Relationship Between Supervised and Unsupervised Criteria for Minimum BER Filtering

Charles C. Cavalcante and João Marcos T. Romano

***Abstract*— In this paper we present a relationship between supervised and unsupervised criteria for minimum bit error rate (BER) filtering. A criterion based on the probability density function (pdf) estimation is used to link the minimum mean square error (MMSE) criterion and the maximum *a posteriori* one in order to obtain a linear filter that minimize the BER. An important analytical relationship of the three criteria is presented and analyzed showing that is not possible to achieve minimum BER without training sequences when the pdf estimation-based criterion is considered.**

***Index Terms*— Minimum BER, MMSE, blind criterion, pdf estimation.**

I. INTRODUCTION

Signal processing is a powerful tool on the design of robust digital communication systems. In particular, the recovering device, called equalizer, plays a key role on the project since the interference can damage the transmitted information. Dealing with the mitigation of interference in transmitted signals, the conception of the equalizer is linked to the choice of an optimization criterion able to recover the original information at the receiver.

A classical strategy for the optimization of the equalizer is the use of a sequence known at the transmitter and the receiver and transmitted periodically in order to minimize the square error given by the difference of the transmitted signal and the recovered one. This is known as the minimum mean square error (MMSE) supervised criteria [1].

When there is no such known sequence available, an unsupervised, or *blind*, processing is employed in order to optimize the equalizer [2]. Blind processing is based on some known statistical characteristics of the transmitted signal that are used to estimate the transmitted symbol at the receiver. Even that most blind algorithm have higher computational complexity they have a lower information complexity since they require less information about the signal than supervised strategies.

Despite its frequent use in supervised strategies, the MMSE is not the optimum solution in practical systems [3]. The minimization of the bit error rate (BER) is more useful due to the importance of such measure in practice. Further, it is known that MMSE does not achieve minimum BER when the equalizer does not have an appropriate length [4].

C. C. Cavalcante is with the GTEL Lab., Dep. of Teleinformatics Engineering, Federal University of Ceará (UFC), Phone/Fax: +55 85 288 9470, C.P. 6007, CEP: 60755-460, Fortaleza-CE, Brazil. (email: charles@gtel.ufc.br)

J. M. T. Romano is with the Digital Signal Processing for Communications (DSPCom) Lab., Dep. of Communications, Faculty of Electrical and Computer Engineering, Campinas State University (UNICAMP), Phone: +55 19 3788 3702, Fax: +55 19 3289 1395, C.P. 6101, CEP: 13083-970, Campinas-SP, Brazil. (email: romano@decom.fee.unicamp.br)

Some works have considered an optimization criterion based on minimum BER [4-7]. However, they rely on a known sequence to minimize the criterion. So the following question arises: *when a training sequence is not available or desired is it possible to perform minimum BER filtering*? This paper aims to provide an answer to this question.

A probability density function (pdf) estimation-based blind criterion was proposed in [8]. Using a parametric model that matches the statistical characteristics of the transmitted signal, the equalizer is designed to minimize the divergence between the pdf of the equalized signal and such parametric model.

In this paper, we present a relationship that shows that it is not possible achieve minimum BER using the proposal in [8]. Using the maximum *a posteriori* (MAP) criterion, which minimizes the BER, we derive a relationship between the MMSE, MAP and the blind criteria proposed in [8]. This new result shows an important property of the blind filtering approach when minimum BER is required.

The rest of the paper is organized is follows. Section II describes the system model. The blind criterion is revisited in Section III and the relationship of the blind criterion and minimum BER approach is presented in Section IV. Finally, our conclusions are stated in Section V.

II. SYSTEM MODEL

The considered base-band system model is depicted in Figure 1.

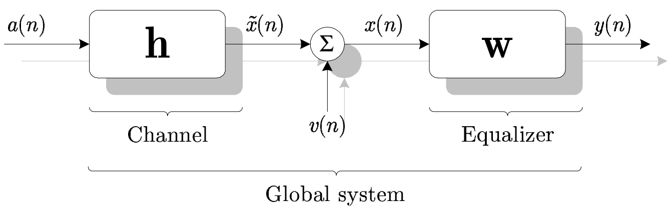

Fig. 1. Base-band system model.

