

Analysis of ITL Criteria in the Context of FIR Channel Equalization

Levy Boccato, Denis G. Fantinato, Daniel. G. Silva, Rafael Ferrari, Aline Neves, Romis Attux

Abstract—In this work, we perform an analysis, in the context of channel equalization, of two criteria that can be considered central to the field of information theoretic learning (ITL): the minimum error entropy criterion (MEEC) and the maximum correntropy criterion (MCC). An original derivation of the exact cost function of these criteria in the scenario of interest is provided and used to analyze their robustness and efficiency from a number of relevant standpoints. Another important feature of the paper is an study of the estimated versions of these cost functions, which raises several aspects regarding parameters of the canonical Parzen window estimator. The study is carried out for distinct channel and noise models, both in the combined response and parameter spaces, and also employs as benchmarks crucial metrics like the probability of bit error. The conclusions indicate under what conditions ITL criteria are particularly reliable and a number of factors that can lead to suboptimal performance.

Index Terms—Filtering, Channel equalization, Information-theoretic learning, Error entropy, Correntropy

I. INTRODUCTION

EQUALIZATION was one of the first applications of adaptive filtering [1], and still is a domain of great interest due to its importance in communication systems, audio and image signal processing, seismology, among others. The classical problem of channel equalization for digital communications consists of using an adaptive filter at the receiver to cope with two factors that affect the

transmission of data through band-limited channels: intersymbol interference (ISI) and additive noise. The three main aspects involved in solving this problem are the choices of the filter structure, of the adaptation criterion, which translates the desired objective into a mathematical function that depends on the filter parameters, and of the strategy to adapt such parameters.

With respect to the filter structure, an important evolution can be seen, starting from very simple linear transversal filters and moving towards to the use of modern intelligent structures such as neural networks [2], [3], [4], [5] and fuzzy filters [6], [7], [8], having as motivation the desire of reaching the optimum performance of the Bayesian equalizer [9].

On the other hand, parameter adaptation according to a specific criterion corresponds to an optimization problem. The first proposed approaches consisted in assuming that the transmitted sequence is known at the receiver. This scenario, known as supervised equalization, allows for the use of an error signal, obtained comparing the filter output with the desired sequence, to adapt the filter coefficients. The zero-forcing condition and Wiener theory [10], based on the classical mean squared error (MSE) criterion, were the pioneer efforts in solving such a problem. When the desired sequence is not available, blind or unsupervised equalization methods are necessary. In this case, only statistical characteristics of the transmitted signal are known at the receiver and the adaptation criteria usually depend on higher-order statistics of the involved signals.

In the last 15 years, a new class of optimization criteria has emerged: instead of using a limited number of statistical moments of the involved signals, the idea was to explore more information about them, especially by working directly with their probability density functions (PDFs). Thus, concepts like entropy and mutual information were

L. Boccato (Corresponding Author), D. G. Fantinato, R. Ferrari and R. Attux are with the School of Electrical and Computer Engineering (FEEC), University of Campinas, SP, Brazil (email: {lboccato,denisgf,attux}@dca.fee.unicamp.br; rferari@decom.fee.unicamp.br)

D. G. Silva is with Department of Electrical Engineering (ENE), FT, University of Brasília, Brasília, DF, Brazil (email: danielgs@ene.unb.br)

A. Neves is with Engineering, Modeling and Applied Social Science Center (CECS), Federal University of ABC (UFABC), Santo André, SP, Brazil (email: aline.neves@ufabc.edu.br)

Digital Object Identifier: 10.14209/jcis.2016.1.

borrowed from Information Theory, leading to the so-called field of *Information Theoretic Learning* (ITL) [11]. In order to achieve practical and efficient methods in such framework, Rényi's definition of entropy, associated with the Parzen window estimation and kernel methods, has been the preferred approach.

Such methods have been recently applied to a variety of problems related to supervised and unsupervised signal processing, such as linear and nonlinear filter adaptation, prediction, feature extraction, equalization and source separation, presenting good results when compared to classical methods under certain scenarios [11]. In the context of supervised equalization, which will be the focus of this work, two criteria have deserved special attention: the Minimum Error Entropy Criterion (MEEC), which tends to eliminate as much uncertainty as possible from the error signal using Rényi's quadratic entropy; and the Maximum Correntropy Criterion (MCC), which seeks to increase the similarity between the desired signal and the adaptive system output through the use of correntropy, a measure directly related to entropy that generalizes the concept of correlation.

Even though good results have been obtained, the preliminary analysis presented in [12] showed that certain precautions are necessary. Such work used the ITL framework together with Extreme Learning Machines (ELM) and, when compared to Wiener solutions (MSE) in certain scenarios, concluded that both provided similar receivers. This intriguing result led the authors to ponder that a more thorough and profound analysis of ITL criteria was necessary in order to elucidate in which cases their use would present advantages with respect to the well-known MSE criterion.

Thus, in this paper, we perform an analytical study of the MEEC and MCC by deriving the exact theoretical formulas of the Rényi's quadratic error entropy and of the correntropy. Such derivations are carried out considering two different noise models: the additive white Gaussian noise (AWGN), which represents the classical model in the context of digital communications [13], and the additive white impulsive noise (AWIN), since it has been explored in important works of the ITL literature [14], [11], characterizing a scenario in which the ITL criteria have shown to offer performance improvements in certain cases. Additionally, the equalizer adopted

in this work consists of a linear finite impulse response (FIR) filter, as it constitutes the classical equalization structure, for which a solid theory has been developed [10], and also due to its mathematical tractability, which is essential for enabling the derivation of the signal PDFs.

The proposed theoretical development provides answers to important questions that hitherto, to the best knowledge of the authors, have not yet been fully addressed in the literature, like: are these criteria suitable to solve the problem of supervised channel equalization, that is, do their minima really present good solutions in terms of reducing ISI or minimizing the probability of errors at the adaptive equalizer output? In addition to that, do these criteria really present advantages with respect to the classical MSE criterion and Wiener solutions? If yes, in what context?

These questions will be treated in this paper by means of an extensive analysis of the criteria theoretical surfaces, which are obtained using the derived exact formulas. Moreover, this development allows the comparison of the exact criteria response with those obtained with the use of Parzen window estimation of the underlying PDFs, a procedure that is always necessary so that efficient algorithms be achieved.

This work is organized as follows: Section II presents the formal definition of the ITL criteria considered in this work, viz., the minimum error entropy and the maximum correntropy criteria, along with the corresponding nonparametric estimators based on kernel functions. Then, Section III begins with the description of the mathematical model of the supervised channel equalization problem, laying the foundations for the derivation of the theoretical expressions of the error entropy and the correntropy for both of the aforementioned noise models. The characteristics of the theoretical ITL criteria are analyzed in detail in Section IV, considering different equalization scenarios with respect to the channel and to the adopted noise, and compared with those associated with the corresponding kernel-based estimators. Additionally, the optimal solutions offered by the ITL criteria are identified and contrasted with the Wiener solution in terms of the error probability each equalizer can achieve, being the performance evaluation based on exact expressions of the error probabilities, which are also derived in this work and shown in A. Finally, concluding remarks and

future perspectives are summarized in Section V.

II. INFORMATION THEORETIC LEARNING

Second-Order Statistics (SOS), such as correlation, variance and the well-known mean squared error are widely employed as optimization criteria associated with several learning algorithms. Notwithstanding, since the initial key theoretical works on unsupervised signal processing, from the 80's until nowadays, new criteria considering higher-order statistical information have been developed in order to obtain more appropriate characteristics to cope with increasingly complex problems.

In this context, ITL [15] comprises the development of adaptation criteria based on concepts that are borrowed from Information Theory. The term ITL was consolidated by Principe and collaborators (although its origin may be associated with [16]), which proposed the adoption of Renyi's quadratic entropy as the key information measure [17]:

$$H_2(X) = -\log \left(\int_{-\infty}^{\infty} f_X^2(x) dx \right), \quad (1)$$

where $f_X(\cdot)$ is the PDF of the random variable X . Although Shannon's classical definition can also be employed within the ITL framework, the quadratic entropy presents the appealing property that it is possible, in association with the Parzen window method for PDF estimation, to define a continuous, differentiable estimator which enables the subsequent development of gradient-based learning algorithms [11].

Assuming that (i) the set $\{x_1, x_2, \dots, x_T\}$ with T independent and identically distributed (i.i.d.) observations of X is available and that (ii) a Gaussian kernel $G_{\sigma_{\text{ker}}^2}(x) = \frac{1}{\sqrt{2\pi\sigma_{\text{ker}}^2}} \exp\left(-\frac{x^2}{2\sigma_{\text{ker}}^2}\right)$ is adopted as the window function for PDF estimation, the standard quadratic entropy estimator is given as follows [11]:

$$\hat{H}_2(X) = -\log \left(\frac{1}{T^2} \sum_{j=1}^T \sum_{k=1}^T G_{2\sigma_{\text{ker}}^2}(x_j - x_k) \right), \quad (2)$$

where σ_{ker}^2 is an important (and sensitive) parameter of the estimator, denoted as the kernel width or size. Since the $\log(\cdot)$ function is monotonic, its usual to consider optimization of (2) only with respect to its argument, which is termed Information Potential [18].

A second important ITL measure is correntropy [19]. It is a similarity metric which "generalizes" correlation, containing second and higher-order moments which are expressed by the kernel used in its definition. If a Gaussian kernel is chosen, the correntropy between two random variables X and Y is defined by

$$\begin{aligned} c(X; Y) &= \mathbb{E}\{G_{\sigma_{\text{ker}}^2}(X - Y)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_{\sigma_{\text{ker}}^2}(x - y) f_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} G_{\sigma_{\text{ker}}^2}(e) f_E(E = e) de, \end{aligned} \quad (3)$$

where $\mathbb{E}(\cdot)$ denotes the statistical expectation operator and $E = X - Y$ is called the error signal. Clearly, in practical adaptive algorithms, the error PDF is not known in advance and just a set $\{e_1, e_2, \dots, e_T\}$ of data is available, leading to the sample estimator of correntropy:

$$\hat{c}(E) = \frac{1}{T} \sum_{i=1}^T G_{\sigma_{\text{ker}}^2}(e_i). \quad (4)$$

After presenting the important definitions of quadratic entropy and correntropy, including the respective canonical estimators, we are ready to introduce the two supervised criteria to be analyzed in this work. Firstly, the minimum error entropy criterion (MEEC):

$$\min_w H_2(E) \quad \text{s.t.} \quad \begin{cases} e(n) = d(n) - f(\mathbf{r}(n), \mathbf{w}), \\ \mathbb{E}\{e\} = 0, \end{cases} \quad (5)$$

where $d(n)$ is the desired signal and $f(\mathbf{r}(n), \mathbf{w})$ is the adaptive system with its respective parameter vector \mathbf{w} . The MEEC has the goal of eliminating as much uncertainty as possible from the error signal, ideally transforming the error PDF into a delta function. When applying the criterion to a learning algorithm, the analytical expression of entropy is typically replaced by the estimator described in (2).

The second criterion is the maximum correntropy criterion (MCC), defined by

$$\max_w c(E) \quad \text{s.t.} \quad e(n) = d(n) - f(\mathbf{r}(n), \mathbf{w}). \quad (6)$$

The principle of MCC is that the maximization of correntropy increases the similarity between the desired signal and the adaptive system output, i.e., the error values become smaller, which correspond

to higher values of the Gaussian kernel, finally leading to an error PDF more concentrated at the origin.

These criteria have found consistent application in the context of several supervised tasks, such as filtering, system identification and classification [11]. Additionally, the ITL methods have attracted significant interest in nonlinear and non-Gaussian framework [17]. Notwithstanding, as we intend to show, there are some aspects that require further clarification, even in the linear case. In that sense, we begin to study them in the specific problem of linear equalization, as described in the following.

III. THEORETICAL DERIVATION OF ITL CRITERIA

In this section, we formally derive the exact expressions associated with the MEEC and the MCC. After defining the signal models explored in the channel equalization problem, we present a general procedure to obtain the PDF of the error signal as a function of the coefficients of the equalizer and of the parameters that describe the scenario, which shall serve as the basis for determining the Rényi's quadratic entropy of the error signal, as well as the correntropy function considering two types of noise models, viz., the AWGN and an AWIN.

A. Mathematical Model of the Channel Equalization Problem

The diagram depicted in Figure 1 indicates the main elements involved in the supervised channel equalization model adopted in this work.

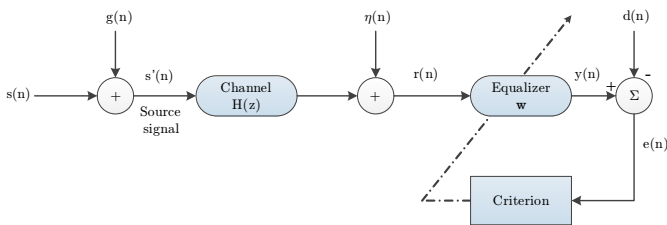


Fig. 1. Block diagram of the supervised channel equalization problem.

