Proceedings of the Institutue of Acoustics

CONSTANT VARIANCE TRANSVERSAL FILTERING FOR ADAPTIVE CHANNEL EQUALIZATION

G. Vázquez, A. Gasull, J.J. Sánchez, M.A. Lagunas

Departament de Teoria del Senyal i Comunicacions
E.T.S.E. Telecomunicació de Barcelona (U.P.C.)
Jordi Girona Salgado, s/n, 08034 Barcelona, SPAIN

1.- INTRODUCTION

This paper addresses the problem of channel equalization in communications and channel identification in the context of sonar. In both cases, the study comes from the optimal linear filtering theory, taking out the useful information contained in the received data, that used to be noisy. In this sense, the equalization of a channel in analog or digital communications consists of the characterization of the channel behaviour in such a way that it is possible to correct its frequency response and besides, to reduce the degradating effects of the noise that we are going to consider additive, but not necesarily gaussian. The objective in the problem of channel identification is clearly different. In this case we search for a model of the channel and the desired information remains just in the coefficients of the model. What is really important is that the two refered topics can not be acomplished by the same techniques. For instance, some kind of gradient-based adaptive algorithms performs much better for equalizing than for identifying the channel, that is, the filtering error exhibits a different convergence-time than the filter weights.

As it is well known, the common strategy for the design of adaptive systems is the minimization of a mean square error function. The gradient-based systems used to be generated from sthocastic cost functions, and the recursive solutions, usually related to the Kalman filter, as the consecuencie of minimizing deterministic expressions of a mean square error. This paper deals with the second case, generating a solution that is obtained under a maximum likelihood criteria, using the complete set of data that is supplied to the filter. The paper shows that a weighted mean square error cost objective deals and thus, it is necessary to weight the filtering error sequence by a factor that, basically, depends on the probability density function of the error sequence and on its first derivate. As it is well known, this kind of information used to be not available and other proposals must be made.

Some kind of data-dependent weighting functions have been refered in the literature, and all them in the context of linear prediction [1] or parametric spectral analysis [8], but never in the context of optimal linear filtering. The paper discusses the design of this weighting factor for including some kind of data-selection mechanism for the final filter weight-vector solution design.

The underlying of the proposal is the search of algorithms achieving better robustness in front of non-gaussianity in the data we are handling, as it happens in a number of actual situations. The main trouble of our proposal is that the M.L. objetive leads to a non-linear problem, but under some specific conditions it is possible to give a recursive and linear estructure to the solution, but as a matter of fact it becomes and open research line with now possible guesses.

The main contribution of this paper is an attractive new Newton adaptive algorithm with a clear improvement in the behaviour for narrow band conditions.

This work was supported by the PRONTIC number 105/88

2.- OBJECTIVE DEFINITION

As it is describes in figure I and figure II given a finite impulse response N-order digital filter with coefficients $\{w_i\}_{1\le i\le N}$, we try to estimate the samples of a known reference sequence (or training sequence) $\{d(k)\}_{0\le k\le n}$ as a linear combination of the received data samples $\{x(k)\}_{0\le k\le n}$. For each reference samples d(k), we commit an error $e_n(k)$ given by the expression :

$$e_n(k) = d(k) - \underline{W}^H(n)\,\underline{X}(k) \qquad (1)$$

that is, the diference between the desired sample and the estimate obtained by the inner product between the coefficient or weight filter vector $\underline{W}(n)$ and the data vector the filter has received $\underline{X}(k)$.

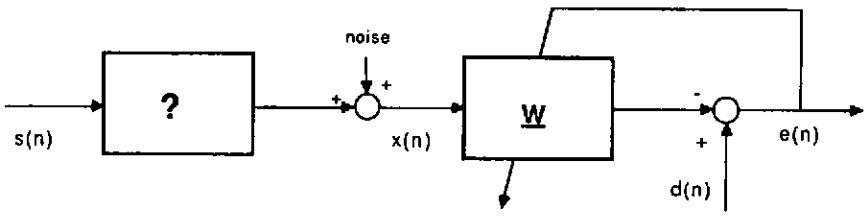

Figure I. Diagram for channel equalization.