The discrete transmitted sequence is represented by:

$$\mathbf{a}(n) = \begin{bmatrix} a(n) \\ \vdots \\ a(n-N-M+1) \end{bmatrix}, \tag{1}$$

where N and M are, respectively, the channel and equalizer lengths. It is assumed that the independent and identically distributed (i.i.d) symbols $a(n) \in \mathcal{A}$, which has cardinality S.

The channel is represented by a FIR filter given by

$$\mathbf{h} = \begin{bmatrix} h_0 \\ \vdots \\ h_{N-1} \end{bmatrix}. \tag{2}$$

The additive noise denoted in vectorial way by $\mathbf{v}(n) = \begin{bmatrix} v(n) & \cdots & v(n-M+1) \end{bmatrix}^T$ is white, Gaussian, uncorrelated from the transmitted sequence and has variance σ_v^2.

The equalizer, which has finite impulse response (FIR) denoted by

$$\mathbf{w}(n) = \begin{bmatrix} w(n) \\ \vdots \\ w(n-M+1) \end{bmatrix}, \tag{3}$$

is fed by the channels outputs $x(n) = \widetilde{x}(n) + v(n)$ where $\widetilde{x}(n) = \sum_{i=0}^{N-1} h_i a(n-i)$ are the noiseless channel outputs.

The equalizer output is denoted in vectorial representation by

$$y(n) = \mathbf{w}^T(n)\mathbf{x}(n), \tag{4}$$

where

$$\mathbf{x}(n) = \begin{bmatrix} x(n) \\ \vdots \\ x(n-M+1) \end{bmatrix}. \tag{5}$$

This model will be used in the rest of this paper.

III. Blind Criterion for PDF Estimation: A Review

In this section, we present the concepts and model of the criterion proposed in [8].

Let $\mathbf{w}_{\text{ideal}}$ be an ideal zero-forcing linear equalizer, the output of which can be written as

$$y(n) = \mathbf{w}_{\text{ideal}}^T \mathbf{x}(n), \tag{6}$$

where

$$\mathbf{x}(n) = \mathbf{H}\mathbf{a}(n) + \mathbf{v}(n) \tag{7}$$

and $\mathbf{H}$ is the $M \times (N + M - 1)$ *convolution matrix* of the channel [9].

Then, using Equation (7) in (6), it is possible to write:

$$\begin{aligned} y(n) &= (\mathbf{H}\mathbf{a}(n) + \mathbf{v}(n))^T \mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^T(n)\mathbf{H}^T\mathbf{w}_{\text{ideal}} + \mathbf{v}^T(n)\mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^T(n)\underbrace{\mathbf{H}^T\mathbf{w}_{\text{ideal}}}_{\mathbf{g}_{\text{ideal}}} + \mathbf{v}^T(n)\mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^T(n)\mathbf{g}_{\text{ideal}} + \vartheta(n) \\ &= a(n-\delta) + \vartheta(n), \end{aligned} \tag{8}$$

where $\mathbf{g}_{\text{ideal}}$ is the ideal system response, δ is a delay and $\vartheta(n)$ is a random variable (r.v.) assumed with independent Gaussian samples [9].

Equation (8) states that the pdf of the signal on the output of the equalizer is a mixture of equiprobable Gaussians (since the transmitted symbols are i.i.d.) given by:

$$p_{Y,\text{ideal}}(y) = \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \cdot \sum_{i=1}^{S} \exp\left[-\frac{|y(n) - a_i|^2}{2\sigma_\vartheta^2}\right] \cdot p(a_i), \tag{9}$$

where the a_i are the possible values of $a(n-\delta)$ that are also symbols of the transmitted alphabet $\mathcal{A}$.

