BLIND CANCELLATION OF INTERSYMBOL INTERFERENCE IN DECISION FEEDBACK EQUALIZERS

Sylvie MARCOS [1], *Sofiane CHERIF* [2] *and Meriem JAIDANE* [2]

[1]Laboratoire des Signaux et Systèmes, CNRS-ESE, Plateau de Moulon,
91192 Gif-sur-Yvette cedex, France, e-mail : marcos@lss.supelec.fr

[2]Laboratoire des Systèmes de Télécommunications, ENIT/ESPTT, Campus Universitaire,
BP 37, Le Belvédère, Tunis, Tunisia. e-mail : elcelab@esptn.esptt.tn

ABSTRACT

This paper presents the cancellation of intersymbols interferences (ISI) as a criterion for the blind adaptation of decision feedback equalizers (DFE). We show that this criterion is an alternative to the decision directed (DD) algorithm. This paper also proposes to replace the hard limiter in the decision device of DFE by a soft decision implemented by a hyperbolic tangent function in order to escape from local minima during the initialization of the blind algorithms. As the theoretical investigation of these criteria and algorithms is difficult, we here analyse some simple examples to illustrate the interest of the proposed solutions.

1. INTRODUCTION

High speed data transmission over band-limited and/or dispersive channels requires the equalization technique to remove the intersymbol interference (ISI) between the successive transmitted data. Equalization is usually performed by a linear transversal equalizer (LTE) followed by a decision device. When the channel transfer function has zeros close to the unit circle, the LTE has to compromise between inverting the channel transfer function and avoiding the noise enhancement. Decision feedback equalizers (DFE), which consist of a recursive filter in conjunction with the decision device in the feedback loop, offer improved performances compared to the LTE in case of difficult channels.

The optimization of the equalizers is achieved by adaptive algorithms derived from the minimum mean square error (MSE) criterion between the equalizer output and the transmitted data sequence. During a preliminary learning phase, a reference data sequence, known by the receiver, is transmitted in order to initialize the adaptation of the equalizer. In some applications, this preliminary learning phase is too constraining.

Blind or self-adaptive equalization which consists in recovering the transmitted data sequence without a reference sequence to be transmitted, has drawn the attention of many researchers. Different classes of techniques have been considered. Among them, the Bussgang-type algorithms [1] are least mean square (LMS) algorithms which include a nonlinearity on the output of the equalizer. Whereas these algorithms (see [2][3], for example) were initially introduced and analysed for LTE, they may, however, be implemented for DFE [4]. The present paper focusses on the blind adaptation of DFE.

The following section introduces the ISI cancellation (ISIC) criterion as an alternative to the decision directed (DD) criterion [4] for the blind optimization of DFE. Based on simple examples, the error surfaces are investigated as well as the convergence properties of the corresponding self-adaptive algorithms. Section 3 presents a procedure for initializing the blind algorithms in order to prevent them from getting stuck in a local minimum. This is obtained by replacing the hard limiter in the decision device by a soft decision function which smoothes the local minima and makes them disappear. Section 4 is the conclusion.

2. BLIND CANCELLATION OF ISI IN DFE

2.1. The ISI cancellation (ISIC) criterion

Let us consider the channel and the DFE in Figure 1. The transmitted symbols a_k are transformed by the channel with coefficients h_i, $i = 0, \dots, N$, into a sequence of received data

$$x_k = \sum_{i=0}^{N} h_i a_{k-i} \qquad (1)$$

We here assume that the a_k are iid and zero-mean and that the channel is causal. The DFE with coefficients g_i, $i = 1, \dots N$, is designed to construct an estimation of the causal ISI so that the signal at the input of the decision device is

$$y_k = x_k - \sum_{i=1}^{N} g_i \hat{a}_{k-i} \qquad \hat{a}_{k-1} = f(y_{k-1}) \qquad (2)$$

where $f(.)$ is the nonlinear function implemented in the decision device. The optimization of the DFE is usually performed by minimizing the MSE $J_0 = \mathrm{E}\left[(a_k - y_k)^2\right]$ between the transmitted symbol and the signal y_k before decision. The corresponding LMS algorithm then requires a preliminary phase during which the transmitted data must be known from the

