SIGNAL PROCESSING III: Theories and Applications
I.T. Young et al. (editors)
Elsevier Science Publishers B.V. (North-Holland)

KNOWN INPUT POWER SPECTRUM IN ADAPTIVE L.M.S. AND A.G. ALGORITHMS

Gregori Vázquez, Antoni Gasull and Miguel A. Lagunas.

E.T.S. Ingenieros de Telecomunicación. U.P.C.
C/ Jordi Girona Salgado, s/n. 08034 Barcelona - Spain.

ABSTRACT. This work deals with the use of previous or colateral information to improve the behaviour of adaptive algorithms. The study is made on gradient-based methods due to the relatively simple and good performances that they use to exhibit.

This paper shows that the complete knowledge of the data at the input of the adaptive filter (and in consequence of its autocorrelation matrix and its inverse) can be used to modify the classic L.M.S. algorithm leading to new expressions for the gradient and for the optimum 'step size', alternative, in some cases, to the Powell expression.

Finally, the description is completed with the comparison between the variation ranges and VLSI implementation cost for this two optimum 'step size' values and a natural generalization set of parameter is obtained.

1. INTRODUCTION

For the sake of simplicity, let us focuss the classic problem of Wiener filtering. Two possible alternatives can be adopted. The first one is the direct use of the optimum Wiener equation and the other is that an adaptive approach could be better under actual situations, where finite arithmetics and no-stationary conditions are used to be imposed.

If we adopt the second possibility, the question is how to use all the previous or colateral information available in a given adaptive algorithm.

From our point of view, there would be two possible choices to reflect these additional information in an adaptive squeme with a quadratic objective. They are the following:

a) To use it to estimate better the parameters or associated functions involved in the adaptive algorithm.

b) To include the colateral information as constrains or just in the minimization process.

This work is driven in both senses. The first one is the most obvious and will be used only to improve the gradient estimate.

On the other hand, the second one is not so direct as the previous, and, in general, it will try to modify the whole structure into the adaptive scheme to satisfy the constrains or the pursued error minimization criterion.

Thus, althrough it seems an atractive possibility, the designer will have to pay attention because, as a matter of fact, often the structure obtained will need a very intensive computation. In our case, only an optimum value for the step size will be searched, keeping the usual adaptive scheme.

2. REVIEW OF THE MINIMUM M.S.E. LINEAR FILTERING: [1], [2], [3], [4].

Let's consider the general scheme given in the figure 1. The objective is to minimize the mean square error (m.s.e.) between a reference signal $y(n)$ and an estimate of this signal at the output of a Q order F.I.R. filter defined by the coefficient vector $\underline{W}$. Thus, given an input data signal $x(n)$, we dispose of a data vector $\underline{X}_n$ and the desired estimation:

$$\hat{y}(n) = \underline{X}_n^T.\underline{W} = \underline{W}^T.\underline{X}_n \qquad (2.1)$$

where:

$$\underline{X}_n^T = (x(n),x(n-1),\ldots,x(n-Q+1)$$

$$\underline{W}^T = (w(0),w(1),\ldots,w(Q-1))$$

The weight vector will be chosen so that the M.S.E. is minimum:

$$\varepsilon^2 = E((y(n)-\hat{y}(n))^2)=E((y(n)-\underline{X}_n^T.\underline{W})^2) \qquad (2.2)$$

and developing the expression:

$$\varepsilon^2 = E(y^2(n))-E(y(n)\underline{X}_n^T)\underline{W} - \\ - \underline{W}^T E(y(n)\underline{X}_n)+\underline{W}^T R_{xx}\underline{W} \qquad (2.3)$$

The minimization of this expression leads to the well known optimal Wiener solution:

$$\underline{W}_{opt} = R_{xx}^{-1}.\underline{P} \qquad (2.4)$$

Where $R_{xx}=E(\underline{X}_n.\underline{X}_n^T)$ is the QxQ autocorrelation matrix of the data sequence x(n) and $\underline{P}=E(y(n).\underline{X}_n)$ is a cross-correlation vector between the data vectors $\underline{X}_n$ and the reference samples y(n) .

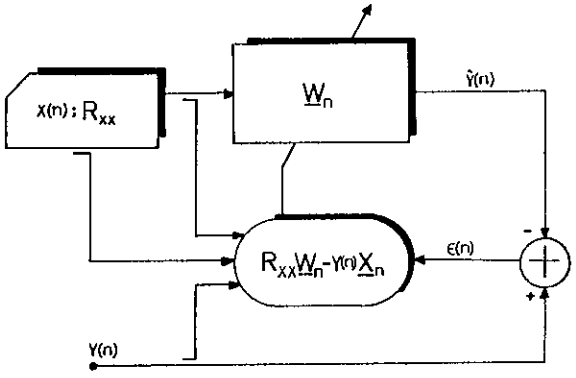

Figure 1. General diagram.