Since the pdf of the equalized signal is known, we desire to construct a criterion that forces the adaptive filter to produce signals with the same (or similar) pdf than the ideal one. It is then interesting to use the well known measure of similarities between strictly positive functions (such as the pdfs), the *Kullback-Leibler Divergence* (KLD) [10].

In order to use the KLD, a parametric model, which is function of the filter parameters, to provide pdf estimation it is constructed [9]. A natural choice is the same model of mixture of Gaussians like the one in Equation (9). Then

$$\Phi(y, \sigma_r^2) = A \cdot \sum_{i=1}^{S} \exp\left(-\frac{|y(n) - a_i|^2}{2\sigma_r^2}\right) \cdot p(a_i), \tag{10}$$

is the chosen parametric model, where σ_r^2 is the variance of each Gaussian in the model and where $A = \frac{1}{\sqrt{2\pi\sigma_r^2}}$. In pattern classification field these kind of parametric functions, which are used to measure similarities against other functions, are called *target functions* [9].

Then, applying KLD to compare Equations (9) and (10) yields:

$$\begin{aligned} D_{p(y)||\Phi(y,\sigma_r^2)} &= \int_{-\infty}^{\infty} p(y) \cdot \ln\left(\frac{p(y)}{\Phi\left(y, \sigma_r^2\right)}\right) dy \\ &= \int_{-\infty}^{\infty} p(y) \cdot \ln\left(p(y)\right) dy - \int_{-\infty}^{\infty} p(y) \cdot \ln\left(\Phi(y, \sigma_r^2)\right) dy, \end{aligned} \tag{11}$$

where $p(y) = p_{Y,\text{ideal}}(y)$.

Minimizing (11) is equivalent to minimizing only the $\Phi\left(y, \sigma_r^2\right)$-dependent term, that is:

$$\begin{aligned} J_{\text{FPC}}(\mathbf{w}) &= -E\left\{\ln\left[\Phi\left(y, \sigma_r^2\right)\right]\right\} \\ &= -E\left\{\ln\left[A \cdot \sum_{i=1}^{S} \exp\left(-\frac{|y(n) - a_i|^2}{2\sigma_r^2}\right)\right]\right\}. \end{aligned} \tag{12}$$

The ***Fitting pdf*** **(FP)** criterion corresponds to minimizing $J_{\text{FPC}}(\mathbf{w})$. Furthermore, it is known that minimizing Equation (12) corresponds to finding the entropy of y if $\Phi\left(y, \sigma_r^2\right) = p_{Y,\text{ideal}}(y)$ [11, p. 59].

A stochastic algorithm for filter adaptation is given by:

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) - \mu_w \nabla J_{\text{FPC}}(\mathbf{w}) \\ \nabla J_{\text{FPC}}(\mathbf{w}(n)) &= \frac{\sum_{i=1}^{S} \exp\left(-\frac{|y(n)-a_i|^2}{2\sigma_r^2}\right)(y(n) - a_i^*)}{\sigma_r^2 \cdot \sum_{i=1}^{S} \exp\left(-\frac{|y(n)-a_i|^2}{2\sigma_r^2}\right)}\mathbf{x}, \end{aligned} \tag{13}$$

where μ_w is the step size.

The adaptive algorithm which uses the proposed criterion will be called ***Fitting pdf Algorithm*** **(FPA)**. Equation (13) shows an important property of the algorithm: it takes into account the phase of the transmitted symbols.

The computational complexity of this algorithm is proportional to the computation of S exponentials which are required by Equation (13). Thus, its complexity is a little higher than other LMS-like algorithms.

Another important point is thats, although the ideal equalizer is known to have infinity length, the use of the FP criterion does not require a long filter to compensate the channel effect. It has been observed, through simulations, that the length of the equalizer for this criterion has the same order of other blind criteria.

• *The parameter* σ_r^2*:*

As shown in the previous section, the parametric model used to update the filter coefficients is also σ_r^2-dependent. This parameter plays an important role since it is the variance of each Gaussian in the parametric model.