0-7803-2431-5/95 $4.00 © 1995 IEEE

receiver. During the running phase, the reference symbol is replaced by the estimated symbol so that the criterion being performed is then $J_1 = \mathrm{E}\left[(\widehat{a}_k - y_k)^2\right]$. Used from the beginning of the adaptation, it gives the DD algorithm [2].
We here consider and analyse the minimization of $J_2 = \mathrm{E}\left[y_k^2\right]$. Under the assumption of iid and zero-mean a_k, this criterion corresponds to the cancellation of the causal ISI before decision. It will be referred to as the ISIC criterion in the following. It can also be viewed as a kind of nonlinear prediction of the received sequence x_k generated by (1). Noting that $J_2 = J_0 + (1-2h_0)\sigma_a^2$, where σ_a^2 is the variance of a_k, it follows that J_0 and J_2 have the same characteristics. The advantage of J_2 over J_0 is that it provides a blind algorithm. Besides, the proposed criterion is valid for any type of modulation of the transmitted signals.
The following is devoted to a simplified analysis of the error surfaces associated with J_1 and J_2.

2.2. Analysis of the error surfaces

We could apply to the ISIC criterion the analysis made in [4] for the DD algorithm. The results are, however, too general to be useful for the comparison of J_1 and J_2. By considering simple systems, we can investigate the surfaces associated with J_1 and J_2 and the conditions of existence of local minima as functions of the characteristics of the channel. Let us consider the case of *i)* iid $a_k \in \{-1, 1\}$ with equal probability, *ii)* a channel with two coefficients, h_0 and h_1, i.e. $N = 1$ and $g_1 = g$, *iii)* $f(.) = sgn(.)$. The cost functions are thus

$$\begin{aligned} J_1(g) &= J_2(g) + 1 - 2h_0\mathrm{E}\left[a_k\widehat{a}_k\right] \\ &\quad -2h_1\mathrm{E}\left[a_{k-1}\widehat{a}_k\right] - 2g\mathrm{E}\left[\widehat{a}_{k-1}\widehat{a}_k\right] \\ J_2(g) &= h_0^2 + h_1^2 + g^2 - 2gh_1\mathrm{E}\left[a_{k-1}\widehat{a}_{k-1}\right] \end{aligned} \quad (3)$$

Clearly, the probability distribution of the data $\widehat{a}_{k-1}$ and $\widehat{a}_k$ also depends on the DFE parameter g. Taking into account that y_{k-1} can only take a finite number of values depending on the channel and DFE parameters and on the data levels, it appears that

$$\mathcal{D} = \{g : y_{k-1} = 0,\ a_{k-1},\ a_{k-2},\ \widehat{a}_{k-2} \in \{-1, 1\}\} \quad (4)$$

defines a set of hyperplanes in the DFE parameter space. These hyperplanes yield a number of manifolds, referred to as polytopes in [4], within which y_{k-1} is in the same decision region, *i.e.*, $f(.)$ takes the same value. Within each manifold, $\widehat{a}_k$ and $\widehat{a}_{k-1}$ are independent of g and the cost functions (3) in the corresponding manifold then reduce to a parabola. Each parabola has a corresponding minimum. This minimum may, however, not belong to the manifold associated with the parabola. If the minimum belongs to the corresponding manifold, it is said to be locally attainable [4] and it yields a local minimum in the resulting error surface.
Within a given polytope, J_1 and J_2 (3) are minimum for

