International Telecommunications Symposium – ITS 2002, Natal, Brazil

Adaptive Criteria Optimization as a Least Squares Problem

Romis Ribeiro de Faissol Attux, João Marcos Travassos Romano

DSPCom – DECOM – FEEC – Unicamp, Campinas, Brazil
{romisri, romano}@decom.fee.unicamp.br

Abstract – **In this work, we propose an approach to the analysis of adaptive filtering criteria, based on an original formulation via least squares. This proposal, though based on some model considerations, possesses great generality and simplicity, encompassing many filtering structures and criteria. The approach is used on the study of three blind techniques (decision-directed, constant modulus and modified Sato) and of its relations to the supervised Wiener criterion. The method allows us to view adaptive techniques in a simple and unusual fashion and brought out theoretical insights on blind receivers and their relationship. The whole theoretical framework has a deep geometric appeal and highlights the relevance of separation conditions to static criteria analysis.**

I. INTRODUCTION

THE last twenty years saw a great effort to put adaptive filtering theory on a solid basis. Criteria, algorithms and alternative filter structures were proposed and analyzed under different approaches, what has revealed astounding features and interconnections. However, as far as blind equalization or unsupervised adaptive filtering are concerned, a lot of aspects remain to be revealed.

In the present paper, we propose a theoretical framework in which adaptive filtering and blind equalization criteria optimization are posed as a least squares problem. The approach provides a simple but quite general ground for analysis and comparison of equalization techniques and offers an original point of view about their essential aspects.

The method is applied to the analysis of four well-known optimization criteria: Wiener, decision-directed, modified Sato [1] and constant modulus [2]. Such analysis deals with the concept of linear separation conditions and leads us to reveal some strong connections between Wiener solutions and the different blind receivers. As a consequence, we also dare to extend the conjecture stated by Johnson et al. in [3], concerning a possible (and strong) link between the constant modulus and Wiener criteria.

The paper is divided as follows: section II poses the necessary background on the general adaptive equalization problem. Section III is totally devoted to establish the proposed theoretical framework. The most important features issues from our approach are discussed in the following sections (IV, V and VI). Finally the section VII summarizes the main conclusions.

II. ADAPTIVE EQUALIZATION

The goal of communication is to assure proper message interchange between a transmitter and a receiver, interconnected by a certain channel. A corresponding simple schema is sketched in Fig. 1.

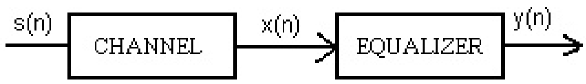

Fig. 1. A Simple Communication System

A relevant practical issue is that the channel may provoke intersymbol interference (IIS), which "spreads" the transmitted pulses, causing undesirable superposition and consequent performance degradation. The most usual counteraction to prevent such distortion from achieving intolerable levels is the introduction of a receiver filter named *equalizer*. Its input-output relation is, in a broad sense, expressed by:

$$y(n) = F[\mathbf{x}(n), \mathbf{w}(n)], \tag{1}$$

where $\mathbf{x}(n)$ is a vector containing the N last channel output samples and $\mathbf{w}(n)$ is the parameter vector at the instant n. This is the general representation of a *finite impulse response (FIR)* structure.

Linear filters are the most common choice for the equalizer role, so that F(.) become:

$$F[\mathbf{x}(n), \mathbf{w}(n)] = \mathbf{w}(n)^H.\mathbf{x}(n), \tag{2}$$

This is the equation of an hyperplane, to which we shall refer, for the sake of simplicity, as a plane.

A crucial problem is *how to adjust filter parameters* in order to attain an optimal condition in some sense. From this matter arises *adaptive filtering theory*, based on several *adaptive criteria*, which can be divided into two classes: *supervised* and *unsupervised*.

To the first class belongs the cornerstone of the whole theory: the Wiener criterion (WIC), defined by the following mean square error (MSE) cost function:

$$J_W = E\{[s(n-d) - y(n)]^2\}, \qquad (3)$$

where d is the equalization delay. If the delay is given *a priori*, this function has a single minimum, called *Wiener solution*. However, if we consider the delay as a degree of freedom, various Wiener solutions will contribute to the function multimodality.