Moreover, σ_r^2 is also important for the convergence rate because it modifies the effective step size, that is, $\mu_{\text{eff}} = \frac{\mu_w}{\sigma_r^2}$. In the classification field this parameter is similar to the *temperature* one in annealing processes [9].

A numerical problem that arises with the use of the FPA is the nonconvergence for very small values of σ_r^2. This is due to the Gaussians being very sharp and much more difficult to fit the data on them. This model have also been considered in [12], where the ideal pdf of the received signal is assumed to be a mixture of impulses and later a Gaussian mixture model is considered in order to make the assumption more realistic and feasible.

IV. Minimum BER: Supervised and Blind Criteria

In order to allow the analysis of a minimum BER criterion, we consider the MAP one.

The MAP criterion aims to maximize the probability of recovering a symbol a_i given that y has been observed in the equalizer output. Then, MAP criterion can written as [13]:

$$J_{\text{MAP}}(\mathbf{w}) = E\left\{\ln\left[p\left(a_i\left|y\right.\right)\right]\right\}, \tag{14}$$

where we are considering the logarithm in Equation (14) in order to simplify computations [14].

Let us write the *a posteriori* probability density functions using the Bayes' rule as [13]:

$$p\left(a_i\left|y\right.\right) = \frac{p\left(y\left|a_i\right.\right)\cdot p\left(a_i\right)}{\sum\limits_{i=1}^{S} p\left(y\left|a_i\right.\right)\cdot p\left(a_i\right)}. \tag{15}$$

Here, we have to change the way we use to present the FPC. The parametric model $\Phi(y)$ represents the sum of probabilities of a possible transmitted signal a_i given that y has been observed. Since there is no knowledge about the transmitted symbol itself, $\Phi(y)$ is then the sum of all conditional probabilities of the received signal y given the transmission of symbol a_i. In other words, we can write:

$$\Phi(y) = \sum_{i=1}^{S} p(y|a_i). \tag{16}$$

Of course, if we assume that the signal is corrupted by AWGN, we obtain Equation (10). Besides, we can use the gradient of J_{FPC} without the stochastic approximation in the form

$$\nabla J_{\text{FPC}}(\mathbf{w}) = -E_Y\left\{\frac{E_{\mathcal{A}}\left\{A\cdot\exp\left(-\frac{|y-a|^2}{\sigma_r^2}\right)[y-a^*]\right\}}{\sigma_r^2\cdot E_{\mathcal{A}}\left\{A\cdot\exp\left(-\frac{|y-a|^2}{\sigma_r^2}\right)\right\}}\mathbf{x}\right\}, \tag{17}$$

where E_Y and $E_{\mathcal{A}}$ stand for expectation with respect to the variables Y and $\mathcal{A}$, respectively.

As in [15], we can define an auxiliary function given by

$$\psi(y,a) = \frac{A\cdot\exp\left(-\frac{|y-a|^2}{\sigma_r^2}\right)}{E_{\mathcal{A}}\left\{A\cdot\exp\left(-\frac{|y-a|^2}{\sigma_r^2}\right)\right\}}, \tag{18}$$

that measures "how sure" is the decision of symbol a since only y has been observed and the signal has a conditional pdf given as a Gaussian.

Then, comparing Equations (15) and (18), we can observe that considering the Gaussian model for the conditional pdf we have the same measure [16].

Using such consideration and supposing that σ_r^2 is chosen appropriately, the MAP criterion can be written as [16]:

$$\begin{gathered} E\left\{\ln\left[p\left(a_i\left|y\right.\right)\right]\right\} = E\left\{\ln\left[\frac{p\left(y\left|a_i\right.\right)\cdot p\left(a_i\right)}{\Phi(y)}\right]\right\} \\ J_{\text{MAP}} = E\left\{\ln\left[p\left(y\left|a_i\right.\right)\right]\right\}\underbrace{-E\left\{\ln\left[\Phi(y)\right]\right\}}_{J_{\text{FPC}}}. \end{gathered} \tag{19}$$