The source signal $s'(n)$ corresponds to an i.i.d. discrete-time random process whose samples result from the addition of a Gaussian noise with zero mean and variance σ_g^2 , represented by $g(n)$, to symbols $s(n)$ taken from the binary alphabet

$\{+1, -1\}$ (BPSK or 2-PAM modulation). Hence, $s'(n) \sim \frac{1}{2}N(-1, \sigma_g^2) + \frac{1}{2}N(+1, \sigma_g^2)$, where $N(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2 . This relatively unusual source model is particularly interesting for two main reasons: (i) the BPSK modulation can be seen as a particular case that is obtained by forcing σ_g^2 to zero, and (ii) it will be useful in the analysis of the behavior of the surface associated with the error entropy criterion, as shall be verified in Section IV.

The transfer function $H(z)$ models the intersymbol interference (ISI) effect of the band-limited channel used in the transmission and is given by $H(z) = \sum_{i=0}^{D-1} h_i z^{-i}$, where $h_i, i = 0, \dots, D-1$, are the coefficients of the impulse response of the (causal) channel with length D .

The transmitted signal is also corrupted by noise, which is represented by an i.i.d random process $\eta(n)$ whose PDF is denoted by $f_\eta(\eta)$, with zero mean and variance σ_η^2 , which leads to a signal-to-noise ratio (SNR), in decibels, given by:

$$\text{SNR} = 10 \log \frac{\mathbb{E}\{s'^2(n)\}}{\sigma_\eta^2}. \quad (7)$$

The device known as equalizer is responsible for estimating the transmitted symbols, an estimation delay being acceptable, based on a collection of samples of the received signal $r(n)$, which can be written as $r(n) = s'(n) * h(n) + \eta(n)$, where $*$ denotes the convolution operation. The structure of the equalizer is of paramount importance, since it establishes an inexorable limit to the attainable equalization performance [20], and affects the complexity of the adaptation process and of the practical implementation.

The classical setup consists of a FIR filter in the role of the equalizer. In this case, the output of the equalizer can be expressed as $y(n) = \mathbf{w}^T \mathbf{r}(n)$, where $\mathbf{r}(n) = [r(n) \ r(n-1) \ \dots \ r(n-K+1)]^T$ is the vector containing the K most recent samples of the received signal and $\mathbf{w} = [w_0 \ \dots \ w_{K-1}]^T$ specifies the coefficients of the equalizer.

Using the convolution matrix of the channel, $\mathbf{H} \in \mathbb{R}^{K \times K+D-1}$, given by

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \dots & h_{D-1} & 0 & \dots \\ 0 & h_0 & \dots & h_{D-2} & h_{D-1} & \dots \\ \vdots & & \ddots & & & \vdots \end{bmatrix}, \quad (8)$$

and the vectors $\mathbf{s}(n) = [s(n) \ \dots \ s(n-K-D+2)]^T$

and $\mathbf{g}(n) = [g(n) \dots g(n - K - D + 2)]$, it is possible to express $\mathbf{r}(n)$ as:

$$\mathbf{r}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{H}\mathbf{g}(n) + \boldsymbol{\eta}(n). \quad (9)$$

Thus, by using (9), the equalizer output is given by:

$$y(n) = \mathbf{w}^T \mathbf{H}\mathbf{s}(n) + \mathbf{w}^T \mathbf{H}\mathbf{g}(n) + \mathbf{w}^T \boldsymbol{\eta}(n). \quad (10)$$

Finally, since the objective is to recover the transmitted information contained in $s'(n-d)$, where d represents the equalization delay, the error signal is given by $e(n) = y(n) - s'(n-d)$.

B. PDF of the Error Signal

The keystone for the derivation of the theoretical expressions of the error entropy and correntropy is the probability density function of the error signal $e(n)$. The strategy we shall adopt for analytically obtaining this PDF consists in, firstly, determining the conditional PDFs of the equalizer output $y(n)$ given that $s(n-d)$ is $+1$ or -1 , and, then, using the law of total probability [21] to find the desired PDF. In this process, it is convenient to interpret the terms related to $\mathbf{g}(n)$ as additional noise elements that can be properly combined with the channel noise $\boldsymbol{\eta}(n)$. Therefore, we define

$$\boldsymbol{\eta}'(n) = \mathbf{w}^T \mathbf{H}\mathbf{g}(n) + \mathbf{w}^T \boldsymbol{\eta}(n) \quad (11)$$

as the term resulting from the filtering process of both the source and channel noises, so that the equalizer output is reduced to

$$y(n) = \mathbf{w}^T \mathbf{H}\mathbf{s}(n) + \boldsymbol{\eta}'(n). \quad (12)$$

Assuming that the transmitted sequence $\mathbf{s}(n)$ of BPSK symbols is known and that the coefficients of the equalizer are fixed, the conditional PDF of $y(n)$ is given by:

$$f_Y(Y = y | \mathbf{s}(n)) = f_{\boldsymbol{\eta}'}(y - \mathbf{w}^T \mathbf{H}\mathbf{s}(n)), \quad (13)$$

where $f_{\boldsymbol{\eta}'}$ denotes the PDF of the $\boldsymbol{\eta}'(n)$, defined in (11).

Actually, it is necessary to express the conditional PDF of $y(n)$ given $s(n-d)$. There are $N_S = 2^{K+D-1}$ possible sequences $\mathbf{s}(n)$ that can be transmitted: half of these sequences contain the symbol $s(n-d) = +1$, while the other half contain $s(n-d) = -1$. Hence, considering that $s(n-d) = +1$, the probability of observing a particular sequence $\mathbf{s}(n)$ is $P(\mathbf{s}(n) | s(n-d) = +1) = \frac{1}{2^{K+D-2}}$,

so that, based on the law of total probability, it is possible to write

$$f_Y(y | s(n-d) = +1) = \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{f_{\boldsymbol{\eta}'}(y - \mathbf{w}^T \mathbf{H}\mathbf{s}_i(n))}{2^{K+D-2}}, \quad (14)$$

where $\mathcal{S}^{\mathcal{D}^+}$ denotes the set of sequences $\mathbf{s}(n)$ whose symbol $s(n-d)$ is equal to $+1$.

Using the same procedure, it is possible to conclude that

$$f_Y(y | s(n-d) = -1) = \sum_{i \in \mathcal{S}^{\mathcal{D}^-}} \frac{f_{\boldsymbol{\eta}'}(y - \mathbf{w}^T \mathbf{H}\mathbf{s}_i(n))}{2^{K+D-2}}, \quad (15)$$

where $\mathcal{S}^{\mathcal{D}^-}$ denotes the set of sequences $\mathbf{s}(n)$ whose symbol $s(n-d)$ is equal to -1 .

Having these two conditional PDFs at hand, the PDF of $y(n)$ can be determined with the aid of the law of total probability:

$$\begin{aligned} f_Y(Y = y) &= \frac{1}{2} \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{f_{\boldsymbol{\eta}'}(y - \mathbf{w}^T \mathbf{H}\mathbf{s}_i(n))}{2^{K+D-2}} \\ &\quad + \frac{1}{2} \sum_{i \in \mathcal{S}^{\mathcal{D}^-}} \frac{f_{\boldsymbol{\eta}'}(y - \mathbf{w}^T \mathbf{H}\mathbf{s}_i(n))}{2^{K+D-2}} \\ &= \sum_{i=0}^{N_S-1} \frac{f_{\boldsymbol{\eta}'}(y - \mathbf{w}^T \mathbf{H}\mathbf{s}_i(n))}{2^{K+D-1}}. \end{aligned} \quad (16)$$

In order to obtain the PDF of the error signal $e(n) = y(n) - s'(n-d) = y(n) - [s(n-d) + g(n-d)]$, it is necessary to analyze the statistical properties of $y'(n) = y(n) - g(n-d)$. Using (10), we can express $y'(n)$ as follows:

$$\begin{aligned} y(n) - g(n-d) &= \mathbf{w}^T \mathbf{H}\mathbf{s}(n) + \mathbf{w}^T \mathbf{H}\mathbf{g}(n) \\ &\quad + \mathbf{w}^T \boldsymbol{\eta}(n) - g(n-d) \\ &= \mathbf{w}^T \mathbf{H}\mathbf{s}(n) + \mathbf{w}^T \boldsymbol{\eta}(n) \\ &\quad + (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T) \mathbf{g}(n), \end{aligned} \quad (17)$$

where $\boldsymbol{\gamma} \in \mathbb{R}^{K+D-1 \times 1}$, and $\gamma_i = \begin{cases} 1, & \text{if } i = d \\ 0, & \text{otherwise} \end{cases}$.

Assuming that $s(n-d)$ is known, the conditional PDF of $y'(n)$ is determined by the PDF of the variable $\Psi = \mathbf{w}^T \boldsymbol{\eta}(n) + (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T) \mathbf{g}(n)$. Thus, the conditional PDFs of $y'(n)$ given that $s(n-d) = +1$ and $s(n-d) = -1$ can be expressed as:

$$f_{Y'}(y' | s(n-d) = +1) = \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{f_{\Psi}(y' - \mathbf{w}^T \mathbf{H}\mathbf{s}_i(n))}{2^{K+D-2}} \quad (18)$$

and

$$f_{Y'}(y'|s(n-d) = -1) = \sum_{i \in \mathcal{S}^{\mathcal{D}^-}} \frac{f_{\Psi}(y' - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n))}{2^{K+D-2}}, \quad (19)$$

respectively.

Since $e(n) = y'(n) - s(n-d)$, when $s(n-d) = +1$, $e(n) = y'(n) - 1$. Hence, the PDF of $e(n)$ conditioned to the knowledge that $s(n-d) = +1$ is the conditional PDF of $y'(n)$ evaluated at the point $e(n) + 1$:

$$f_E(e|s(n-d) = +1) = f_{Y'}(e+1|s(n-d) = +1). \quad (20)$$

Analogously,

$$f_E(e|s(n-d) = -1) = f_{Y'}(e-1|s(n-d) = -1). \quad (21)$$

Finally, the PDF of the error signal can be written as:

$$f_E(E = e) = \frac{1}{2} f_{Y'}(e+1|s(n-d) = +1) + \frac{1}{2} f_{Y'}(e-1|s(n-d) = -1). \quad (22)$$

By replacing (18) and (19) in (22) and defining

$$\beta_i = \begin{cases} +1, & \text{if } i \in \mathcal{S}^{\mathcal{D}^+} \\ -1, & \text{if } i \in \mathcal{S}^{\mathcal{D}^-} \end{cases}, \quad (23)$$

the PDF of the error signal is given by:

$$f_E(e) = \frac{1}{2^{K+D-1}} \sum_{i=0}^{N_S-1} f_{\Psi}(e + \beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)). \quad (24)$$

C. Error Entropy

In this section, we shall determine the theoretical expressions for the error entropy considering the AWGN and the AWIN noise models.

1) *Gaussian Noise*: As demonstrated in Section III-B, in order to obtain the exact PDF of the error signal, it is necessary to identify the PDF of the variable $\Psi = \mathbf{w}^T \boldsymbol{\eta}(n) + (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T) \mathbf{g}(n)$. In the AWGN case, we consider that $\boldsymbol{\eta}(n) \sim N(0, \sigma_{\eta}^2)$.

It can be shown that the linear combination of independent Gaussian random variables gives rise to another Gaussian random variable [21]. By using this property, we can conclude that the two terms that compose Ψ are Gaussian random variables with distributions $N(0, \sigma_{\eta_f}^2)$ and $N(0, \sigma_{g_f}^2)$, respectively,

where $\sigma_{\eta_f}^2 = \mathbf{w}^T \sigma_{\eta}^2 \mathbf{w}$ and $\sigma_{g_f}^2 = \sigma_g^2 (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T) (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T)^T$. Additionally, since Ψ is also a linear combination of Gaussian random variables, $\Psi \sim N(0, \sigma_{\eta_f}^2 + \sigma_{g_f}^2)$.

Therefore, the PDF of the error signal for the AWGN case can be written as:

$$f_E(e) = \sum_{i=0}^{N_S-1} \frac{G_{\sigma_{\Psi}^2}(e + \beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n))}{2^{K+D-1}}, \quad (25)$$

where $G_{\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ and $\sigma_{\Psi}^2 = \sigma_{\eta_f}^2 + \sigma_{g_f}^2$.