$$\begin{aligned} g_{1,opt} &= h_1\mathrm{E}\left[a_{k-1}\widehat{a}_{k-1}\right] + \mathrm{E}\left[\widehat{a}_{k-1}\widehat{a}_k\right] \\ g_{2,opt} &= h_1\mathrm{E}\left[a_{k-1}\widehat{a}_{k-1}\right] \end{aligned} \quad (5)$$

respectively. It then appears that the DD cost function may have more local minima than the proposed one. This is illustrated by the following examples.
Examples: Let us consider the manifold where the decisions are all correct, i.e., $\widehat{a}_k = a_k$, so that

$$\mathrm{E}\left[a_{k-1}\widehat{a}_{k-1}\right] = 1 \quad \text{and} \quad \mathrm{E}\left[\widehat{a}_{k-1}\widehat{a}_k\right] = 0$$

and the cost functions (3) become

$$\begin{aligned} J_1(g) &= 1 + h_0^2 + h_1^2 + g^2 - 2gh_1 - 2h_0 \\ J_2(g) &= h_0^2 + h_1^2 + g^2 - 2gh_1 \end{aligned} \quad (6)$$

They both have the same minimum $g = h_1$ and this minimum is attainable since it is easy to check that it correponds to correct decisions.
Now, when $g = 0$ and $0 < h_0 < h_1$ (non minimum phase channel) we can establish that

$$\mathrm{E}\left[a_{k-1}\widehat{a}_{k-1}\right] = 0 \quad \mathrm{E}\left[a_{k-1}\widehat{a}_k\right] = 1 \quad \mathrm{E}\left[\widehat{a}_{k-1}\widehat{a}_{k-1}\right] = 0$$

and consequently

$$\begin{aligned} J_1 &= 1 + h_0^2 + h_1^2 + g^2 - 2h_1 \\ J_2 &= h_0^2 + h_1^2 + g^2 \end{aligned} \quad (7)$$

It then follows that J_1 and J_2 have the same attainable minimum $g = 0$ which then corresponds to a local minimum of the global cost functions. Figure 2a exhibits the simulated cost functions J_1 and J_2, and the theoretical curves (6)(7) for J_2 when $h_0 = 0.5$, $h_1 = 1$. Reasoning in the same way, it is easy to check that when $0 < h_1 < h_0$ (minimum phase channel), $g = 0$ does no longer correspond to an attainable minimum. Figure 2b exhibits the cost functions J_1 and J_2, and the theoretical curves (6)(7) for J_2 when $h_0 = 1$, $h_1 = 0.5$. Figure 3 exhibits an example, corresponding to $h_0 = 0.3$, $h_1 = 0.4$, where two minima are attainable for J_2 and five minima are attainable for J_1.

2.3. Self-adaptation of the DFE

Because of the recursive structure of the DFE, the computation of the gradient of J_1 and J_2 would theoretically require an infinite memory of the data [6]. The DD and ISIC algorithms

$$\begin{aligned} g_{k+1} &= g_k + \mu(\widehat{a}_k - y_k)\widehat{a}_{k-1} \\ g_{k+1} &= g_k + \mu y_k \widehat{a}_{k-1} \end{aligned} \quad (8)$$

derived from the cost functions J_1 and J_2, respectively, are approximations of the actual LMS algorithm. It has been established in [4] that the convergence of the DD algorithm within a given polytope is ensured when the step-size μ is small. For the ISIC algorithm, we establish from

$$\begin{aligned} \mathrm{E}\left[g_{k+1}^2\right] &= (1-2\mu)\mathrm{E}\left[g_k^2\right] + \mu\mathrm{E}\left[y_k^2\right] \\ &\quad + 2\mu h_1\mathrm{E}\left[g_k a_k \widehat{a}_{k-1}\right] \\ \mathrm{E}\left[y_k^2\right] &= h_0^2 + h_1^2 + \mathrm{E}\left[g_k^2\right] - 2h_1\mathrm{E}\left[g_k a_k \widehat{a}_{k-1}\right] \end{aligned} \quad (9)$$

and under the hypothesis that $E\left[a_{k-i}\widehat{a}_k\right]$ vanishes when i tends to infinity, and for small values of μ, that

$$\lim_{k\to\infty} \mathrm{E}\left[g^2(k)\right] \approx \mu\frac{h_0^2+h_1^2}{2} \qquad (10)$$