It is worth mentioning that the conditional probability $p\left(y\left|a_i\right.\right)$ concerns the assumed model to the signal at the output of the equalizer and we are also assuming an ideally equalized signal in presence of additive Gaussian noise. We then have:

$$p\left(y\left|a_i\right.\right) = \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}}\exp\left(-\frac{|y-a_i|^2}{2\sigma_\vartheta^2}\right). \tag{20}$$

Therefore, we can rewrite Equation (19) as

$$\begin{gathered} J_{\text{MAP}} = -\frac{1}{2\sigma_{\vartheta^2}}E\left\{|y-a_i|^2\right\} + \ln\left[\frac{1}{\sqrt{2\pi\sigma_\vartheta^2}}\right] + J_{\text{FPC}} \\ J_{\text{FPC}} - J_{\text{MAP}} = \frac{1}{2\sigma_{\vartheta^2}}E\left\{|y-a_i|^2\right\} - \ln\left[\frac{1}{\sqrt{2\pi\sigma_\vartheta^2}}\right]. \end{gathered} \tag{21}$$

Now, we need to explore the right side of Equation (21) in order to provide an appropriate relationship.

Equation (16), as we stated before, corresponds to the sum of all probabilities of symbol a_i given the observation y.

This is due to the blind processing, when there is no information about the transmitted symbol in a given time instant. In the case of supervised processing, the transmitted symbol is known at each time instant and there is no need of computing the contribution of all symbols from alphabet $\mathcal{A}$. Thus, the parametric model for the supervised case is given by [16]:

$$\Phi(y) = p(y(n)|a(n)). \tag{22}$$

Considering Equation (22), and also assuming the ideally recovered signal immersed in AWGN, that is, the conditional probability as given by Equation (20), we can write $J_{\text{FPC}}(\mathbf{w})$ in Equation (12) for the supervised case as:

$$\begin{aligned} J_{\text{FPC}}(\mathbf{w}) &= -E\left\{\ln\left[A \cdot \exp\left(-\frac{|y(n)-a(n)|^2}{2\sigma_r^2}\right)\right]\right\} \\ &= \frac{1}{2\sigma_r^2}E\left\{|y(n)-a(n)|^2\right\} - \ln[A]. \end{aligned} \quad (23)$$

Clearly, the cost function in Equation (23) is the MMSE cost function up to scaling and translation effects. However, the optimization of Equation (23) with respect to $\mathbf{w}$ provides the same solution than the classical MMSE cost function given as [1]:

$$J_{\text{MMSE}}(\mathbf{w}) = E\left\{|y(n)-a(n)|^2\right\}.$$

Therefore, the cost function in Equation (23) will be denoted J_{MMSE} as it stands for the MMSE in the supervised case.

Observing Equation (21) and comparing the right side with Equation (23) we can see that it is the same. Thus, the following relationship can be given considering $\sigma_r^2 = \sigma_v^2$ [16]:

$$J_{\text{MAP}} = J_{\text{FPC}} - J_{\text{MMSE}}. \quad (24)$$

Equation (24) provides an important issue about relationships of blind and supervised criteria for minimum BER filtering using the FPC criterion. It shows that when there is no knowledge about the transmitted signal, the FPC does not achieve minimum BER. So, it is not possible to perform minimum BER filtering with this criterion without the knowledge of the transmitted sequence. Further, since the criteria are defined as positive functions we can also write the following inequality for the FPC and MAP:

$$J_{\text{FPC}} \geq J_{\text{MAP}}, \quad (25)$$

showing that achieving the minimum for J_{FPC} does not necessarily imply achieve minimum BER.

V. Conclusions and Perspectives

In this paper we have presented a relationship between supervised and unsupervised criteria aiming a minimum bit error rate filtering.

The unsupervised criteria is based on the approach of estimation of the probability density function of the signal on the equalizer output using a parametric model. The Kullback-Leibler divergence is used to minimize the divergence of the equalizer output pdf and the parametric model.