Having obtained the error PDF, it is possible to express the Rényi's quadratic entropy associated with the error signal, defined in (1), as follows:

$$H_2(E) = -\log \left(\int_{-\infty}^{\infty} \left[\sum_{i=0}^{N_S-1} \frac{G_{\sigma_{\Psi}^2}(e + \beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n))}{2^{K+D-1}} \right]^2 de \right). \quad (26)$$

The term inside the integral can be written as the product of two equivalent sums indexed by i and j . Then, the multiplicative constants can be placed outside the integral, as well as both sums, yielding the following expression:

$$H_2(E) = -\log \left(\left(\frac{1}{2^{K+D-1}} \right)^2 \sum_{i=0}^{N_S-1} \sum_{j=0}^{N_S-1} \int_{-\infty}^{\infty} G_{\sigma_{\Psi}^2}(e + \beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)) \times G_{\sigma_{\Psi}^2}(e + \beta_j - \mathbf{w}^T \mathbf{H} \mathbf{s}_j(n)) de \right). \quad (27)$$

The expression of $H_2(e)$ in (27) can be simplified by exploring a property of Gaussian functions [11]:

$$\int_{-\infty}^{\infty} G_{\sigma_i^2}(x - x_i) G_{\sigma_j^2}(x - x_j) = G_{\sigma_i^2 + \sigma_j^2}(x_i - x_j). \quad (28)$$

Hence, the error entropy can be expressed as:

$$H_2(E) = -\log \left(\left(\frac{1}{2^{K+D-1}} \right)^2 \sum_{i=0}^{N_S-1} \sum_{j=0}^{N_S-1} G_{2\sigma_{\Psi}^2}(\beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n) - \beta_j + \mathbf{w}^T \mathbf{H} \mathbf{s}_j(n)) \right) \quad (29)$$

By defining $\nu_i = \beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)$, we obtain the final expression for the Rényi's quadratic entropy of the error signal in the AWGN case:

$$H_2(E) = -\log \left(\left(\frac{1}{2^{K+D-1}} \right)^2 \sum_{i=0}^{N_S-1} \sum_{j=0}^{N_S-1} G_{\sigma_{2\Psi}^2}(\nu_i - \nu_j) \right). \quad (30)$$

2) *Impulsive Noise*: The second type of noise we shall consider can be described by the following PDF [11]:

$$f_{\eta}(\eta) = p G_{\alpha^2 \sigma_1^2}(\eta) + (1 - p) G_{\alpha^2 \sigma_2^2}(\eta), \quad (31)$$

where p represents the probability of taking a sample according to a Gaussian distribution with zero mean and a small variance σ_1^2 , $(1 - p)$ is the probability of the noise sample being generated according to a Gaussian distribution with zero mean and a large variance σ_2^2 , and $\alpha = \sqrt{\frac{\sigma_{\eta}^2}{p\sigma_1^2 + (1-p)\sigma_2^2}}$ is a scaling factor used to ensure that the overall noise variance is σ_{η}^2 , leading to a SNR as given by (7). The shape of the noise PDF becomes an approximation of an impulse by adopting $p \gg (1 - p)$ and $\sigma_1^2 \ll \sigma_2^2$.

The derivation of the theoretical expressions for the MEEC and the MCC requires the exact PDF of the error signal. As discussed in Section III-B, the error PDF is determined by the statistical properties of $\Psi = \mathbf{w}^T \boldsymbol{\eta}(n) + (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T) \mathbf{g}(n)$, with the difference that, now, the first term of Ψ consists of a linear combination of AWIN samples.

It can be shown that the PDF of the sum of independent random variables is obtained by means of the convolution between the PDFs of the original variables [21]. This property can be employed to determine the PDF of the error signal. However, due to the characteristics of the noise model, it is not viable to derive a generic expression for $f_E(e)$, as occurred with the AWGN case, since the number of elements that result from the required convolutions significantly grows as the number of coefficients of the equalizer is increased. Therefore, we shall restrict the derivation of the expressions to the case in which the equalizer has only two coefficients ($K = 2$).

Let $\mathbf{w} = [w_0 \ w_1]^T$ be the vector containing the coefficients of the equalizer. Thus, the PDF of the first term of Ψ , viz., $\eta_f(n) = \mathbf{w}^T \boldsymbol{\eta}(n)$, is obtained via the convolution between the PDFs of $w_0 \eta(n)$ and $w_1 \eta(n - 1)$. Due to the linear character of the convolution, the Gaussian functions present in (31) are convolved with themselves, so that the PDF of $\eta_f(n)$ contains four terms resulting from the convolutions. Moreover, since the convolution between Gaussian functions also yields a Gaussian function whose mean and variance are the sums of the individual means and variances, respectively, the PDF of the filtered noise $\eta_f(n) = \mathbf{w}^T \boldsymbol{\eta}(n)$ is given

by:

$$\begin{aligned} f_{\eta_f}(\eta_f) &= p^2 G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_1^2}(\eta_f) \\ &+ p(1 - p) G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2}(\eta_f) \\ &+ (1 - p)p G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2}(\eta_f) \\ &+ (1 - p)^2 G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_2^2}(\eta_f). \end{aligned} \quad (32)$$

The second term of Ψ , viz., $(\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T) \mathbf{g}(n)$, is associated with the distribution $N(0, \sigma_{g_f}^2)$, where $\sigma_{g_f}^2 = \sigma_g^2 (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T) (\mathbf{w}^T \mathbf{H} - \boldsymbol{\gamma}^T)^T$, as indicated in Section III-C1. Hence, the PDF of Ψ can be written as follows:

$$\begin{aligned} f_{\Psi}(\psi) &= f_{\eta_f}(\eta_f) * G_{\sigma_{g_f}^2}(g_f) \\ &= p^2 G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_1^2 + \sigma_{g_f}^2}(\psi) \\ &+ p(1 - p) G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2 + \sigma_{g_f}^2}(\psi) \\ &+ (1 - p)p G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2 + \sigma_{g_f}^2}(\psi) \\ &+ (1 - p)^2 G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_2^2 + \sigma_{g_f}^2}(\psi), \end{aligned} \quad (33)$$

where $*$ denotes the convolution operator.

Then, using (24), the PDF of the error signal in the presence of impulsive noise can be expressed as follows:

$$\begin{aligned} f_E(E = e) &= \frac{1}{2^{K+D-1}} \sum_{i=0}^{N_S-1} \\ &\left(p^2 G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_1^2 + \sigma_{g_f}^2}(e + \nu_i) \right. \\ &+ p(1 - p) G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2 + \sigma_{g_f}^2}(e + \nu_i) \\ &+ (1 - p)p G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2 + \sigma_{g_f}^2}(e + \nu_i) \\ &\left. + (1 - p)^2 G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_2^2 + \sigma_{g_f}^2}(e + \nu_i) \right). \end{aligned} \quad (34)$$

The exact expression for the Rényi's quadratic entropy associated with the error signal can be obtained by following the same steps performed in Section III-C1, with the difference that, in the AWIN case, the integral of the square of $f_E(e)$ gives rise to 16 terms, each of them being the result of the integral of the product between two Gaussian functions, which can be determined with the aid of the property shown in (28). Hence, the error entropy

is given by:

$$\begin{aligned}
 H_2(e) = & -\log \left[\left(\frac{1}{2^{K+D-1}} \right)^2 \sum_{i=0}^{N_S-1} \sum_{j=0}^{N_S-1} \right. \\
 & \left(p^4 G_{2(\sigma_1^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_{g_f}^2)}(\nu_i - \nu_j) \right. \\
 & + p^3 (1-p) G_{2\sigma_1^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p^3 (1-p) G_{\sigma_1^2 \alpha^2 w_0^2 + 2\sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_0^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + (1-p)^2 p^2 G_{\sigma_1^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p^3 (1-p) G_{2\sigma_1^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & \left. + (1-p)^2 p^2 G_{2(\sigma_1^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_1^2 + \sigma_{g_f}^2)}(\nu_i - \nu_j) \right. \\
 & + (1-p)^2 p^2 G_{\sigma_1^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p(1-p)^3 G_{\sigma_1^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_0^2 + 2\sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p^3 (1-p) G_{\sigma_1^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_0^2 + 2\sigma_1^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p^2 (1-p)^2 G_{\sigma_2^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p^2 (1-p)^2 G_{2(\sigma_2^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_{g_f}^2)}(\nu_i - \nu_j) \\
 & + p(1-p)^3 G_{2\sigma_2^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p^2 (1-p)^2 G_{\sigma_1^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p(1-p)^3 G_{\sigma_1^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_0^2 + 2\sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & + p(1-p)^3 G_{2\sigma_2^2 \alpha^2 w_0^2 + \sigma_1^2 \alpha^2 w_1^2 + \sigma_2^2 \alpha^2 w_1^2 + 2\sigma_{g_f}^2}(\nu_i - \nu_j) \\
 & \left. + (1-p)^4 G_{2(\sigma_2^2 \alpha^2 w_0^2 + \sigma_2^2 \alpha^2 w_1^2 + \sigma_{g_f}^2)}(\nu_i - \nu_j) \right]. \quad (35)
 \end{aligned}$$

D. Correntropy

In the following, we present the mathematical derivation of the correntropy expressions considering both noise models.

1) *Gaussian Noise*: Based on (25), which provides the PDF of the error signal, the correntropy between the equalizer output signal and the desired signal $d(n) = s'(n-d)$, defined in (3), is given by the following expression:

$$\begin{aligned}
 c(y(n); d(n)) = & \frac{1}{2^{K+D-1}} \sum_{i=0}^{N_S-1} \int G_{\sigma_{\text{ker}}^2}(e) \\
 & \times \left[G_{\sigma_{\Psi}^2} \left(e + \beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n) \right) \right] de. \quad (36)
 \end{aligned}$$

By using the property presented in (28), we can obtain the final expression of the correntropy:

$$c(y(n); d(n)) = \frac{1}{2^{K+D-1}} \sum_{i=0}^{N_S-1} G_{\sigma_{\Psi}^2 + \sigma_{\text{ker}}^2}(\beta_i - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)). \quad (37)$$

2) *Impulsive Noise*: By repeating the same procedure of the previous section, using the error PDF defined in (34) as well the property in (28), the theoretical expression for correntropy in the AWIN case can be written as follows:

$$\begin{aligned}
 c(y(n); d(n)) = & \frac{1}{2^{K+D-1}} \sum_{i=0}^{N_S-1} \\
 & \left((1-p)^2 G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_2^2 + \sigma_{g_f}^2 + \sigma_{\text{ker}}^2}(\nu_i) \right. \\
 & + p(1-p) G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2 + \sigma_{g_f}^2 + \sigma_{\text{ker}}^2}(\nu_i) \\
 & + (1-p)p G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2 + \sigma_{g_f}^2 + \sigma_{\text{ker}}^2}(\nu_i) \\
 & \left. + p^2 G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_1^2 + \sigma_{g_f}^2 + \sigma_{\text{ker}}^2}(\nu_i) \right), \quad (38)
 \end{aligned}$$

3) *Discussion*: Once we have derived the exact expressions of the error entropy and correntropy for both the AWGN and the AWIN cases, it is interesting to contrast them with those associated with the corresponding kernel-based estimators shown in Section II, since we may obtain relevant elements concerning the expected behaviors of both the analytical and the estimated versions of the ITL criteria, which shall, then, be verified through the simulations carried out in different scenarios in Section IV.

An immediate difference to be remarked is related to the type of information that each version of the ITL criteria has access to: while the theoretical expressions of the error entropy and correntropy explore information about the set of possible transmitted sequences $\mathbf{s}_i(n)$, $i = 0, \dots, 2^{K+D-1} - 1$, and require knowledge of properties related to the noise and to the source, the estimators are based on a set of T observations of the error signal and use a single parameter, viz., the kernel size (σ_{ker}^2), whose value must be adequately selected [11].

By comparing the analytical error entropy related to the AWGN case, whose expression is given by (30), with the corresponding estimator using Gaussian kernel, defined in (2), it is possible to observe another important difference: in the former, the variances of the Gaussian functions are modified according to the coefficients of the equalizer (\mathbf{w}), since they affect $\sigma_{\eta_f}^2$ and $\sigma_{g_f}^2$, whereas, in the latter, all the Gaussian functions adopt the same variance value that is proportional to the chosen kernel size (σ_{ker}^2), regardless of the filter coefficients. A similar observation can be made with respect to the correntropy, although it is important to emphasize that the theoretical correntropy, defined in (37), also

includes a parameter σ_{ker}^2 .

On the other hand, in the AWIN case, the similarities between the theoretical and estimated versions of the error entropy and correntropy become less pronounced, since the theoretical entropy and correntropy explicitly take into account the PDF of the error signal, which, in turn, is directly influenced by the noise model, whereas the corresponding estimators are expressed as functions of the kernels, which, in this work, have been chosen to be Gaussian functions.

It is important to emphasize that we are focusing on the case of linear FIR filters as equalizers, whose mathematical simplicity allows the derivations of the exact cost functions of the MEEC and the MCC. Hence, the analysis and observations to be followed cannot be directly extended to nonlinear structures. Even though the ITL framework is recognized for their advantages in the nonlinear and/or non-Gaussian case, we highlight the relevance of this study as an useful tool – which shall consider not only the Gaussian but also the impulsive noise –, for a more detailed understanding of the ITL criteria.

Having these remarks in mind, we proceed to the detailed analysis of the theoretical ITL criteria considering different channel equalization scenarios.