Many criteria belong to the second class. Three are of particular interest here, namely, the *decision-directed criterion* (DDC), the *modified Sato criterion* (MSC) and the *constant modulus criterion* (CMC). Their cost functions are respectively given by [4]:

$$J_{DD} = E\{[\mathrm{dec}[y(n)] - y(n)]^2\}, \qquad (4)$$

where dec (.) is a quantizer mapping,

$$J_{MSC} = E\{[R_1 - |y(n)|\,]^2\}, \qquad (5)$$

where R_1 is defined by:

$$R_1 = \frac{E[|\,s(n)\,|^2]}{E[|\,s(n)\,|\,]} \qquad (6)$$

and

$$J_{CM} = E\{[R_2 - |y(n)|^2\,]^2\}, \qquad (7)$$

with R_2 defined as:

$$R_2 = \frac{E[|\,s(n)\,|^4]}{E[|\,s(n)\,|^2]} \qquad (8)$$

The last two decades have seen a great deal of work that launched and scrutinized the foundations of blind equalization theory. The works of Lucky, Sato, Godard and many others brought out new criteria, which have been the subject of important theoretical analysis [1]–[5]. A remarkable need arose: to understand not only the criteria *per se*, but also their interconnections. This is certainly one field of considerable interest nowadays and we think that it holds the key to important insights. The contribution of the present work comes in this sense, although many opened questions remain to be investigated.

III. Criteria Optimization as a Least Squares Problem

Most communication channels are modeled as a FIR filter. This gives rise to a great number of interesting properties, many of which have been described in several works concerning the "equalization as a classification task" framework [6].

Now, if the transmitted signal s(n) belongs to a finite alphabet and the channel is a FIR filter, it is straightforward that x(n) be constrained to a finite set of possible values. It holds:

$$x(n) = \Phi[s(n-a_1), s(n-a_2), \ldots, s(n-a_p)] \qquad (9)$$

where p is a memory index. This condition will be always assumed in the sequel. The set of possible p-uples is finite due to s(n) character. Besides, if x(n) has a limited repertoire of values, the same property holds for the vector: $\mathbf{x}(n) = [x(n)\ x(n-1) \ldots x(n-k+1)]^T$.

An example can clarify these points.

Example 1

Suppose that s(n) is binary (+1 / -1) and i.i.d. (independent and identically distributed). The channel is linear, noiseless and has an impulse response given by $h(n) = \delta(n) + 0.6\delta(n-1)$. Let us assume a two-coefficient equalizer, what requires a two-element input vector $\mathbf{x}(n) = [x(n)\ x(n-1)]$. From this information, we build Table 1.

The channel produces eight possible vectors, each one associated to a triple of transmitted samples. The distribution of these points is plotted in Fig. 2.

TABLE 1
POSSIBLE INPUT COMBINATIONS

x(n)	**x(n-1)**	**s(n)**	**s(n-1)**	**s(n-2)**
1.6	1.6	1	1	1
1.6	0.4	1	1	-1
0.4	-0.4	1	-1	1
0.4	-1.6	1	-1	-1
-0.4	1.6	-1	1	1
-0.4	0.4	-1	1	-1
-1.6	-0.4	-1	-1	1
-1.6	-1.6	-1	-1	-1

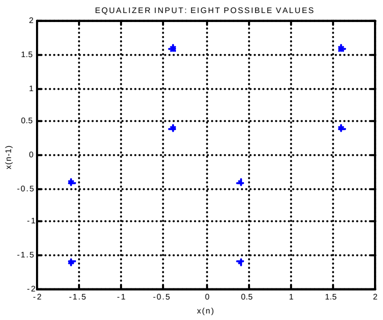

Fig. 2. Possible Input Vectors

We now return to the theoretical discussion to access how the resulting pattern of the mentioned distribution will determine the equalizer performance.