Figure 4a exhibits the convergence of the ISIC algorithm initialized at $g=0$ when $h_0=0.5$, $h_1=1$ for $\mu=0.01$ and $\mu=0.1$. The variance of g_k is found to be equal to 6.510^{-3} and 0.01, respectively, which is in agreement with 610^3 and 0.06 found with (10).
The simulations have also shown that when $\mu > 0.045$ the ISIC algorithm can escape from the local minimum $g=0$ towards the global minimum corresponding to $g=h_1$. By assuming that g_k is Gaussian and zero–mean with variance (10) in the polytope around $g=0$, it is easy to compute that the probability to escape from the local minimum $g=0$ bounded by $[h_0-h_1, h_1-h_0]$ is greater than 0.98 when $\mu > 0.1$ which yields a satisfying condition on μ compared to the simulation results. We can observe in Figure 4b that the DD algorithm cannot escape from the local minimum. For the ISIC algorithm, when $h_0=0.3$ and $h_1=0.4$, the approximation yields that μ should be greater than 0.02 to escape from the local minimum which is confirmed by the simulation in Figure 5 where $\mu=0.025$. The DD again converges towards a local minimum.
The following is devoted to a modification of the decision function in order to ensure the convergence to the global minimum for any step–size.

3. SOFT DECISIONS IN ADAPTIVE DFE

3.1. The error surfaces with soft decisions

We here suggest to introduce a soft decision at the beginning of the DFE adaptation. This is achieved by implementing a hyperbolic tangent function $f_p(x) = \tanh(px)$, with slope p at the origin, instead of $f(x) = \mathrm{sgn}(x)$. This choice is motivated by the fact that the influence of a false decision when forming the control error in the adaptive algorithm should be reduced. Figure 6 exhibits that this soft decision allows to smooth the local minima as p decreases, until they disappear, while the global minimum is only slightly translated.
A rough theoretical approximation of the cost function J_2 can be established by assuming that y_k at the input of the decision device is a zero–mean Gaussian process. This hypothesis is far from being true in the case of data transmission. It becomes, however, ever more true as the data take a larger number of values and as the channel is longer. With this hypothesis, a simplified version of the Price theorem [5] can be used to derive a recursive computation of the variance σ_k^2 of y_k,

$$\begin{aligned}\sigma_k^2 = h_0^2 + h_1^2 + g^2\tfrac{2}{\pi}Arcsin\left[\tfrac{2p^2\sigma_{k-1}^2}{1+2p^2\sigma_{k-1}^2}\right] \\ -\tfrac{4h_1 g}{\pi}Arcsin\left[\sqrt{\tfrac{2p^2\sigma_{k-1}^2}{1+2p^2\sigma_{k-1}^2}}c\right]\end{aligned} \qquad (11)$$

where $c = \sqrt{2/\pi}h_0/\sigma_{k-1}^2$ is a correlation coefficient constrained to be smaller than 1. Consequently J_2 can be approximated by the limit of the above recurrence.
For example, we can see in Figure 6 the approximation (11) of J_2 when $p=4$ and $p=1$. Although the global minimum does no longer exist, the approximated error surface and the actual one coincide for the rest of the curve. This suggests that the rough hypothesis of y_k being Gaussian is ever more true far from the global minimum.