IV. ANALYSIS

In this section, we shall study the characteristics of the theoretical error entropy and correntropy, as well as evaluate the performance of the optimum equalizers associated with these criteria and with the classical mean squared error. Additionally, we shall also discuss the effects of the main parameters of the problem, e.g., the signal-to-noise ratio, and of the criteria, such as the correntropy kernel size (σ_{ker}^2), on the surfaces associated with the theoretical ITL criteria, and, finally, contrast the observed behaviors with those associated with the error entropy and correntropy estimators.

A. The Combined Channel + Equalizer Impulse Response Analysis

In the problem of channel equalization, from an analytical standpoint, the qualitative evaluation of a criterion performance can be preliminarily done by considering a combined channel+equalizer impulse

response, which is a simpler but efficient methodology to identify the basic features and adequacy of a criterion to the problem in hand. Such analysis is known to be very useful as it is able to reveal the fundamental conditions necessary to achieve the desired solutions [22], [23], [24].

The main idea behind the combined channel+equalizer impulse response – or simply, combined impulse response – is to allow a simplification in the model of the channel and equalizer systems (by considering them as a single linear system) so as to facilitate the analysis of the criterion. Undoubtedly, the simplification of the communication model is an abstraction that implies in a reduction of the accuracy of the performed representation, as it ignores the existence of an attainable set of solutions, and, for this reason, this type of analysis is very clarifying but not complete. In this sense, we first present the analysis of the combined impulse response and, next, a more detailed study of the ITL criteria.

Mathematically, the combined channel+equalizer impulse response is represented by the convolution between the channel and the equalizer impulse responses:

$$c(n) = h(n) * w(n), \quad (39)$$

or, in terms of the Z-transform, $C(z) = H(z)W(z)$. In simple terms, (39) provides a single and simpler linear system to be analyzed. On the other hand, this simplification does not encompass the presence of noise $\eta(n)$, which must be disregarded from the original formulation.

Having this definition in view, additional considerations are necessary in the derivation of the error PDF shown in Section III-B. However, in order to avoid new conditions that make the analytical calculation of the ITL criteria impractical, we propose only slight modifications on the previous formulation, as shown in the following. First, the combined impulse response can be easily inserted in the analytical formulation by considering $\mathbf{c} = \mathbf{H}^T \mathbf{w}$ and, secondly, to remove the effect of the noise $\eta(n)$, one possible approach is that of defining its variance as $\sigma_{\eta}^2 = 0$. Indeed, this last assumption will cause the suppression of the noise effect, however, there are some limitations of the formulation. In the case in which the source is discrete, it is necessary that $\sigma_g^2 = 0$, and, considering the error entropy cost defined in (40), if $\sigma_g^2 = \sigma_{\eta}^2 = 0$, we know

that $H_2(E) = -\infty$. Hence, to circumvent this indeterminacy, we propose two alternatives: (i) to approximate $\sigma_\eta^2 \rightarrow 0$, by attributing to it a very small value – i.e., $\sigma_\eta^2 = \epsilon = 1 \times 10^{-10}$ – and adopting $\sigma_g^2 = 0$; or (ii) to adopt $\sigma_\eta^2 = 0$ and approximate $\sigma_g^2 \rightarrow 0$, by assuming $\sigma_g^2 = \epsilon$. We highlight that these two possibilities will be considered in the following analysis for the error entropy criterion. For continuous sources, as $\sigma_g^2 \neq 0$, we can simply assume $\sigma_\eta^2 = 0$. Regarding correntropy, (37), the cancellation of the noise effect can always be achieved with $\sigma_\eta^2 = 0$, because, even if the source is discrete, it suffices that σ_{ker}^2 be different from zero.

Finally, with these relatively simple modifications in the formulation of the PDF error considering the combined impulse response, we obtain valid definitions of the error entropy and the correntropy costs, respectively:

$$H_2(E) = -\log \left(\left(\frac{1}{2^{K+D-1}} \right)^2 \sum_{i=0}^{N_S-1} \sum_{j=0}^{N_S-1} G_{2\sigma_\Psi^2}(\hat{v}_i - \hat{v}_j) \right), \quad (40)$$

and

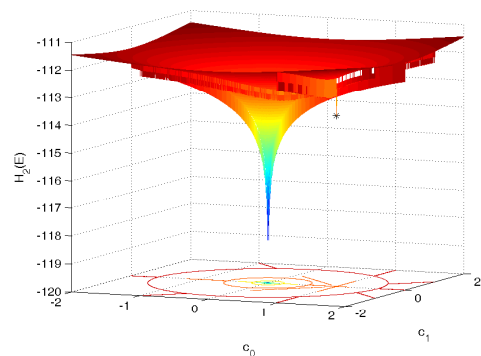
$$c(y(n); d(n)) = \frac{1}{2^{K+D-1}} \sum_{i=0}^{N_S-1} G_{\sigma_{g_f}^2 + \sigma_{\text{ker}}^2}(\beta_i - \mathbf{c}^T \mathbf{s}_i(n)), \quad (41)$$

where $\hat{v}_i = \beta_i - \mathbf{c}^T \mathbf{s}_i(n)$.

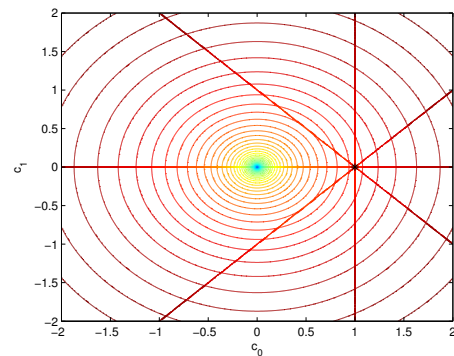
In order to obtain perfect channel equalization, it is necessary that the equalizer be the inverse of the channel, which implies in $C(z) = z^{-d}$, being d an arbitrary delay, a condition also referred to as *zero-forcing* (ZF) [10]. The adequacy of the criterion to the equalization problem demands that its optima meet the ZF condition. In this sense, we will perform our analysis of the entropy and correntropy criteria by obtaining their surfaces as a function of the combined impulse response, which, for the sake of visualization, will be assumed to be represented by a vector of two coefficients $\mathbf{c} = [c_0 \ c_1]^T$.

1) *Analytical Error Entropy*: Our study of the analytical entropy criterion for the combined channel+equalizer impulse response will encompass two representative cases regarding the source signal: the first one assumes a discrete BPSK transmitted source, while the second one assumes a continuous source. Furthermore, for both cases, we consider a reference signal with delay $d = 0$; therefore, it is expected, under the assumption of a complete adequacy to the problem of equalization, a single solution equal to $C(z) = 1$ or, in vector notation, $\mathbf{c} = [1 \ 0]^T$.

Starting with the classical BPSK source, we follow the first approach for the entropy cost in (40) by assuming $\sigma_g^2 = 0$ and $\sigma_\eta^2 = \epsilon$. Under these circumstances, we varied each of the coefficients of the combined response vector \mathbf{c} from -2 to $+2$, allowing us to obtain the surface of the analytical entropy cost and its contours in function of \mathbf{c} , as illustrated in Figure 2. To facilitate the visualization, we represented the desired solution $\mathbf{c} = [1 \ 0]^T$ as an asterisk (*) in the figure. From the figures, it



(a) Entropy Surface - $\sigma_g^2 = 0$



(b) Entropy Contour - $\sigma_g^2 = 0$

Fig. 2. Error entropy surface and contour for the discrete source as a function of the combined impulse response coefficients.

is possible to note that, for this case, the Rényi's entropy is a multimodal function, in which we highlight three relevant aspects to be discussed: (i) the expected solution $\mathbf{c} = [1 \ 0]^T$ is the second best minimum of the criterion, which indicates that entropy, besides the error minimization, also may lead to different (inadequate) solutions; (ii) there are four line segments (clearer in Figure 2(b)) that present lower values of measured entropy (but higher than the local minimum) that are not minima; and (iii) the global minimum is the trivial solution, i.e., $\mathbf{c} = [0 \ 0]^T$, which is not a desirable solution to

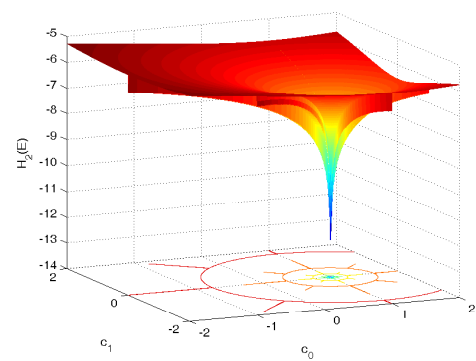
the problem of equalization.

We begin the discussion by jointly analyzing the aspects (i) and (ii). An evidence of the observed behavior of entropy for this case is linked to the resulting number of discrete states associated with the error signal $e(n)$. For instance, in the case (i), the combined channel+equalizer system $\mathbf{c} = [1 \ 0]^T$ works as a merely bypass, hence, the output signal $y(n)$ is equal to $s(n)$ and the error signal is $e(n) = y(n) - s(n) = 0$, which means that there is a single state for the error. It is important to mention that, as we assumed that $\sigma_\eta^2 \rightarrow 0$, the error signal presents a very small variance. In the case (ii), we have lower values of entropy that, together, form two diagonals, one horizontal and one vertical line segment. The vertical segment is related to a combined system of the type $\mathbf{c} = [1 \ c_1]^T$, which results in $e(n) = \pm c_1$, i.e., two possible error states; the horizontal line can be represented by the set of combined impulse response of the type $\mathbf{c} = [c_0 \ 0]^T$, which also results in two error states: $e(n) = \pm(c_0-1)$; for the two diagonals, in turn, we have that $\mathbf{c} = [c_0 \ c_0-1]^T$ and $\mathbf{c} = [c_0 \ -c_0+1]^T$, which both will imply in three possible values for $e(n)$, to say, 0, $-2c_0+2$ and $2c_0-1$. Again, it is important to emphasize that the error presents a small variance and is not precisely discrete. Additionally, Figure 2(a) also shows that the lower the number of discrete states for the error, the smaller will be the associated value of entropy. Nevertheless, slight variations of the entropy values can be observed along the discussed line segments, since their effects are combined with the concentric ellipses observed in Figure 2(b).

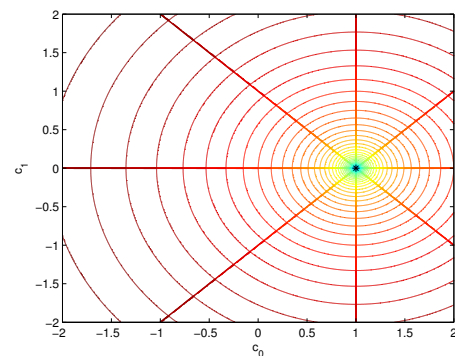
Finally, the case of the trivial solution (iii) is an important aspect that is very likely to happen in communications problems. Since the transmitted signal is usually modulated according to a set of discrete symbols, the source presents, consequently, a discrete distribution. However, the error entropy criterion is aimed to deal with continuous distributions and, when applied to the equalization problem, the interaction with discrete probabilities is unavoidable, possibly resulting in discontinuities in the cost function. More specifically, this is exactly the reason for the appearance of the minimum at the trivial solution, because when the coefficients of the combined impulse response \mathbf{c} are null, the error signal will be discrete with exact symbols -1 and $+1$, so that the error entropy becomes equal to the entropy of the source. From the point of

view of the presented formulation, this corresponds to setting both variances σ_g^2 and σ_η^2 to zero, and, from (40), $H_2(E) \rightarrow -\infty$ at the trivial solution, an indetermination.

For the second approach of the error entropy when dealing with a discrete source, i.e., with $\sigma_g^2 = \epsilon$ and $\sigma_\eta^2 = 0$, we obtained the error entropy surface and its contours, as exhibited in Figure 3. Now, for the trivial solution $\mathbf{c} = [0 \ 0]^T$, the resulting error signal $e(n)$ is constituted of two narrow Gaussians, centered in -1 and $+1$ (since $\sigma_g^2 \neq 0$), and we no longer observe the previous discontinuity. Hence, the error entropy criterion



(a) Entropy Surface - $\sigma_g^2 = \epsilon$



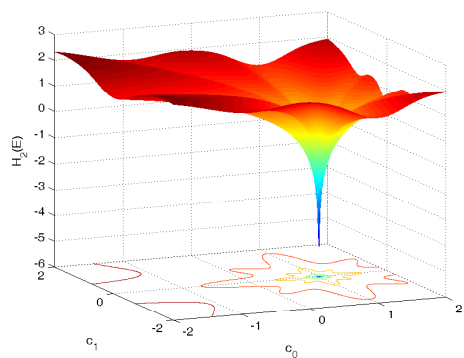
(b) Entropy Contour - $\sigma_g^2 = \epsilon$

Fig. 3. Error entropy surface and contour for an approximated discrete source.