Indeed, many criteria to be used in adaptive filtering (supervised or not) take the following form:

$$J = E\{[k(n) - G(\mathbf{w}, \mathbf{x})]^2\}, \tag{10}$$

as can be promptly confirmed by an inspection of equations (3, 4, 5, 7). In (10), k(n) contains information about or an estimate of the transmitted signal and G(.) is related to the filtering structure.

From our previous discussion about the input vector, we conclude that, for a given **w**, G(**w**, **x**) is a discrete random variable. We may thus rewrite (10):

$$J = \sum_{i=1}^{X} \text{prob}(\mathbf{x}_i).[k_i - G(\mathbf{w}, \mathbf{x}_i)]^2 \tag{11}$$

The operator prob(.) simply stands for probability of occurrence and k_i is a value associated to $\mathbf{x}_i$. So, in (11) we have a general model for equalization criteria in our working conditions. In this article, we deal exclusively with equiprobable and i.i.d. sources. This allows us to rewrite (11) as:

$$J = \frac{1}{X} \sum_{i=1}^{X} [k_i - G(\mathbf{w}, \mathbf{x}_i)]^2 \tag{12}$$

In (12) holds the crucial feature of our proposal. If the cost functions in adaptive filtering can be expressed as in (12), it means that they can be view as a *least squares problem*. Then, from now we use the expression *least squares approach (LSA)*. Clearly, the LSA must not be confused with the well-established theory of *least squares algorithms*. They are essentially different and deal with distinct aspects of the filtering problem.

The cost function model (12) has a great degree of generality. It encompasses problems so diverse as, for instance, linear equalizer optimization via CMC or neural network optimization via WIC. In all the cases, all we need to know about the structure and the criterion are the corresponding values of k_i and G(.).

It is our belief that this simplicity can be very useful to analyze the "criterion-in-itself" as well as the liaisons between different techniques. So, let us see how each criterion fits the general LSA model.

Wiener

Corresponding settings are $k_i = s_i$ and $G(\mathbf{w}, \mathbf{x}_i) = F(\mathbf{w}, \mathbf{x}_i)$.

Modified Sato

The choices must be $k_i = R$ and $G(\mathbf{w}, \mathbf{x}_i) = |F(\mathbf{w}, \mathbf{x}_i)|$.

CM

To obtain the CM criterion, it is necessary to pose $k_i = R_2$ and $G(\mathbf{w}, \mathbf{x}_i) = |F(\mathbf{w}, \mathbf{x}_i)|^2$.

DD

The DD criterion arises when one makes $k_i = \text{dec}[F(\mathbf{w}, \mathbf{x})]$ and $G(\mathbf{w}, \mathbf{x}_i) = F(\mathbf{w}, \mathbf{x}_i)$.

Discussion

The original aim of equalization is represented by the zero-forcing condition (ZFC) that, in the LSA formulation, would be:

$$F(\mathbf{w}, \mathbf{x}_i) = s_i \text{ for all } i \tag{13}$$

WIC is the only criterion that can be straightly associated to this goal (we consider no source correlation). However, we do not wish to create a wrong notion that other criteria cannot achieve a ZFC. It is possible, but always through "indirect" ways.

A close study of the "choices" of k_i and G(.) in each one of the previous criteria reveals that the blind techniques use two basic expedients: i-a function G(.) different of F(.) and k_i equal to a constant (MSC, CMC) or ii- G(.) = F(.) and k_i as a nonlinear estimate of s_i (DDC). In both cases, they are using some kind of artifice to obtain the equalizer without the knowledge of the transmitted signal.

IV. Relationship Between Wiener And DD

In order to well pose the mentioned relationship, the first two subsections, IV.A and IV.B, recalls some useful results, while IV.C contains the new analytical work and the obtained results. From now on we assume that the equalizer is a linear filter and that the transmitted signal is zero-mean and binary.

A. Planes And Linear Separation

As shown in (2), linear filters have an input-output relationship identical to the equation of a plane. Besides it is easily shown that this plane always crosses the origin.