This criterion presents some interesting properties that are based on the evaluation of the conditional probabilities of received signal over all possible transmitted ones. As a result, the presented approach allows to achieve a relationship between the supervised and unsupervised criteria.

The obtained relationship states that minimum BER is not attained with the blind criteria because it requires the instantaneous knowledge of the transmitted signal. Thus, such relationship is given in terms of the minimum mean square error and FPC criteria for achieving maximum *a posteriori* probabilities.

The main perspective for future works is the investigation and proposal of a semi-blind criterion that can possibly minimize BER and some other metric (e.g. Kullback-Leibler one) with a good compromise of computational and information complexity.

Acknowledgements

The authors would like to thank Dr. Renato R. Lopes by his useful comments that improved a lot writing of the paper.

References

[1] S. Haykin, *Adaptive Filter Theory*, 3rd ed., ser. (Prentice-Hall Information and System Sciences Series). Prentice-Hall, 1996.

[2] Z. Ding and Y. G. Li, *Blind Equalization and Identification*, ser. (Signal Processing and Communications Series). New York, USA: Marcel Dekker, 2001.

[3] J.-M. Brossier, *Signal et Communication Numérique: égalisation et synchronisation*, ser. (Collection Traitement du Signal). HERMES, 1997.

[4] C.-C. Yeh and J. R. Barry, "Adaptive Minimum Bit-Error Rate Equalization for Binary Signaling," *IEEE Transactions on Communications*, vol. 48, no. 7, pp. 1226–1235, July 2000.

[5] J. D. Laster, "Robust GMSK Demodulation Using Demodulator Diversity and BER Estimation," Ph.D. dissertation, Faculty of the Virginia Polytechnic Institute and State University, Blacksburg, Virginia, March 1997.

[6] K. A. Phillips, "Probability Density Function Estimation Applied to Minimum Bit Error Rate Adaptive Filtering," Master's thesis, Virginia Polytechnic Institut and State University, May 1999.

[7] R. C. de Lamare and R. Sampaio-Neto, "Adaptive Channel Equalisation Using Minimum BER Gradient-Newton Algorithms," in *Proceedings of IEEE Int. Symposium on Telecommunications (ITS 2002)*, Natal, RN - Brazil, September 2002.

[8] C. C. Cavalcante, F. R. P. Cavalcanti, and J. C. M. Mota, "A PDF Estimation-Based Blind Criterion for Adaptive Equalization," in *Proceedings of IEEE Int. Symposium on Telecommunications (ITS 2002)*, Natal, Brazil, September 2002.

[9] C. C. Cavalcante, "Neural Prediction and Probability Density Function Estimation Applied to Blind Equalization," Master's thesis, Federal University of Ceará (UFC), Fortaleza-CE, Brasil, February 2001.

[10] S. Haykin, Ed., *Unsupervised Adaptive Filtering*, ser. (Series on Adaptive and Learning Systems for Signal Processing, Communications and Control). John Wiley & Sons, 2000, vol. II: Blind Deconvolution.

[11] C. M. Bishop, *Neural Networks for Pattern Recognition*. UK: Oxford University Press, 1995.

[12] R. Amara, "Égalisation de Canaux Linéaires et Non Linéaires, Approche Bayésienne," Ph.D. dissertation, Université Paris XI Orsay, France, Novembre 2001.

[13] J. G. Proakis, *Digital Communications*, 3rd ed., ser. (Electrical Engineering Series). McGraw-Hill, 1995.

[14] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*. Prentice Hall International, 1987.

[15] J. Sala-Alvarez and G. Vázquez-Grau, "Statistical Reference Criteria for Adaptive Signal Processing in Digital Communications," *IEEE Transactions on Signal Processing*, vol. Vol. 45, no. No. 1, pp. 14–31, January 1997.

[16] C. C. Cavalcante, "On Blind Source Separation: Proposals and Analysis of Multi-User Processing Strategies," Ph.D. dissertation, State University of Campinas (UNICAMP) – DECOM, Campinas, SP - Brazil, April 2004.