3.2. Adaptation of the DFE with soft decisions

By replacing the sign function by a hyperbolic tangent function with slope p giving a unique global minimum in the cost function, we can verify that the ISIC algorithm converges towards this minimum for any step–size of the algorithm. For example, we can see in Figure 7 that the algorithm, initialized at $g=0$, converges for $p=1$ with $\mu=0.01$ whereas it did not converge for this step–size with the sign function (compare with Figure 4).
If the global minimum of the error surface with the soft decision is in the polytope corresponding to the global minimum of the error surface with the hard limiter, we can start the algorithm with a small slope in the decision function and, after a few iterations, we can switch to the algorithm with the hard limiter. The result can be seen in Figure 7.
It is worth noting that another advantage of using soft decisions is that we can use, for the adaptation of the DFE, the finite memory recursive LMS (FMRLMS) algorithms [6] which were introduced to adapt nonlinear recursive filters and which were found to be useful in difficult situations, like convergence near unstable regions or in a non–stationary context [7].

4. CONCLUSION

In this paper, we proposed the ISIC criterion for the self–adaptation of DFE as an alternative to the DD algorithm. Through simple examples, we have shown that the ISIC criterion is less subject to local minima than the DD criterion and that the ISIC algorithm is less subject to getting stuck in the local minima than the DD algorithm. These results should, however, be further analysed in a more general context. Besides, we proposed to replace the hard limiter in the decision device by soft decisions during the initialization of the blind algorithms. This was shown to reduce the influence of false decisions and to smooth the local minima so that the global convergence may occur for any step–size and any initialization.

5. REFERENCES

[1] J.J. Bussgang, "Cross correlation functions of amplitude–distorted Gaussian signals", *Tech. Rep. 216, MIT*, Cambridge, MA, 1992.

[2] O. Macchi and E. Eweda, "Convergence analysis of self–adaptive equalizers", *IEEE Trans. on Inf. Theory*, Vol.IT-30, 1984.

[3] A. Benveniste and M. Goursat, "Blind equalizers", *IEEE Trans. on Com.*, Vol. COM-32, 1984.

[4] R.A. Kennedy et al., "Blind adaptation of decision feedback equalizers : gross convergence properties", *Int. J. of Adapt. Control and Signal Proc.*, Vol. 7, 1993.

[5] R. Price,, "A useful theorem for nonlinear devices having Gaussian inputs", *IEEE Trans. on Inf. Theory*, Vol.IT-4, 1958.

[6] S. Marcos et al., "A unified framework for gradient-based algorithms used for filter adaptation and neural network training", *Intern. J. of Circuit Theory and Applic.*, Vol. 20, 1992.

[7] Ch. Vignat et al., "Analysis of gradient-based adaptation algorithms for linear and nonlinear recursive filters", *Proc. of ICASSP*, San Fransisco, 1992.

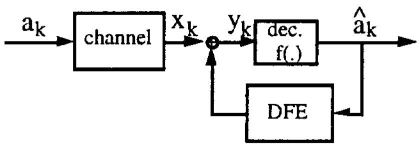

Figure 1 : Decision feedback equalization

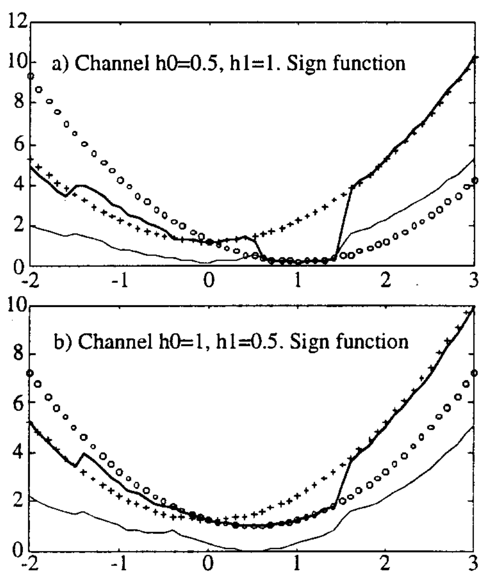

Figure 2 : J1 (–) and J2 (–) wrt g and theoretical curves for J2 in polytopes g=0 (+) and g=h1 (o).