In general, without additional information about R_{xx} and $\underline{P}$, the designer has to use estimates of both R_{xx} and $\underline{P}$. The most familiar approach is the Steepest Descent Method based in the gradient of the error given in (1.2):

$$\underline{W}_{n+1} = \underline{W}_n - \mu\underline{\nabla}_n \qquad (2.5)$$

with $\underline{\nabla}_n = R_{xx}\underline{W}_n - \underline{P}$ the exact gradient vector of the M.S.E. and μ the step size.

But, the knowledge about R_{xx} and $\underline{P}$ is partial and thus, the exact gradient is substituted by an estimate ($\hat{\underline{\nabla}}_n$):

$$\underline{W}_{n+1} = \underline{W}_n - \mu\hat{\underline{\nabla}}_n \qquad (2.6)$$

Widrow proposed as gradient $\hat{\underline{\nabla}}_n = e(n).\underline{X}_n$ (i.e. the so-called instantaneous gradient) [3], and L.J. Griffiths to $y(n)\underline{X}_n - \underline{P}$ [2] in an adaptive array context. As it is seen, the Widrow approach is a complete instantaneous estimate and it presents a very computation simplicity. For this reasons, this proposal will be used in one of our final algorithms.

3. DESCRIPTION OF THE METHOD:

In some cases, the data sequence {x(n)} is exactly known (i.e. active sonar, time delay estimation, ...), and in consequence, its autocorrelation matrix and its inverse matrix can be considered as data. Thus, under this condition, it seems that the best alternative is directly the Wiener equation for the filter coefficient estimation, that is:

$$\underline{W}_n = R_{xx}^{-1}.\hat{\underline{P}}_n \qquad (3.1)$$

where: $\hat{\underline{P}}_n$ is a vector cross-correlation estimate, for instance, an instantaneous one:

$$\hat{\underline{P}}_n = y(n).\underline{X}_n \qquad (3.2)$$

but it is clear that this method doesn't make enough use of the past information. Among many alternatives, a possible one could be the following smoothed version:

$$\underline{W}_n = \alpha.\underline{W}_{n-1} + (1-\alpha)R_{xx}^{-1}.\hat{\underline{P}}_n \qquad (3.3)$$

where α is a constant such that $0<\alpha<1$. However, in actual situations, the evaluation of the expressions will be made with finite arithmetics and between them, basically, in fixed point. On the other hand, the election of parameter α is a difficult issue because it will affect strongly to the convergence speed and it is not clear.

Thus, even in this case, it seems to be reasonable the election of adaptive expressions based on the gradient of the mean square error.

Under the conditions of the problem, our proposal for the gradient estimate is:

$$\hat{\underline{\nabla}}_n = R_{xx}.\underline{W}_n - y(n).\underline{X}_n \qquad (3.4)$$

instead of the reported by Widrow and Griffiths in other contexts. If the prior knowledge of R_{xx} is not complete, an adaptive actualitation could be made by the successfully employed expression:

$$\hat{R}_{xx}(n) = (1-\beta)\hat{R}_{xx}(n-1) + \beta\underline{X}_n^*\underline{X}_n^T;\ 0<\beta<1 \qquad (3.5)$$

Where: * denotes complex conjugate, expression that has been verified, recently, into the Window Methods of Spectral Estimation [5], [6], giving a good physical sense to its ordinary use.

At this moment, it is necessary to describe the 'step size' (μ) actualitation in each iteration. As it is well known, the election of this parameter will affect to the convergence time, the stability and many other parameters of the algorithm. In this sense, for the L.M.S. method, it is shown that a sufficient condition to guarantee stability is that $0<\mu<1/\lambda_{max}$, where λ_{max} is the maximum eigenvalue of R_{xx}.

Another strategy is the so-named Accelerated Gradient (A.G.) approachs. In this way, Powell argued that a suitable election of μ is such that the M.S.E. (2.3) is minimized in each iteration. The obtained expression is:

$$\mu_n = \frac{\nabla_n^T \nabla_n}{\nabla_n^T R_{xx} \nabla_n} \qquad (3.6)$$

Thus, the optimum is just the invers of the Rayleigh quotient, which it is ensured to be bounded by the inverses of the maximum and minimum eigenvalues of R_{xx}:

$$\frac{1}{\lambda_{max}} < \mu_n < \frac{1}{\lambda_{min}} \qquad (3.7)$$

Our proposal differs in the minimization objective. Assuming R_{xx} exactly known, we consider that the best error to be minimized is the M.S.E. associated to the weights, given by:

$$\varepsilon'^2 = E((\underline{W}-\underline{W}_{opt})^T(\underline{W}-\underline{W}_{opt})) \qquad (3.8)$$

For this new objective, the optimum 'step size' (μ_n') is found to be:

$$\mu_n' = \frac{\nabla_n^T R_{xx}^{-1} \nabla_n}{\nabla_n^T \nabla_n} \qquad (3.9)$$

relation that can be expressed in many other equivalent forms through the use of the various gradient estimates.

The analysis of the new quotient (3.9) shows that its variation range is bounded by the same values (3.7) that for Powell's relation (3.6). As it is seen, this optimum needs the knowledge of the inverse of R_{xx}, and thus, it will be a usefull relation only in the problem under study.