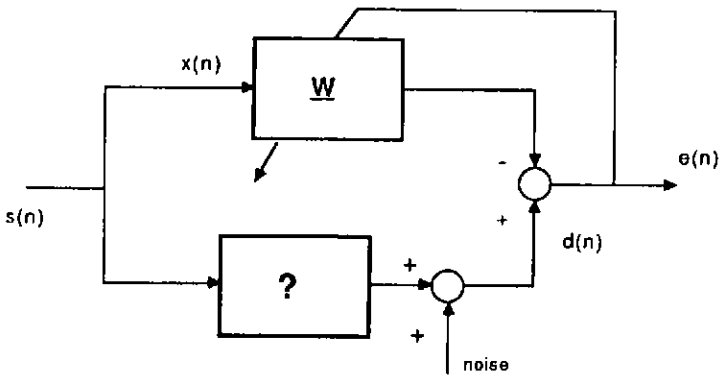

Figure II. Diagrama for channel identification

The vector solution $\underline{W}(n)$ will be designed in such a way that the filter behaviour will be optimal under a given criteria. Let's consider a meassure of the likelihood degree of the error sequence given a filter coefficients vector $\underline{W}(n)$ and conditioned to the received data. Thus, we will use the following conditioned joint probability density function :

$$\Phi(e_n(.);\underline{W}(n)/X(.)) = p(e_n(0),e_n(1), \ldots, e_n(n)/x(0), x(1), \ldots, x(n)) \quad (2)$$

the design equation for the filter coefficient vector under a M.L. criteria becomes :

$$W(n) \qquad \max_{\underline{W}(n)} \quad \Phi(e_n(.);\underline{W}(n)/X(.)) \quad (3)$$

If the error sequence $\{e_n(k)\}$ is distributed following the probability density function $f(e;\underline{W}(n))$, the cost function (3) leads to the conditional log likelihood function given by :

$$\underline{W}(n) \qquad \max_{\underline{W}(n)} \sum_{k=0}^{n} \ln f(e_n(k);\underline{W}(n)) \quad (4)$$

where the error term $e_n(k)$ is defined in (1)

It is easy to prove that (4) is equivalent to the following weighted least square problem :

$$\underline{W}(n) \qquad \min_{\underline{W}(n)} \sum_{k=0}^{n} |e_n(k)|^2 \Gamma_n (e_n(k)) \quad (5)$$

where each error term becomes weighted by the factor $\Gamma_n(.)$ that depends on the distribution of the error and on its first derivate :

$$\Gamma_n(e) = - \frac{f'(e;\underline{W}(n))}{e\ f(e;\underline{W}(n))} \quad (6)$$

Unfortunately, it is difficult to known the statistic behaviour of the error, either of the data we are handling. It is very common to try a characterization from second order statistics and in most of the cases it is not sufficient. Thus, not to know this lateral information makes the computation of the weighting factor (6) impossible. As we will see, for most of the usual distributions this factor tries to penalize the higher energy terms in the expression (5) considering them unlikely ones.

3.- STATISTIC KNOWLEDGE. DESIGN OF THE WEIGHTING FACTOR

From equations (5) and (6), the evaluation of the weighting factor implies the resolution of a transcendental equation. It is possible to see that for a number of non-gaussian symmetric density functions decreasing monotonically with |e|, that the weighting factor is also a monotonically decreasing function with |e|, symmetric and positive. As a consecuence, the effect of the weighting factor is to penalize the higher order terms in the accumulated square error (5), keeping unaffected the error samles with smaller energy.

Kay [1] proposes a non linear function of the error for the design of the weighting factor (6), in the form of a "time domain" Butterworth filter, that is :

$$\Gamma(e) = \frac{a}{1 + |\frac{e}{e_c}|^{\beta}} + b \qquad (7)$$

where parameters $\{a,b,e_c,\beta\}$ are determined for each statistic distribution to approach.

The method consists in a two step one, where first, the weighting factor is assumed to be constant supplying an initial filter vector solution. Then the "a posteriori" error samples are computated and the least square error problem is re-initialized with each error term weighted by expression (7).

The main handicaps for this proposal are that the computation cost increase, the determination of parameters $\{a,b,e_c,\beta\}$ is not clear enough and finally, that it is not possible to give a recursive diagram to the algorithm.

The underlying of our proposal is that in spite of the non-gaussianity for the involved data, the error (1) tends to gaussianity when the filter order increasses. Under this condition it is possible to evaluate the weighting factor (6) :

$$\Gamma_n(e(k)) = \frac{1}{\sigma_n^2(k)} \qquad (8)$$

Where $\sigma_n^2(k)$ is the variance of the error sample $e_n(k)$:

$$\sigma_n^2(k) = E(|e_n(k)|^2) \qquad (9)$$

The consecuence of substituting (8) in (5) is that the square error terms are normalized by the estimate of the error sample variance, leading to constant variance error samples. If the measure of the variance if high, the sample is removed in the objetive function (5) and it is classified as an unlikely one. The problem now is how to estimate the variance (9) of each error sample.