From (2), it comes that the linear equalizer structure can only perform linear separation. In fact, suppose that we transmit a series of +1 and –1 pulses. The quantizer will map the equalizer output onto +1 if it is positive and onto –1 if it is negative. The equation $\mathbf{w}^H.\mathbf{x} = 0$ establishes a decision boundary X^0 in the equalizer parameters space. Two regions could be formed: X^+, in which y(n) > 0 and X^-, in which y(n) < 0. Every point of X^+ will be reconstructed as a +1 pulse, and every point of X^- as -1.

It is important to remark that there are infinite possible values of **w** that lead to the same regions. These solutions differ only by a gain, which is automatically adjusted by all the presented criteria.

B. Equalization Delay Influence

Equalization delay influence is an aspect present in several works on the "equalization as a classification task" framework. It is also important in the context of the LSA, as we intend to show.

Let us turn back to the model presented in example 1. Suppose that we wish to employ the Wiener criterion. One immediate question arises: which equalization delay should be chosen? This question may sound, at a first sight, quite pointless. However, if we carefully consider the results in Table 1, its importance becomes patent.

Suppose that we chose d = 0. This means that our values of s_i in (12) will be given by the column s(n) in Table 1. In Fig. 3, we label the points $\mathbf{x}_i$, presented in Fig. 2, according to their corresponding values of s_i. We also plot the regions X^-, X^+ and X^0 for the Wiener solution. Their parameters are set to make y(n) as close as possible to +1 in the X^+ points and to –1 in the X^- points.

This delay configuration allows linear separation and it is not surprising at all that the Wiener solution is associated to a *linear separation condition (LSC)*. It is reasonable to state that the Wiener solution will produce a LSC when the last is attainable, as previously demonstrated in [7].

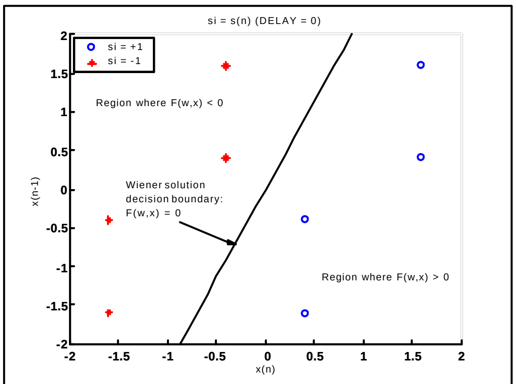

Fig. 3. Situation for s_i = s(n)

This approach helps us to see why, in this model, it is not possible to obtain perfect (zero-forcing) equalization. The eight points ($\mathbf{x}_i$, s_i) are not coplanar, what accounts for a non-zero residual MSE. Perhaps other structures can achieve this goal but not a linear filter. Finally, this solution accounts for a residual error of 0.0870.

Let us now study the second possibility, that is s_i = s(n-1). The Fig. 4 brings the labeled points and the Wiener solution boundary. Again we have a LSC, and the residual error is 0.2417. This reveals that this point distribution was less favorable than the previous one, what is apparent from the figure.

We finally consider the last possibility, s_i = s(n-2). In Fig. 5, we have the labeled points and the Wiener boundary. In this case, *we have not a possible LSC*. As expected, the minimum has a large residual error: 0.6713. This illustrates the connection between good classification and the least squares problem inherent to adaptive filtering.

After this case study, we are ready to search for relations between the DD and Wiener criteria. We will consider, without loss of generality, that the signal assumes values +1 and –1.

C. Linear Separation and Minima Analysis

As seen in section III, DD and Wiener functions have distinct k_is. Two crucial questions are: can the costs coincide? And how?

A natural procedure is to look for situations where k_i = dec[F(**w**, $\mathbf{x}_i$)] and k_i = s_i coincide for every possible value of i. In these situations, the cost associated to a solution **w** would be the same. Since we are working with blind criteria, it is also acceptable to have k_i = -s_i.

Now let us pose two important results as the basis of our subsequent analysis.

