td>17.1175</td>
<td>30.9573</td>
<td>23.0328</td>
<td>22.9061</td>
<td>33.3782</td>
<td>33.5513</td>
</tr>
<tr>
<td>Singular</td>
<td>1.3885</td>
<td>27.4027</td>
<td>0.3921</td>
<td>4.3040</td>
<td>27.5275</td>
<td>2.0704</td>
</tr>
<tr>
<td>Values</td>
<td></td>
<td></td>
<td></td>
<td>2.85e-6</td>
<td>27.3861</td>
<td>3.51e-6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.54e-6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. AR prediction coefficients (first block of rows) and singular values (second block of rows), computed setting a prediction order of 2 (first 3 columns) and of 4 (last 3 columns) by using respectively the noiseless signal, the signal plus impulsive noise with the standard FBLS algorithm and the signal plus impulsive noise with the proposed algorithm.

Fig. 1. Spectrum obtained using an AR model of order 4, estimated by the standard FBLS algorithm applied to the test signal.

Fig. 2. Weight vector obtained by the proposed algorithm applied to the test signal, in the case of order equal to 4 and K=0.5.

Fig. 3. Spectrum obtained using an AR model of order 2, estimated by the proposed algorithm applied to the test signal.

Fig. 4. Spectrum obtained using an AR model of order 4, estimated by the proposed algorithm applied to the test signal.