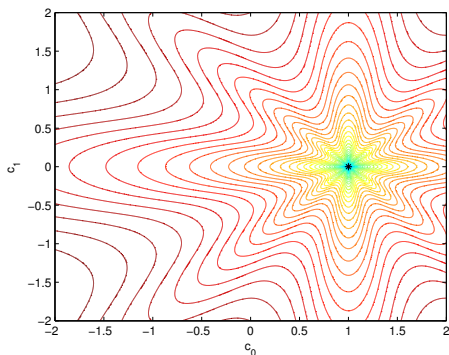
presents a single minimum at the desired solution $\mathbf{c} = [1 \ 0]^T$ and shows itself to be adequate to the problem of equalization. However, it is important to mention that the approach $\sigma_g^2 \rightarrow 0$ implies in a modification in the source signal, which is no longer strictly discrete. Rigorously, the condition $\sigma_g^2 \rightarrow 0$ is considered as a limit case of the continuous distribution, so that this assumption can be simply interpreted as an abstraction to help the understand-

ing of the trivial solution effect over the entropy criterion when dealing with purely discrete sources, which, as shown, presents some inadequacies.

In the properly continuous scenario, we consider the continuous source $s'(n)$ with $\sigma_g^2 = 0.05$ – here, σ_η^2 is equal to zero – in (40), and repeat the procedure to obtain the surface of the error entropy cost. The results are presented in Figure 4. In comparison with the previous case in which



(a) Entropy Surface - Continuous Source



(b) Entropy Contour - Continuous Source

Fig. 4. Error entropy surface and contour for the continuous source.

$\sigma_g^2 = \epsilon$, the result is similar, with a single minimum located at the desired combined channel+equalizer impulse response, and, instead of the line segments, there are smoothed curls.

Therefore, we can notice that the presence of noise terms, like σ_η^2 or σ_g^2 , can provoke modifications in the error entropy surface considering the combined impulse response. Indeed, when we assumed $\sigma_g^2 = \epsilon$, the noise is associated with the source (which becomes continuously distributed) and the undesirable trivial solution disappears. Interestingly, this idea is closely related to the deterministic kernel annealing [25], where the kernel size can be adjusted to reduce the number of minima in the

cost function. This topic will be further discussed ahead.

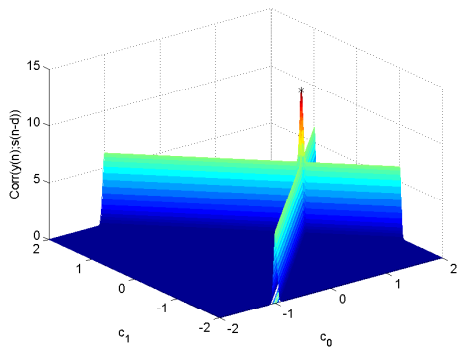
So far, the combined impulse response analysis for the error entropy criterion revealed some important features that, until the present moment, were not acknowledged via the usual estimator-based approaches. Motivated by these results, we will extend this analysis to the correntropy criterion.

2) *Analytical Correntropy*: The correntropy criterion is known to present some similarities with the error entropy criterion, but, in certain cases, these two entities can behave considerably differently [11]. In that sense, the analysis with respect to the combined channel+equalizer impulse response can provide us a clearer view of the correntropy criterion, revealing some points in common with the error entropy.

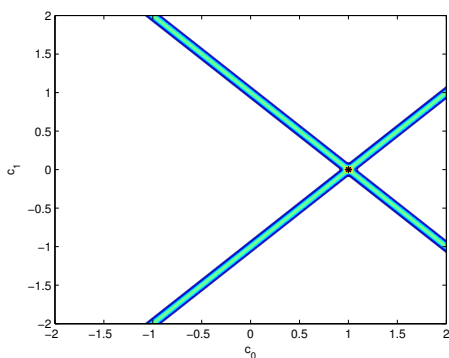
As in the previous case, we will consider the two types of source: the discrete and the continuous signal. Again, the reference signal will not be delayed – i.e., $d = 0$. According to (41), the parameters to be adjusted here are the variance of the Gaussian effect over the source signal, σ_g^2 , and the variance of the Gaussian kernel, σ_{ker}^2 .

In the first case, we consider the discrete source by adopting $\sigma_g^2 = 0$ and, for the variance of the Gaussian kernel, we assume $\sigma_{ker}^2 = 0.001$. It is important to note that, as a first difference from entropy, the presence of the σ_{ker}^2 guarantees that correntropy shall not diverge – as can be deduced from (41) – and no further assumptions are necessary. In order to obtain the correntropy surface, we vary the coefficients of the combined impulse response vector from -2 to $+2$. The results are displayed in Figure 5. Interestingly, correntropy presents a global maximum (since this is a maximization problem) at the desired combined impulsive response, i.e., $\mathbf{c} = [1 \ 0]^T$, which means that the criterion is adequate to the problem of equalization, even for discrete sources. Correntropy also presents the “line segments” observed in the entropy cost, but only the diagonal ones. However, differently from entropy, the value of correntropy is constant along the diagonal lines defined by the vectors $\mathbf{c} = [c_0 \ c_0 - 1]^T$ and $\mathbf{c} = [c_0 - c_0 + 1]^T$ – except in the neighborhood of $\mathbf{c} = [1 \ 0]^T$ –, which can cause a first-order search algorithm to get stuck at non-desired solutions.

In the case of the continuous source, we adopt $\sigma_g^2 = 0.05$ and $\sigma_{ker}^2 = 0.01$. The resulting surface of correntropy and the contours are shown in Figure



(a) Correntropy Surface - Discrete Source



(b) Correntropy Contour - Discrete Source

Fig. 5. Correntropy surface and contour for the discrete source.

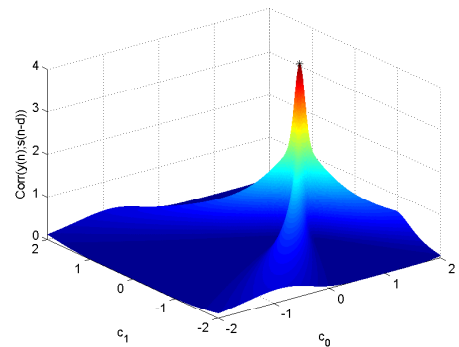
6. It is possible to see that correntropy keeps the maximum at $\mathbf{c} = [1 \ 0]^T$ and, additionally, the values of correntropy along the “line segments” are no longer constant, but incremental in the direction of the desired solution $\mathbf{c} = [1 \ 0]^T$.

It is important to mention that the variation of σ_{ker}^2 mainly causes changes in the smoothness of the correntropy surface, but the general shape of the cost remains the same.

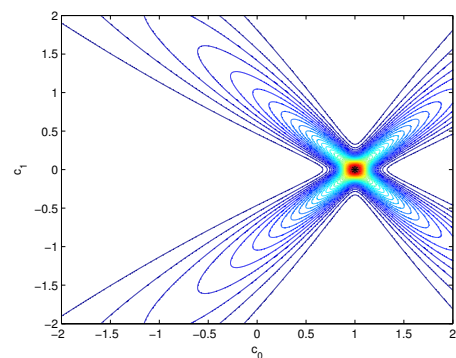
In general, we can affirm that correntropy, just like entropy, is more adequate for scenarios where the sources are associated to continuous distributions. Nevertheless, when dealing with purely discrete sources, correntropy seems to be a more robust criterion than entropy, since it does not suffer from the problem of the trivial solution.

B. First Specific Scenario: Minimum-Phase Channel and AWGN

The scenario we shall consider here is characterized by a minimum-phase channel whose transfer function is $H(z) = 1 + 0.6z^{-1}$. The noise present in



(a) Correntropy Surface - Continuous Source



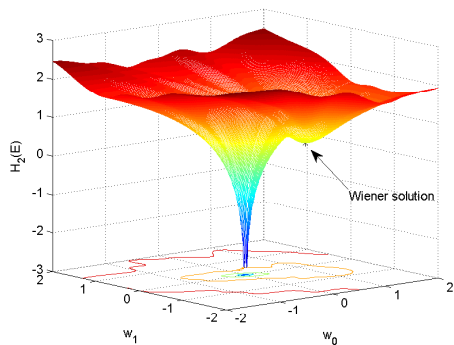
(b) Correntropy Contour - Continuous Source

Fig. 6. Correntropy surface and contour for the continuous source.

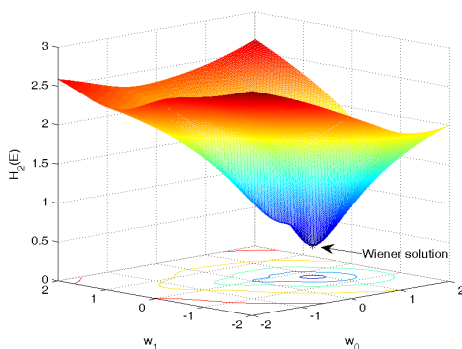
the transmission is AWGN and the desired equalization delay is $d = 0$. In order to visualize the surfaces associated with the ITL criteria, the equalizer is a FIR filter with two coefficients, i.e., $\mathbf{w} = [w_0 \ w_1]^T$.

1) *Error Entropy*: Initially, we display in Figure 7 the surfaces of the theoretical error entropy as a function of the equalizer coefficients considering an SNR of 13 dB and two particular conditions regarding the source model: (1) $\sigma_g^2 = 0$, which means that $s'(n)$ is a discrete source associated with the BPSK modulation, and (2) $\sigma_g^2 = 0.15$. Additionally, we also show the position of the corresponding Wiener solutions with an asterisk (*).

Some interesting remarks can be drawn from Figure 7. Firstly, it is possible to observe that the optimum solution is different depending on the source nature. When the source signal consists of symbols belonging to the BPSK modulation, the trivial solution, $\mathbf{w} = [0 \ 0]^T$ offers the minimum value for the error entropy. In light of the analysis of the combined channel + equalizer impulse response performed in Section IV-A, this undesirable phenomenon was expected. However, when the source



(a) Discrete source



(b) Continuous Source

Fig. 7. Surfaces of error entropy considering the first scenario with $SNR = 13$ dB for both discrete and continuous sources.

becomes continuous through the addition of the Gaussian component $g(n)$, the optimum located at the origin tends to disappear. In fact, as σ_g^2 increases from zero, the values of the error entropy at the origin progressively increase, until the local optimum at the vicinity of the Wiener solution becomes the solution of minimum entropy, as occurs in Figure 7(b).

Secondly, we can notice in Figure 7 that the surface of the theoretical error entropy is usually multimodal. This fact becomes more evident if we observe the contour of the error entropy exhibited in Figure 8 for the discrete source. Excluding the optimum located at the origin, there are, at least, three local optima.

Interestingly, one of these optima is relatively close to the Wiener solution, which indicates that the filter coefficients that minimize the mean squared error also offer, to a certain extent, an attractive solution in terms of the error entropy. However, depending on the SNR value and/or the noise model, this remark may not be valid. In order to elucidate this question, we present in Figure 9 the

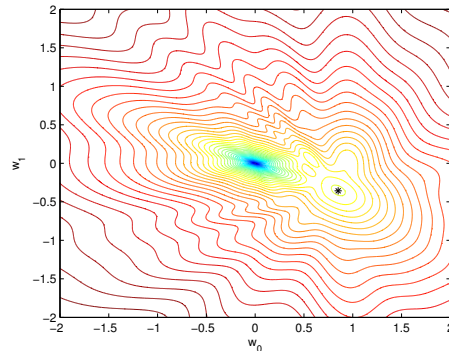


Fig. 8. Contour of error entropy considering the first scenario with $SNR = 13$ dB for the discrete source.

value of the gradient of the error entropy calculated at the Wiener solution as a function of the signal-to-noise ratio.

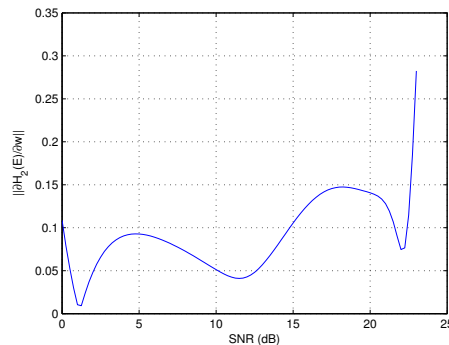
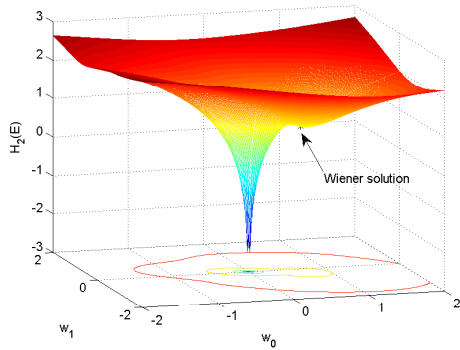


Fig. 9. Gradient of the theoretical error entropy evaluated at the Wiener solution as a function of the SNR .