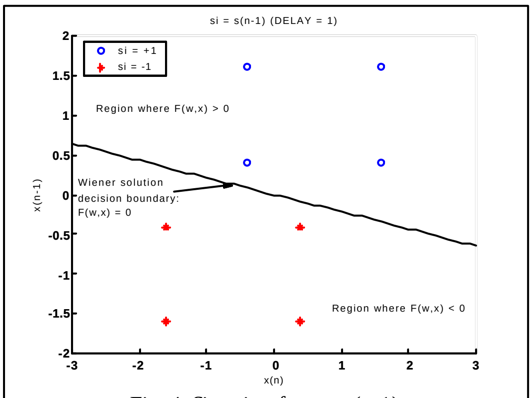

Fig. 4. Situation for $s_i = s(n-1)$

Result 1: If, for a given **w**, dec[F(**w**, $\mathbf{x}_i$)] = s_i or dec[F(**w**, $\mathbf{x}_i$)] = $-s_i$, for every possible i, then $J_w = J_{DD}$ at this point. The proof comes from immediate substitution in (12).

Result 2: The conditions dec[F(**w**, **x**)] = s_i and dec[F(**w**, **x**)] = $-s_i$, for every possible I, only hold if **w** leads to a LSC.

As dec[F(**w**, **x**)] = +1 in X^+ and dec[F(**w**, **x**)] = -1 in X^-, it is imperative that the distinctly labeled points fall in distinct regions. As the regions are divided by a linear boundary X^0, a LSC is necessary.

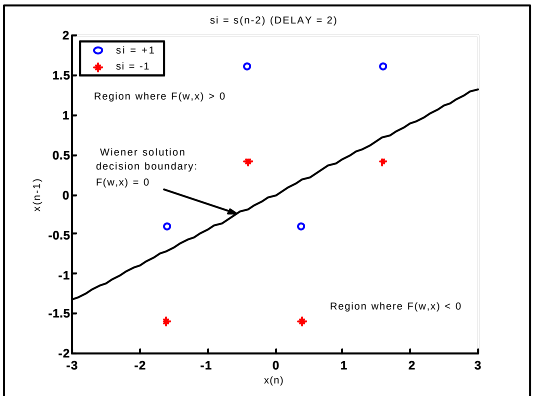

Fig. 5. Situation for $s_i = s(n-2)$

Result 3: If **w** leads to a LSC, this condition will also hold for a vicinity in the parameter space.

Notice that $y(n) = \mathbf{w}^H.\mathbf{x} = \mathbf{x}^H.\mathbf{w}$. Given a certain **x**, y(n) is a function whose domain is the parameter space. The well-known continuity of a plane assures us that, for every **w**, there will be a vicinity in which no change of y(n) sign shall happen. We conclude the proof by taking the most restrictive region among all.

From results 1, 2 and 3 a central statement follows:

Every Wiener minimum that leads to a LSC will also be a DD minimum and so will, consequently, its symmetrical.

This suggests Wiener minimum separation in two classes: the ones who lead to a LSC and the ones who do not. We will refer to the first class as LSC minima and to the second as NLSC minima.

No coincidence can be guaranteed between a NLSC minimum and a DD minimum. Indeed, the rule is that *equality shall not occur*. This explains the notion of *spurious minimum*, that is, a minimum that has no connection with a Wiener receiver. They simply cannot be related to a LSC.

Notice another interesting aspect: the DD cost function will have LSC regions of equality to the Wiener cost function. This allows us to think about polytopes in the context of LSA.

Another point is that for multilevel modulations, there shall occur multiple frontiers. However, many results can be extended to this more general case.

V. Relationship Between CM Criterion And Wiener Criterion

Until now, the studied criteria had a G(.) that was a plane or, at least, "plane-like". This is not the case when the CMC is in question. *Its G(.) has a parabolic shape*, what is a remarkable difference.

The point is: until now, we were using "plane-like" G(.) functions to optimize the parameters of a plane (the linear filter). With the CMC, *we use a parabolic G(.) function to obtain the parameters of a plane*. This accounts for the well-known difference between a CMC minimum and a close Wiener solution.

To probe further, we will briefly turn to another criterion: the MSC. It has k_i = R and G(**w**, **x**) = |F(**w**, **x**)|. Without loss of generality, let us consider R = 1. In this case, k_i = 1 and the function to be optimized is the modulus of a plane, to which we will refer to as a V-plane.