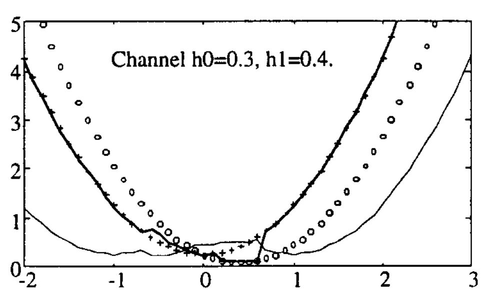

Figure 3 : J1 (–) and J2 (–) wrt g and theoretical curves for J2 in polytopes g=0 (+) and g=h1 (o).

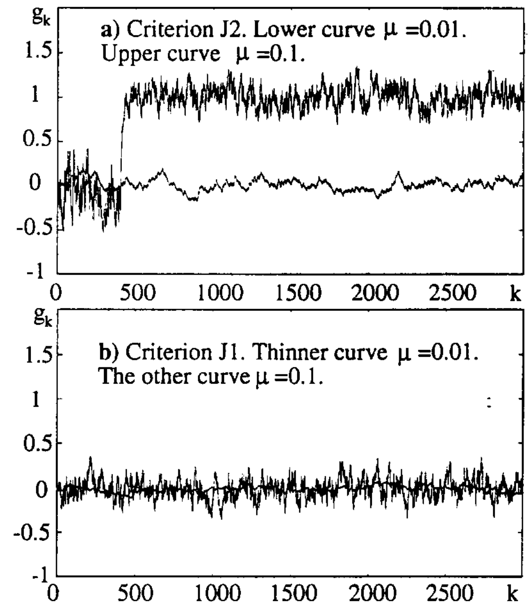

Figure 4 : Convergence of parameter gk. Channel h0=0.5, h1=1. Sign function.

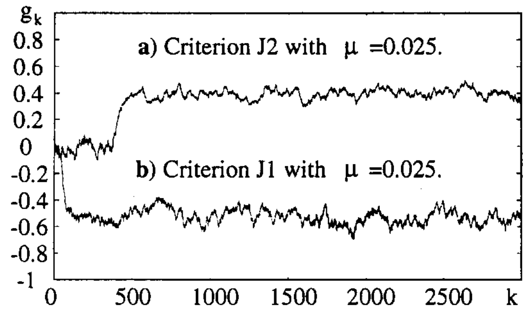

Figure 5 : Convergence of parameter gk. Channel h0=0.3, h1=0.4. Sign function.

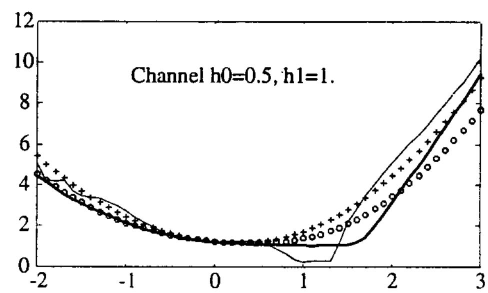

Figure 6 : J2 with soft decision p=4 (–) p=1 (–). Theoretical curves : p=4 (+) p=1 (o).

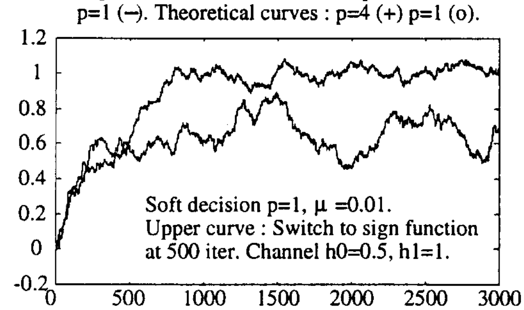

Figure 7 : Convergence of parameter gk.