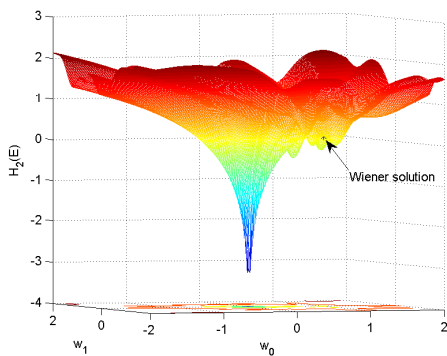
As we can observe, the magnitude of the gradient of the error entropy for the Wiener filter remains with relatively small values until the SNR of 15 dB, which suggests that the Wiener solution is, to a certain degree, near a local optimum of the error entropy. On the other hand, as the SNR increases, the tendency is that the magnitude of the gradient becomes higher, which means that the Wiener solution seems to be moving apart from locally optimal solutions in the sense of error entropy.

This evidence motivates an analysis of the impact that the SNR has on the surface of the error entropy. Thus, Figure 10 presents the surfaces of the theoretical error entropy considering the discrete source case for two different values of SNR : 5 and 20 dB, which are associated with the low and high SNR regimes.

It is possible to see in Figure 10 that the general



(a) $SNR = 5$ dB



(b) $SNR = 20$ dB

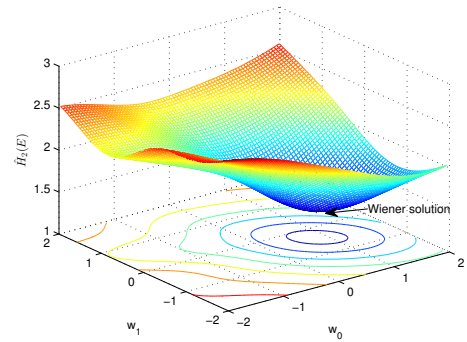
Fig. 10. Surfaces of error entropy for the SNR values of 5 dB and 20 dB considering the first scenario.

aspect of the surface of the error entropy is significantly modified by the SNR : when the SNR is 5 dB, the surface of the error entropy becomes smoother and there are only two visible local optima, viz., those located at the origin and near the Wiener solution; on the other hand, for the SNR of 20 dB, several other local optima emerge and the Wiener solution is not as close to a local optimum as in the previous cases, which is in accordance with the behavior indicated by the analysis of the gradient.

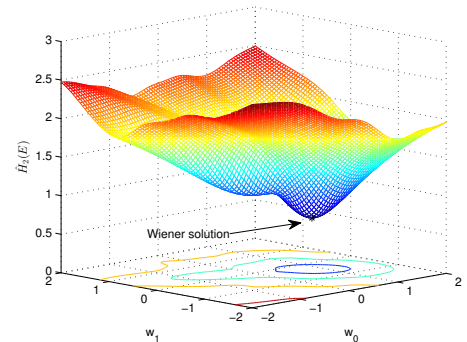
The observations pointed out so far evince that the MEEC presents a more complex behavior when compared with the MSE due to the potential occurrence of multiple optima in the corresponding entropy surface. Moreover, depending on the parameters of the scenario, such as the SNR , the optimum solution may or may not be close to the Wiener solution. The question as to whether the minimum error entropy solution yields a more efficient equalizer than that associated with the Wiener solution shall be addressed in Section IV-B3.

Since we have observed the main characteristics

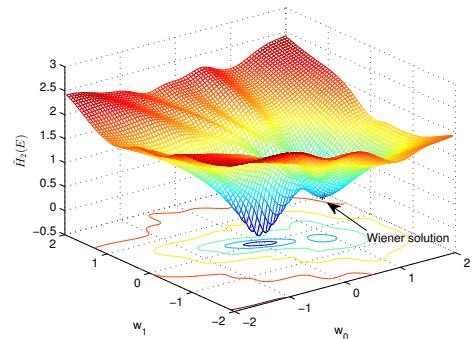
of the theoretical error entropy, we can now contrast them with those related to the entropy estimator, defined in (2). Considering the discrete source, which represents the practical case from the communications standpoint, and a set of 2500 samples of the error signal $e(n)$, we display in Figure 11 the surface of the estimated error entropy for the SNR of 13 dB and for three values of σ_{ker}^2 : 1, 0.15 and 0.01.



(a) $\sigma_{ker}^2 = 1$



(b) $\sigma_{ker}^2 = 0.15$



(c) $\sigma_{ker}^2 = 0.01$

Fig. 11. Surfaces of the error entropy estimator as a function of the kernel size considering the first scenario.

It is possible to observe in Figure 11 that the kernel size has a significant impact on the behavior

of the estimator. In fact, the surface shown in Figure 11(a) presents a smooth decay towards the point with minimum estimated entropy, which is close to the Wiener solution. Moreover, in accordance with the analysis of [11], the surface has approximately a quadratic character around the optimum solution. On the other hand, when the kernel size becomes smaller, the surface of the estimated error entropy may present multiple optima and relatively more abrupt variations. Additionally, the optimum solution is now located at the origin, as shown in Figure 11(c).

Interestingly, the behavior of the surface of the error entropy estimator as a function of the adopted kernel size is quite similar to that associated with the theoretical error entropy surface when we modify the variance of the continuous source component (σ_g^2). This similarity suggests that it may be possible to establish a connection between the kernel size employed by the entropy estimator and the characteristics of the source considered in the theoretical error entropy, particularly the variance of the continuous component (σ_g^2). In order to elucidate this connection, we display in Figure 12 the surface of the theoretical error entropy considering $\sigma_g^2 = 0.01$.

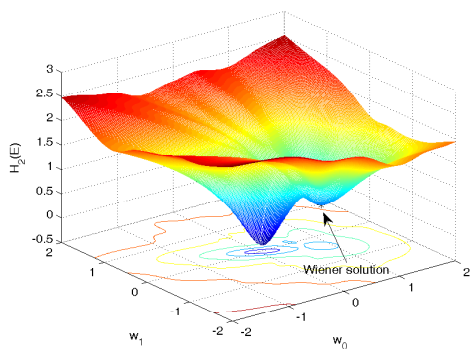


Fig. 12. Surface of the theoretical error entropy considering the first scenario with $SNR = 13$ dB and a continuous source with $\sigma_g^2 = 0.01$.

By comparing Figures 7(b) with 11(b), and 12 with 11(c), it is possible to notice the evident resemblance between the surfaces of the entropy estimator and the theoretical error entropy. Indeed, when larger values of σ_{ker}^2 are adopted in the entropy estimation, the corresponding surface mirrors the characteristics of the surface associated with the theoretical error entropy for a continuous source with a larger value of σ_g^2 . On the other hand, as σ_{ker}^2

decreases, the observed modifications in the surface of the estimated error entropy are quite similar to those verified for the theoretical error entropy as σ_g^2 becomes smaller, including the emergence of $\mathbf{w}^T = [0 \ 0]$ as the minimum entropy solution.

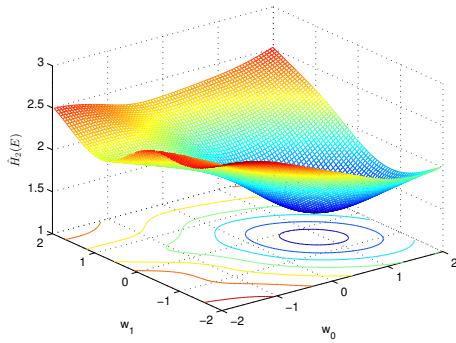
Therefore, based on these evidences, we can affirm that even though the kernel-based entropy estimator has access to error samples that are obtained in the transmission of a discrete source, its surface can be seen as an approximation of that associated with the theoretical error entropy for a continuous source. In this sense, the entropy estimator only approximates the theoretical error entropy for the same discrete source model in the limit when $\sigma_{ker}^2 \rightarrow 0$, which would ideally demand an infinite number of samples in order to achieve an adequate – unbiased and with minimum variance – estimation [11]. Nonetheless, the surface of the theoretical error entropy for the discrete source can still be used as a reference to the entropy estimator when σ_{ker}^2 is significantly reduced as it reveals the expected asymptotic behavior.

It is important to stress that the impact of the kernel size on the characteristics of the surface associated with the estimated error entropy has already been discussed in [11] in light of the concept of convolution smoothing [26], but, here, we have provided an alternative interpretation by establishing a connection with the analytical error entropy based on a continuous source model.

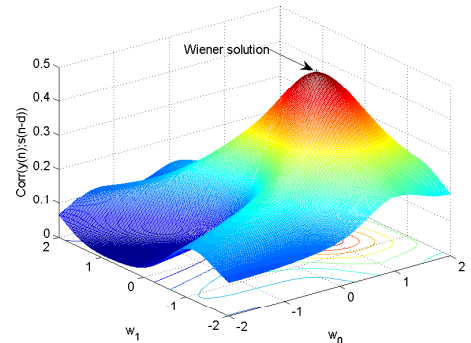
Finally, it is pertinent to mention that the surface of the error entropy estimator for a given value of σ_{ker}^2 is modified as the SNR varies in a manner that resembles the behavior of the theoretical entropy considering the correct source model: if σ_{ker}^2 is sufficiently small, the behavior of the estimator can be related to that of the theoretical entropy for the discrete source¹, while, if σ_{ker}^2 is large, the association is with the theoretical entropy for the continuous source with a large value of σ_g^2 . This tendency can be confirmed by Figure 13, which displays the surfaces of the error entropy estimator for the SNR of 20 dB considering the kernel sizes of 1 and 0.01, respectively.

Clearly, the surface of the estimated entropy is not significantly modified when $\sigma_{ker}^2 = 1$, a fact

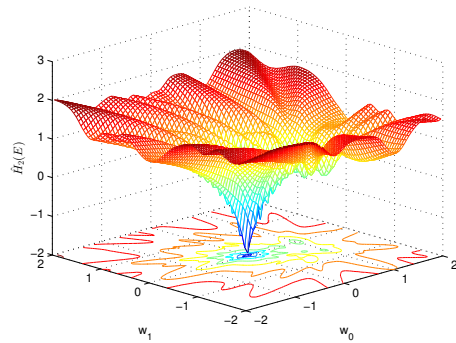
¹Ideally, this correspondence could only be established when $\sigma_{ker}^2 \rightarrow 0$, but, as aforementioned, the theoretical entropy for the discrete source still serves as an useful reference to analyze the behavior of the entropy estimator for small values of σ_{ker}^2 .



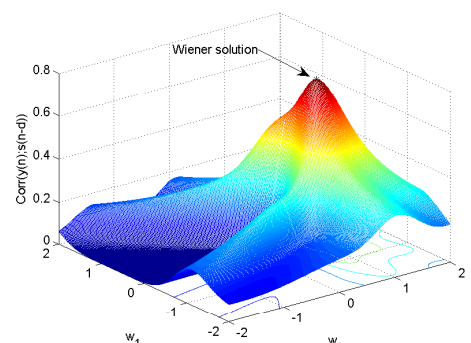
(a) $\sigma_{\text{ker}}^2 = 1$



(a) $\sigma_{\text{ker}}^2 = 0.5$



(b) $\sigma_{\text{ker}}^2 = 0.01$



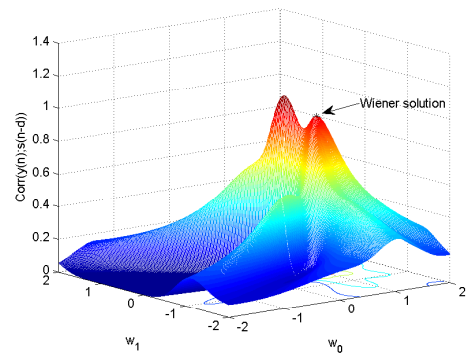
(b) $\sigma_{\text{ker}}^2 = 0.1$

Fig. 13. Surfaces of the error entropy estimator as a function of the kernel size considering an SNR of 20 dB.

that could also be observed by analyzing the surface of the theoretical error entropy for the continuous source case. On the other hand, when σ_{ker}^2 is small, we notice a behavior similar to that verified in Figure 10 in the context of the theoretical error entropy for a discrete source: the number of local optima is increased and the Wiener solution becomes more distant from the closest local optimum.

2) *Correntropy*: Now, we turn our attention to the MCC. The first question we shall analyze involves the impact of the kernel size (σ_{ker}^2), which is part of the definition of the correntropy measure, on the corresponding surface. Therefore, we present in Figure 14 the surfaces of the theoretical correntropy considering a SNR of 13 dB and three different kernel sizes: 0.5, 0.1 and 0.01, respectively. We shall restrict the analysis of the correntropy to the discrete source (BPSK modulation) case, since the continuous source model is not as useful to explain the effect of the kernel size used by the estimator as it has proved to be in the context of the error entropy.

As we can observe, for $\sigma_{\text{ker}}^2 = 0.5$, the corren-



(c) $\sigma_{\text{ker}}^2 = 0.01$

Fig. 14. Surfaces of the theoretical correntropy as a function of the kernel size considering an SNR of 13 dB. The asterisk (*) indicates the position of the Wiener solution.

trophy surface is relatively smooth and the optimum solution is quite similar to the Wiener solution, which is in accordance with the fact that correntropy converges to the MSE in the limit when $\sigma_{\text{ker}}^2 \rightarrow \infty$ [11].