4.- CONSTANT VARIANCE CRITERIA

Let's proceed as in the Kalman filter theory. If we denote the ideal solution by $\underline{W}^*$, the variance of an error sample $e_n(k)$ is given by :

$$\sigma_n^2(k) = E(|e_n(k)|^2) = \xi^2_{min} + \underline{X}^H(k)\,\underline{\underline{K}}(n-1)\underline{X}(k) \qquad (10)$$

where $\underline{\underline{K}}(.)$ is the coefficients vector covariance matrix :

$$\underline{\underline{K}}(n) = E((\underline{W}(n) - \underline{W}^*)(\underline{W}(n) - \underline{W}^*)^H) \qquad (11)$$

and ξ^2_{min} is the mean power of the error for the ideal solution :

$$\xi^2_{min} = E\,(|e_n(k)|^2) \Big|_{\underline{W}(n) = \underline{W}^*} \qquad (12)$$

Thus, the maximum likelihood criteria leads to the following least square problem :

$$\underline{W}(n) \qquad \min_{\underline{W}(n)} \sum_{k=0}^{n} \frac{|e_n(k)|^2}{\sigma_n{}^2(k)} \qquad (13)$$

where $\sigma_n{}^2(k)$ is given by (10). The physical meaning of the constant variance criteria is clear. Recovering the definition of the error (1), we have that given a known data vector $\underline{X}(k)$ and a known sample of the reference sequence d(k), the error term is only a function of the coefficient vector $\underline{W}(n)$.

Besides, vector $\underline{W}(n)$ is obtained in the minimization of the deterministic objective (13), where the complete set of data samples and reference samples are considered. The question is the relative to the degree of dependence or the sensibility exhibed by the error (1) to any change in the coefficient vector $\underline{W}(n)$. If one data vector $\underline{X}(k)$ is degradated, it shows a different structure than the rest of the data, and an increase in the variance of its associated error is expected. As a consequence, the contribution of the error term due to this data is removed in the cost function. If this idea is true, the error variance estimate will depent on the data vector we are using in the evaluation of the error and also of the rest of the data and the relation between them. In this way, the statistic behaviour of the complete data set must be reflected in the expression of the variance.

Until now, all the development has been rigorous and exact. The last aspect to define is the evaluation of the coefficient vector covariance matrix of the filter [11].

Due to the fact that matrix (11) is positive definite, the estimate of the error variance (7) appears, except a constant factor, as an inner product of the considered data vector $\underline{X}(k)$, in the metric of the coefficient vector covariance matrix (11).

"Data selection" will take this matrix (11) as a reference in such a way that we hope that it will include the information about the statistic behaviour of all the previous received data vectors, that is, all them except the data vector we are selecting X(k).

It is very common to consider the error variance as a constant one to avoid the computation of matrix (11). This is equivalent to assume that the coefficient vector covariance matriz is a null one, and this is absurd. It is evident that the succesive updating of the coefficients improves the quality of the solution we obtain and this making better of

the equalizer must be reflected in the coefficient vector covariance matrix. Therefore, the inclussion of the data selection gets a clear meaning.

5.- WEIGHT-VECTOR COVARIANCE MATRIX ESTIMATION

The coefficient vector covariance matrix is obtained in the resolution of a transcendental equation and then without a high computation cost. Our proposal begins considering that the searched solution respons to a Newton diagram, as in Kalman filtering, that is :

$$\underline{W}(n+1) = \underline{W}(n) + \mu\, \underline{\underline{R}}^{-1}(n)\, e^{*}(n)\, \underline{X}(n) \qquad (14)$$

where 'μ' is the 'step-size' and $\underline{\underline{R}}(n)$ is an ergodic estimate of the data autocorrelation matrix :

$$\underline{\underline{R}}(n) = \sum_{k=0}^{n} \underline{X}(k)\, \underline{X}^{H}(k) \qquad (15)$$

In this case, it is easy to obtained an expression for the covariance matrix (11), that is:

$$\underline{\underline{K}}(n) = \xi^{2}_{min}\, [1 + \underline{X}^{H}(k)\, \underline{\underline{R}}^{-1}(n-1)\underline{X}(k)] \qquad (16)$$

Thus, the variance for an error sample (10) becomes to be :

$$\sigma^{2}_{n}(k) = \xi^{2}_{min}\, [1 + \underline{X}^{H}(k)\, \underline{\underline{R}}^{-1}(n-1)\underline{X}(k)] \qquad (17)$$

An the final cost function is given by

$$\underline{W}(n) \qquad \min_{\underline{W}(n)} \sum_{k=0}^{n} \frac{|e_{n}(k)|^{2}}{1+\underline{X}^{H}(k)\, \underline{\underline{R}}^{-1}(n-1)\underline{X}(k)} \qquad (18)$$

For the moment, it is not known a recursive solution for (18) and a block analysis is required. In spite of the non-recursivity of the solution, matrix $\underline{\underline{R}}^{-1}(n-1)$ is known for the 'n' instant due to the fact that it has been computated for obtaining $\underline{W}(n-1)$ in the previous update, and the global computation is not so intensive.