Result 4: The WIC and MSC costs do coincide when evaluated at a LSC Wiener minimum.

Considering a LSC, we have the regions X^+ and X^-, created by the Wiener minimum. In the X^+ region, both G(.) functions coincide and so do the k_i. In the X^- region, $G_{MSC}(.) = -G_{WIC}$. However, $k_{iMSC} = -k_{iWIC}$, what preserves the error value. This concludes the proof.

From results 3 and 4, we see that the same minima division holds for the MSC, what was expected. Now we will try to relate these results to the CMC.

We studied the MSC because it is, in form, very close to the CMC. Their difference lies primarily on the function G(.). If parameter change can be related to a deformation, then maybe we can find a direct relationship between WIC and CMC minima.

VI. CONJECTURING ON WIENER / CMC RELATION

This is a section devoted to conjectures on the relation between WIC and CMC minima. Of course, we will formulate them by means of the LSA.

The first conjecture we state is: *every CMC minimum is close to a LSC Wiener minimum*.

We dare to affirm such a thing inspired by a conjecture raised by Johnson et al. in [3], proposed in the context of fractionally-spaced filtering. Their conjecture is that the *CMC minima are always close to the best possible Wiener solutions*. In other words, the CMC "chooses the best delays".

The results presented by Li et al. [5] also were encouraging. In their work, blind criteria minima are also divided in two classes: length- and cost-dependent. So a last question to be opened as a perspective of further studies could be: Is there some kind of correspondence between those two classes and the LSC and NLSC minima?

VII. CONCLUSIONS

The main result present in this work was the establishment of a LSA to the analysis of adaptive equalization criteria under a certain system model. The approach is quite flexible, encompassing different structures and criteria. It provides a rather simple view on the techniques and the problem as a whole.

The study of the DDC under the LSA gave rise to some interesting results concerning separation and cost function structure and also provided a solid basis to relate these minima to the WIC minima also under the separation concept.

The study of the CMC minima allowed us to establish some connection between these and WIC minima, using the MSC as a liason.

At last, we forged a new conjecture about the connection CMC-WIC based on a previous one, made by Johnson et al. in [3]. Subsequent work on the LSA may clarify this proposition bases.

ACKNOWLEDGMENTS

The authors wish to thank Charles Cavalcante, for h is generous expert contribution, and Aline Neves, Cristiano Panazio, Danilo Zanatta Filho and Maurício de Castro for their inestimable help. Romis Attux thanks FAPESP (process no. 01/00680-4) for its financial support.

REFERENCES

[1] – Y. Sato, "A Method of Self-Recovering Equalization for Multi-level Amplitude Modulation", IEEE Trans. on Comm., COM – 23, pp. 679-682, 1975.

[2] – D. N. Godard, "Self-recovering Equalization and Carrier Trackind in Two-dimensional Data Communications Systems", IEEE Trans. on Comm., COM – 28, pp. 1867-1875, 1980.

[3] – C. R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, R. A. Casas, "Blind Equalization using the Constant Modulus Criterion: a Review", Proceedings of the IEEE, Vol. 86, No. 10, pp. 1927-1950, 1995.

[4] – S. Haykin, *Adaptive Filter Theory*, 3rd Edition, Prentice Hall, 1996.

[5] – Y. Li, K. J. Ray Liu, Z. Ding, "Length- and Cost-dependent Local Minima of Unconstrained Blind Equalizers", IEEE Trans. on Signal Processing, Vol.43, No. 9, pp. 2120 – 2129, 1995.

[6] – J. R. Montalvão Filho, J. C. M. Mota, B. Dorizzi, "Some Theoretical Limits for Linear and Nonlinear Equalizers", Journal of the Brazilian Telecommunications Society (SBrT), Vol. 3, No. 2, pp. 22-31, 1999.

[7] - J. R. Montalvão Filho, J. C. M. Mota, B. Dorizzi, C. C. Cavalcante, "Reducing Bayes Equalizer Complexity: a New Approach for Clusters Determination", Proc. of the ITS 1998, pp. 428-433, São Paulo, Brazil, 1998.