On the other hand, when $\sigma_{\text{ker}}^2 = 0.01$, the surface presents another peak, which, indeed, offers the maximum value of correntropy. The emergence of this alternative optimum solution may be related to

the fact that the MCC boils down to maximizing the value of the error PDF at the origin, i.e., $f_E(e = 0)$, when $\sigma_{\text{ker}}^2 \rightarrow 0$ [11]. Thus, depending on the value of σ_{ker}^2 , the concept of correntropy may establish a different objective to the filter design and, ultimately, may lead to distinct solutions for the equalizer.

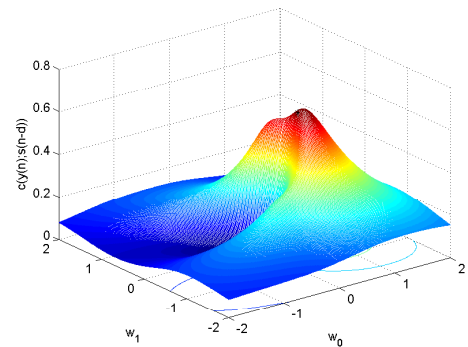
In this context, it is natural to wonder whether the alternative solutions that correntropy may offer actually represent better options for the equalizer when compared with the Wiener solution. This important question shall be addressed in Section IV-B3 taking into account the theoretical probability of error associated with the optimal solutions of the adaptation criteria discussed in this work.

The second aspect to be analyzed is related to the impact of the SNR on the surface of the theoretical correntropy. In order to elucidate this question, Figure 15 displays the correntropy surfaces, along with the corresponding contours, for the SNR values of 5 and 20 dB, considering that $\sigma_{\text{ker}}^2 = 0.01$. We verified that, for large values of σ_{ker}^2 , only slight modifications occur in the correntropy surface, so that the main characteristics observed in Figures 14(a) and 14(b) are preserved. So, for the sake of brevity, we decided to omit their presentation.

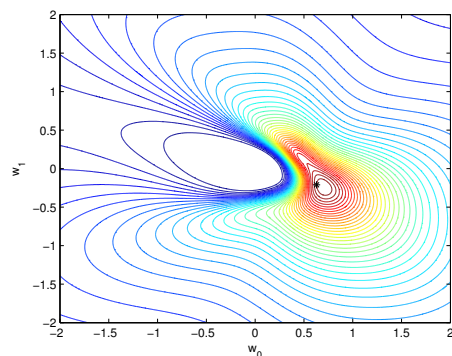
It is possible to notice in Figure 15 that the surface of the theoretical correntropy can be significantly modified due to the SNR : for low values of the SNR , the peak located near the Wiener solution is the preferred solution by the MCC, whereas, for high values of SNR , the other peak, which also appeared in Figure 14(c), is the optimum. This behavior is, to a certain extent, similar to that associated with the theoretical error entropy, since there is the possibility of dealing with relatively complex and multimodal surfaces depending on the SNR .

Finally, as performed in the context of the error entropy, it is pertinent to establish a comparison between the theoretical correntropy and the corresponding estimator, defined in (4). Thus, we exhibit in Figure 16 the surfaces of the estimated correntropy for three values of $\sigma_{\text{ker}}^2 - 0.25, 0.05$ and 0.01 -, considering the SNR of 13 dB. The values of the correntropy have been estimated using 2500 samples of the error signal.

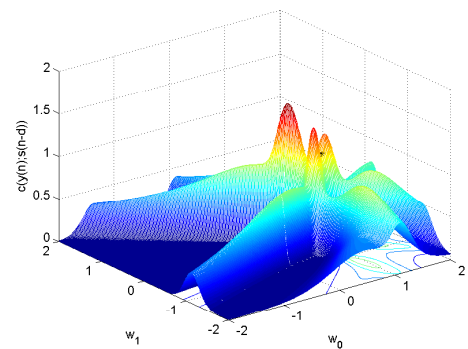
The behavior of the correntropy estimator as the kernel size is decreased follows the same pattern observed for the theoretical correntropy: for large



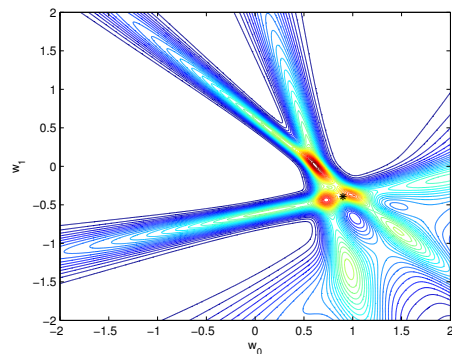
(a) $SNR = 5$ dB



(b) $SNR = 5$ dB



(c) $SNR = 20$ dB



(d) $SNR = 20$ dB

Fig. 15. Surfaces and contours of the theoretical correntropy as a function of the SNR considering $\sigma_{\text{ker}}^2 = 0.01$. The asterisk (*) indicates the position of the Wiener solution.

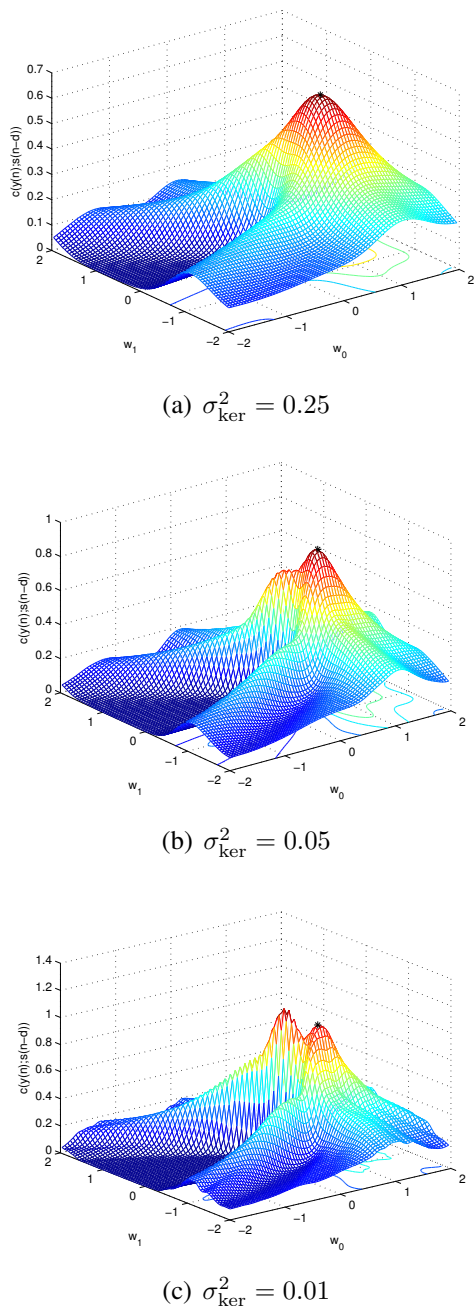


Fig. 16. Surfaces associated with the correntropy estimator as a function of the kernel size considering the first scenario.

values of σ_{ker}^2 , the surface is smoother and the global optimum is relatively close to the Wiener solution; on the other hand, for small values of σ_{ker}^2 , the surface presents additional local optima and a different peak provides the maximum value of correntropy.

3) *Probability of Error:* After analyzing the features of the surfaces associated with the theoretical error entropy and correntropy, as well with the corresponding estimators, we are now interested

in assessing the performance of the optimal solutions offered by the MEEC and the MCC in the equalization problem. The fundamental performance measure in the context of digital communications is the probability of error (also called error rate). In the Appendix, we formally derive the exact expressions of the probability of error for any FIR equalizer considering the AWGN and the AWIN scenarios.

Hence, the performance comparison in terms of the probability of error involves two steps: (1) the identification of the optimum filter coefficients provided by the MEEC, MCC and MSE; and (2) the calculation of the analytical probability of error by substituting the obtained vectors w in Equation (51). It is important to remark that, in the first step, we perform a search for the point that minimizes the theoretical error entropy, i.e., for which the gradient of the error entropy is zero, and that is not located at the vicinity of the trivial solution, since it does not correspond to an actual solution for the equalization problem; in the second step, the values of the probability of error for a given vector w are analytically obtained from the error distribution in Equation (51) – please refer to the appendix and theoretical expressions of error probability.

Table I presents the values of the analytical probability of error associated with the Wiener filter and with the optimal solutions of the MEEC and the MCC considering two values of SNR and, in the case of correntropy, different choices for the kernel size.

TABLE I
PROBABILITY OF ERROR (P_e) ASSOCIATED WITH THE OPTIMAL SOLUTIONS OF MSE, MEEC AND MCC, CONSIDERING DIFFERENT VALUES OF SNR AND σ_{ker}^2 .

Criterion		SNR	
		13 dB	20 dB
Wiener		5.76×10^{-3}	7.87×10^{-7}
MEEC		5.80×10^{-3}	3.30×10^{-4}
MCC	$\sigma_{\text{ker}}^2 = 0.1$	5.83×10^{-3}	9.44×10^{-7}
	$\sigma_{\text{ker}}^2 = 0.01$	6.51×10^{-2}	3.30×10^{-4}

The results shown in Table I indicate that the solutions offered by the MEEC and the MCC were not capable of achieving probabilities of error smaller than those associated with the Wiener solution. With respect to the MEEC, we notice that for the SNR of 13 dB, the probability of error was quite close to that obtained by the Wiener filter. On the other hand,

for the SNR of 20 dB, the value of P_e achieved by the Wiener filter is significantly smaller than that related to the MEEC.

This fact can be explained in view of the surfaces shown in Figures 7 and 10. For the SNR of 13 dB, the solution that gives the minimum value of error entropy (excluding the trivial solution) lies in the same region of the Wiener filter, which means that the corresponding coefficient vectors \mathbf{w} are quite similar, so that the probabilities of error obtained in both cases are almost equivalent. On the other hand, for the SNR of 20 dB, the solution that yields the minimum value of error entropy is located in a different region, in the vicinity of the point $\mathbf{w} = [0.6115 \ 0.014]^T$. So, it is expected that the probability of error obtained by the MEEC solution differs from that of the Wiener filter, which is associated with the point $\mathbf{w} = [0.8996 \ -0.3929]^T$. Unfortunately, the equalizer designed by the MEEC is not attractive in view of the significant degradation of performance when compared with the Wiener solution.

A similar phenomenon can be observed in Table I with respect to the MCC. When the kernel size is relatively large, we know from Figure 14 that the optimum solution in the sense of maximum correntropy is quite similar to the Wiener filter, which explains the fact that the equalizer designed according to the MCC reached almost the same values of P_e associated with the Wiener solution. However, when σ_{ker}^2 is small, the MCC and MSE solutions can be significantly different and, consequently, the probabilities of error as well.

For the specific value of $\sigma_{ker}^2 = 0.01$ and the SNR of 20 dB, the optimum solution of the MCC is given by $\mathbf{w} = [0.613 \ 0.014]^T$, being almost equivalent to that obtained by the MEEC in the same scenario. Although this solution yields a significantly worse performance when compared with the Wiener solution, it has an interesting effect: it maximizes the error PDF at the origin. This is confirmed by Figure 17, which displays the PDFs of the error signal, computed via (25), associated with the Wiener filter and the MCC optimum solution.

As we can observe, the PDF of the error signal generated by the MCC optimum solution presents a higher amplitude at the origin. However, this PDF is not as concentrated around the origin as that associated with the Wiener filter, which certainly increases the possibility of occurrence of a decision

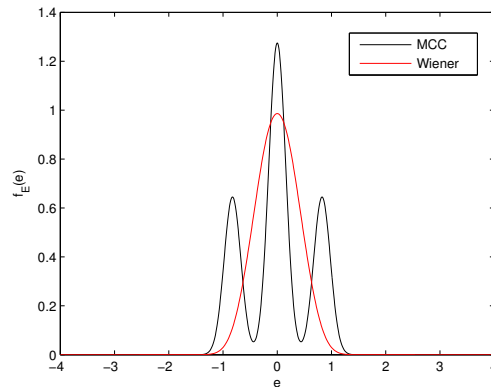


Fig. 17. PDFs of the error signal associated with the Wiener filter and the MCC optimum solution considering an SNR of 20 dB and $\sigma_{ker}^2 = 0.01$.

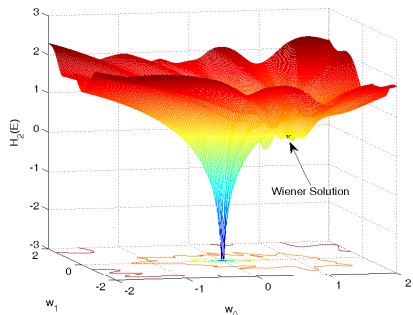
error, as was, in fact, verified and shown in Table I.