The gradient in (3.9) can be evaluated by the proposed relation (3.4), but for the sake of simplicity, other possibility is the use of the Wiener's gradient, leading to an equivalent approach such that:

$$\mu_n' = \frac{\underline{X}_n^T R_{xx}^{-1} \underline{X}_n}{\underline{X}_n^T \underline{X}_n} \qquad (3.10)$$

that represent the most simplified estimate.

The comparison between the Powell's step size (3.6) and the new quotient suggests the following natural generalitation:

$$\mu_n^K = \frac{\nabla_n^T R_{xx}^K \nabla_n}{\nabla_n^T R_{xx}^{K+1} \nabla_n} \qquad (3.11)$$

This complete set of possible values are shown to satisfy the bound conditions (3.7) for each K, too, and particularly leads to (3.6) and (3.9) for K=0 and K=-1, respectively.

Again for (3.11) any approach could be adopted for the gradient, suplying different estimates of (3.11) like in (3.10).

Under our approach, the gradient noise, coefficients error, missadjustment error and convergence rate have shown a better performance with respect other algorithms. Besides, the A.G. propossal presents the same computation cost as for Powell's one, and the same structure.

Finally, a complete study of (3.11) for different K-values will be presented in further papers.

4. ALGORITHMS AND IMPLEMENTATIONS

As it has been described, the main proposed scheme consists in the iterative evaluation of (2.6), where the gradient is given by (3.4) and the step size by (3.9).

It is known that the A.G. Methods present a hard computation effort, nevertheless, using VLSI techniques, it is not a trouble. Keeping it in mind, our scheme can be modified to achieve maximum simplicity. For intance, the gradient (3.4) could be substituted by the simpler one given by Widrow, leading to relation (3.10). The analysis of this way shows that many systolic realizations can be adopted, and among them, the CORDIC one seems to be the best under low sampling rate assumption.

One possible implementation is shown in fig. 2, making use of only two of the six basic CORDIC elements. Note in fig. 2, that normalizations needed to ensure convergence in CORDIC elements can be applied jointly to the reference signal y(n). So, it is just needed a constant factor, and it yields a very compact realization.

The network to evaluate the value (3.10), that is not in fig. 2, will only requiere a normalized matrix-vector and vector-vector product and no-particular comment will be necessary.

5. CONCLUSIONS

This paper has considered the use of prior information into adaptive algorithms. It has shown that the complete knowledge of the input data autocorrelation matrix or power spectrum density function can be used to improve the gradient estimate and defining a better criterion that Powells' to optimize the step size in each iteration. As a consecuence of the found relation a generalization of the step size computation has been presented. The proposal has shown an improvement in the main parameters of the adaptive system.

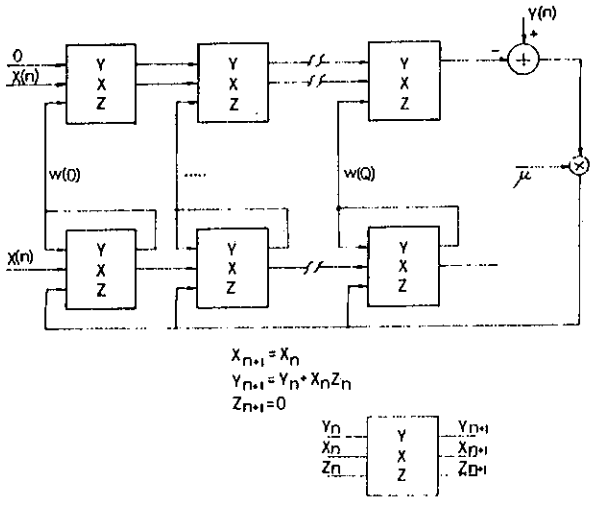

Figure 2. Systolic Implementation.

REFERENCES

|1| R.A. Monzingo and T.W. Miller. "Gradient-Based Algorithms". John Wiley & Sons, Inc. 1980.

|2| L.J. Griffiths. "A Simple Adaptive Algorithm for Real-Time Processing in Antenna Arrays". Proc. of the IEEE, Vol. 57, No. 10. Oct. 1969.

|3| B. Widrow et al. "Adaptive Antenna Systems". Proc. IEEE, Vol. 55, pp.2143-2159. Dec. 1967.

|4| B. Widrow et al. "A Comparison of Adaptive Algorithms Based on the Methods of Steepest Descent and Random Search". IEEE Trans. Antennas and Prop., Vol. AP-24, No. 5, Sep. 1976.

|5| M. Bertran & C. Nadeu. "On the Inclusion of Prior Information in the Window Method of Spectral Estimation". Proc. of MELECON'85/Vol. II, pp.59-61. Oct. 1985. Madrid.

|6| R.A. Monzingo and T.W. Miller. "Recursive Methods for Adaptive Array Processing". John Wiley & Sons, Inc. 1980.

|7| J.S. Walther. "A Unified Algorithm for Elementary Functions". 1971, Spring JCC, pp. 379-385.

|8| H.M. Ahmed et al. "A VLSI Speech Analysis Chip Set Based on Square-Root Normalized Ladder Forms". ICASSP 1981, pp.648-653.