Nevertheless, our proposal includes another alternative, much more simple and efficient. Approaching expression (17) by :

$$\sigma_{n}^{2}(k) = \xi^{2}_{min}\, (1 + \underline{X}^{H}(k)\, \underline{\underline{R}}^{-1}(k-1)\, \underline{X}(k)) \qquad (19)$$

the objetive can be writen in a recursive way and it presents a recursive resolution :

$$\underline{W}(n) \qquad \min_{\underline{W}(n)} \sum_{k=0}^{n} \frac{|e_n(k)|^2}{1+\underline{X}^H(k)\underline{\underline{R}}^{-1}(k-1)\underline{X}(k)} \qquad (20)$$

In this approach the covariance matrix (11) is sample by sample updated and it tries to emphasises the more recent received data in front of the older ones, which is logical in an adaptive system because the quality of the solution improves in the succesive updating.

6.- FINAL ALGORITHM AND LAST COMMENTS

The objective (20) can be solved as an exact least square one. Minimizing with respect to the coefficients and ordering the terms of the expression, the final updating equation is given by :

$$\underline{W}(n+1) = \underline{W}(n) + \frac{1}{\eta^2(n) + \underline{X}^H(n)\,\underline{\underline{R}}^{-1}(n-1)\underline{X}(n)}\;\underline{\underline{R}}^{-1}(n-1)\;e^*_n(n)\;\underline{X}(n) \qquad (21.a)$$

with:

$$\eta^2(n) = 1 + \underline{X}^H(n)\,\underline{\underline{R}}^{-1}(n-1)\;\underline{X}(n) \qquad (21.b)$$

where $\underline{\underline{R}}(.)$ is evaluated by :

$$\underline{\underline{R}}(n) = \sum_{k=0}^{n} \frac{1}{\eta^2(k)}\;\underline{X}(k)\;\underline{X}^H(k) \qquad (21.c)$$

which inverse matrix is updated by the matrix inversion-lemma, that is :

$$\underline{\underline{R}}^{-1}(n) = \underline{\underline{R}}^{-1}(n-1) - \frac{\underline{\underline{R}}^{-1}(n-1)\underline{X}(n)\underline{X}^H(n)\,\underline{\underline{R}}^{-1}(n-1)}{\eta^2(n) + \underline{X}^H(n)\,\underline{\underline{R}}^{-1}(n-1)\underline{X}(n)} \qquad (21.d)$$

the analysis of the final solution proofs all the initial considered hypothesis, which are the following :

(1)- From (11), (16) and (21.c) the coefficient vector covariance matrix only depends on the previous received data vector and it doesn't include the data vector under selection.

(2)- The final coefficient-vector updating equation (21.a) is a Newton diagram.

(3)- Parameter $\eta^2(n)$ in (21.b) is a data selector into the updating equation, increasing the speed of the adaptive algorithm for data vectors which are parallel to the signal eigenvectors of the data autocorrelation matrix (21.c) and 'bracking' the update when the supplied data vectors are parallel to the noise subspace.

7.- REFERENCES

[1] S.Kay & D. Sengupta
Statistically computationally efficient estimation of non gaussian autoregresive process
Procc. of ICCASP-87, pp. 45-48, Dallas, U.S.A. 1987

[2] B.D.O. Anderson & J.B. Moore
Optimal Filtering
Prentice-Hall, 1979

[3] G. Vázquez
Power Dunctions and Data Transforms in Adaptive Filtering
Ph. D. dissertation, Politechnique University of Catalonia, March 1988 (Spanish)

[4] D. Goddard
Channel Equalization Using a Kalman Filter for Fast Data Transmission
IBM J. Res. Development, May 1974

[5] D. Widrow & S.D. Stearns
Adaptive Signal Processing
Prentice-Hall, 1985

[6] S. Haykin
Adaptive Filter Theory
Prentice-Hall, 1986

[7] S.U.H. Qureshi
Adaptive Equalization
Proc. IEEE, vol. 73, No. 9, Sept. 1985

[8] B. Toplis & S. Pasupathy
Tracking Improvements in Fast RLS Algorithms Using a Variable Forgetting Factor
IEEE Trans.ASSP, Vol. 36, No. 2, pp. 206-227, FEb. 1988

[9] B.I. Helme & C.L. Nikias
Improved Spectrum Performance Via a Data-Adaptive Weighted Burg Technique
IEEE Trans. ASSP, Vol. ASSP-33, No.4, pp.903-910, Aug. 1985