C. Second Specific Scenario: Minimum-Phase Channel and AWIN

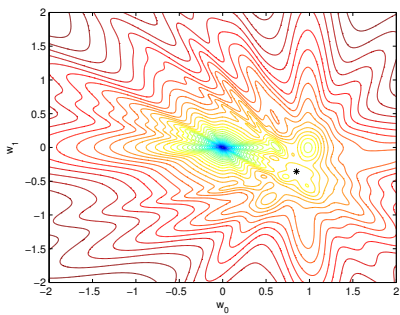
Similarly to the previous analysis, we consider again the scenario with the minimum-phase channel $H(z) = 1 + 0.6z^{-1}$. However, the additive noise affecting the transmitted signal is now the AWIN (with parameters $p = 0.9$ and $\sigma_1^2 \approx 0.0342$ – the other parameters were adjusted depending on the SNR level), which may lead to different behavior of the error entropy and the correntropy criteria in comparison to the AWGN case. It is important to remark that our analysis considers the analytical costs functions, which may present different behavior from the usual adaptive gradient-based algorithms. In that sense, we present in the following some of the noteworthy results considering the impulsive noise.

1) *Error Entropy:* We start by presenting in Figure 18 the surface of the theoretical error entropy and its associated contours for the scenario in question with a BPSK modulated source ($\sigma_g^2 = 0$) and an SNR level of 13 dB. In comparison to the correspondent AWGN case (Figure 7(a)), it is possible to observe that the general cost shape remains the same with the trivial solution $\mathbf{w} = [0 \ 0]^T$ as the minimum error entropy value. Nevertheless, we can note – specially from the contours in Figure 18(b) – that the number of local minima is increased. Again, the Wiener solution is close to one of the local minima, however, the local optima that best reduces entropy

is nearby $\mathbf{w} = [0.6 \ 0]^T$, i.e., a solution that only works as a scale factor over the received signal.



(a) Entropy Surface - Discrete source



(b) Entropy Surface Contours - Discrete source

Fig. 18. Surface of error entropy and its contours considering the second scenario with $SNR = 13$ dB for discrete sources.

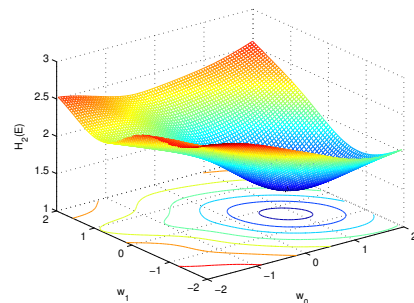
Interestingly, the variation of the SNR level for the impulsive noise causes a similar effect as that of the AWGN: when the SNR level is low, the error entropy surface becomes smoother with a reduced number of minima, as presented in Figure 10(a); when the SNR level is higher, the number of minima increases, like in Figure 10(b), where the SNR level was of 20 dB (being $\mathbf{w} \approx [0.6 \ 0]^T$ the local optima with the lower entropy associated). In the vicinity of the Wiener solution, we observe that the lower the SNR level, the closer is the local optimum of the error entropy criterion to the Wiener solution, which is the same tendency found in the AWGN case.

The particular case of a continuous source is also considered by adjusting $\sigma_g^2 = 0.15$ (and an SNR level of 13 dB). The obtained error entropy surface is much similar to that presented in Figure 7(b) – and, for this reason, we have opted for its omission –, which leads to two main results: first, as already mentioned, the nature of this source reduces the effect of the trivial solution, privileging a local

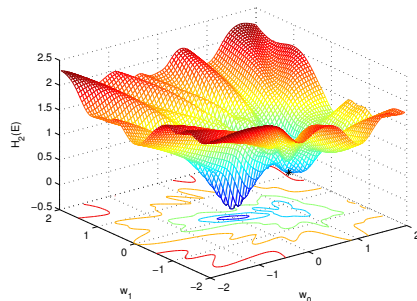
minimum close to the Wiener Solution; second, we note that this type of source also mitigates the peculiar AWIN effect, making the surface similar to that observed in the AWGN case.

From these brief results of the AWIN scenario, we can summarize that the MEEC presents a larger number of local minima in comparison to the AWGN case and, by promoting some adjustments like reducing the SNR level or changing the type of source, we note a ‘Gaussianization’ effect that smooths the entropy surface – in a very similar way of that observed in the AWGN scenario.

The next step involves the comparison of the presented results to the entropy estimator for the AWIN scenario. In order to do so, we consider the discrete source ($\sigma_g^2 = 0$) and the previous minimum-phase channel. In addition, it is assumed that the entropy estimator has at disposal 2500 samples of the error signal $e(n)$ and that the SNR level is of 13 dB. Figures 19(a) and 19(b) illustrate the surfaces of the estimated entropy for the kernel sizes $\sigma_{ker}^2 = 1$ and $\sigma_{ker}^2 = 0.01$, respectively.



(a) $\sigma_{ker}^2 = 1$



(b) $\sigma_{ker}^2 = 0.01$

Fig. 19. Surfaces of the error entropy estimator as a function of the kernel size considering the second scenario.

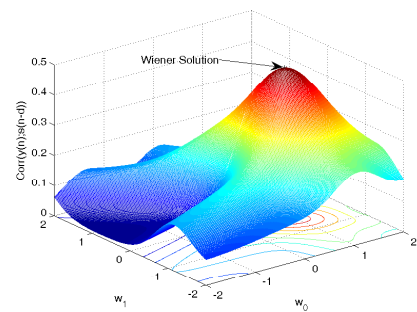
It is possible to notice from Figure 19 that the action of the kernel size strictly follows the indicated pattern discussed in Section IV-B1 for the

AWGN case (Figure 11), i.e., for kernel sizes close or larger than unity, the entropy surface becomes smoother while the global optimum at the origin vanishes; on the other hand, for small kernel sizes, the estimated entropy surface becomes much alike the theoretical entropy for discrete sources. We can also say that changing the SNR levels also leads to results that resemble those obtained for the AWGN case. As an example, for the SNR of 20 dB, the estimated entropy surfaces follows the same general tendencies of that of Figure 13 – and, for that reason, these figures are omitted from this work.

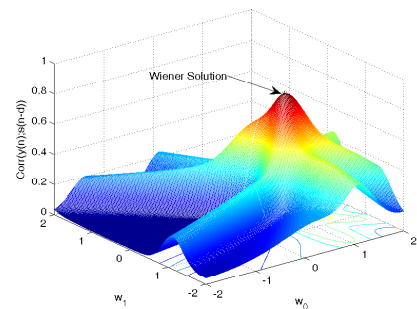
2) *Correntropy*: Following the sequence of the analysis, we now focus on the MCC for a scenario with the impulsive noise and the discrete source. We start by analyzing the effect of the kernel size assuming three cases: σ_{ker}^2 equal to 0.5, 0.1 and 0.01. For each case, we obtained the corresponding theoretical correntropy surfaces, as displayed in Figure 20, for the SNR of 13 dB.

It is possible to note from Figures 20(a) and 20(b) that the presence of AWIN is practically indistinguishable to the case when the noise is the AWGN – since there is great resemblance to the Figures 14(a) and 14(b). Notwithstanding, in Figure 20(c) (for $\sigma_{ker}^2 = 0.01$), other peaks arise on the theoretical correntropy surface that do not match the ones observed in the AWGN case (Figure 14(c)). This indicates that correntropy increases its sensibility – e.g., to the noise effect – only when the kernel size σ_{ker}^2 assumes relatively small values. In that sense, it is an important task to evaluate the quality of the observed solutions for different kernel sizes from the standpoint of the problem of channel equalization. This will be done in Section IV-C3 in terms of the theoretical probability of error.

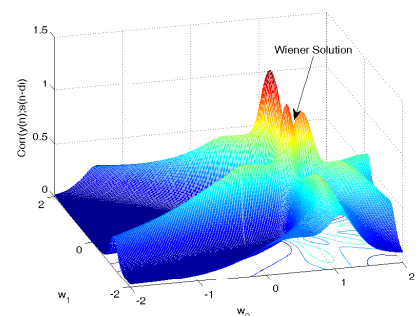
The effect of the SNR level on the theoretical correntropy surface for the AWIN is also a worthy study case. As in the AWGN scenario, we considered the SNR values of 5 and 20 dB (and a kernel size of $\sigma_{ker}^2 = 0.01$) to illustrate the analysis. Figure 21(a) shows the obtained correntropy surface and its associated contours for the SNR level equal to 5 dB. We can notice that, differently from the AWGN case, the peak located near the Wiener solution is not the preferred solution by the MCC, but the peak located at the vicinity of $\mathbf{w} = [0.60]^T$. Nevertheless, if we continue to decrease the SNR level, the peak located near the Wiener solution becomes more pronounced and the situation becomes similar to the



(a) $\sigma_{ker}^2 = 0.5$



(b) $\sigma_{ker}^2 = 0.1$



(c) $\sigma_{ker}^2 = 0.01$

Fig. 20. Surfaces of the theoretical correntropy as a function of the kernel size considering an SNR of 13 dB for the second scenario. The asterisk (*) indicates the position of the Wiener solution.

AWGN case. For a SNR level of 20 dB, the noise effect is weakened and the correntropy surface is very similar to that presented in Figures 15(c) and 15(d) – for this reason, we opted for their omission.

At last, to analyze the estimated correntropy cost function, we present in Figure 22 the surfaces for the SNR level of 13 dB and the kernel sizes $\sigma_{ker}^2 = 0.05$ and $\sigma_{ker}^2 = 0.01$. The correntropy values were estimated from 2500 samples of the error signal.

As can be viewed in Figure 22, the smoothing pattern observed by decreasing/increasing σ_{ker}^2 in the theoretical correntropy is also present in the estimated correntropy – moreover, their surfaces

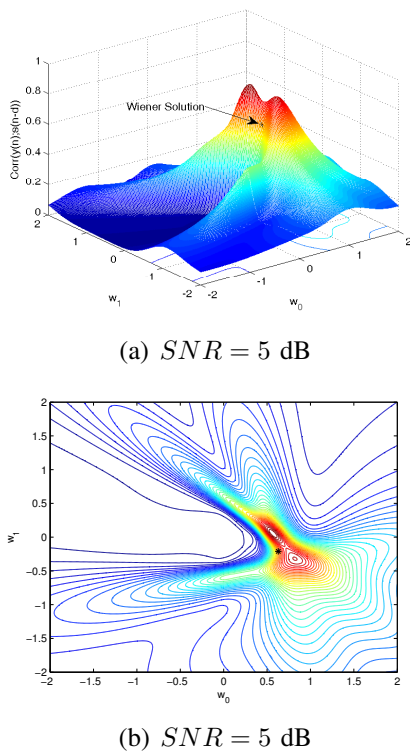


Fig. 21. Surface and contour of the theoretical correntropy for $SNR = 5$ dB considering $\sigma_{ker}^2 = 0.01$. The asterisk (*) indicates the position of the Wiener solution.

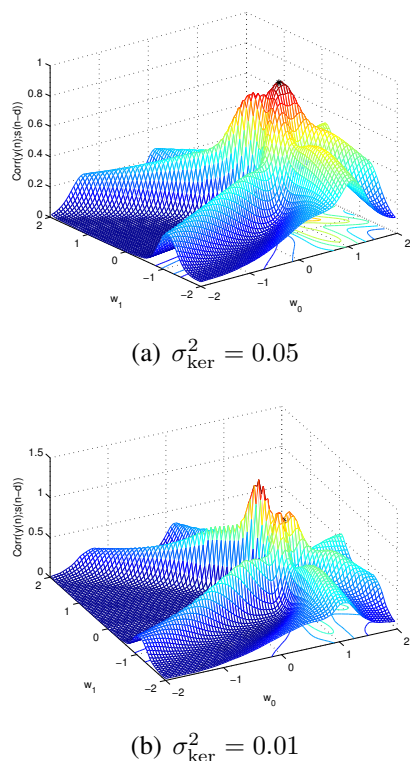


Fig. 22. Surfaces associated with the correntropy estimator as a function of the kernel size considering the second scenario.

also tend to be similar. It is also possible to note the emergence of local optima solutions different from the Wiener solution (as observed in Figure 20(c)).

3) *Probability of Error*: Under the same assumptions made for the AWGN scenario, we evaluate the performance of the solutions obtained by each criterion for the AWIN case in terms of the analytical probability of error P_e . The obtained results for the solutions of the Wiener filter, the MEEC and MCC – as well as the lowest P_e attainable (using as cost function Equation (54), in the Appendix) – are presented in Tables II and III for each scenario. Observe that we have decided to present more than one solution for the MEEC and MCC, since their cost functions present several local optima in the AWIN case. The solutions are ordered from the best to the worst optima (top to bottom) according to the corresponding criterion. For a better distinction, we also included the angle $\theta = \arctan(w_1/w_0)$ associated with each solution, which will be helpful since a gain factor does not change the error probability – i.e., the decision boundary remains the same.

TABLE II
PROBABILITY OF ERROR (P_e) ASSOCIATED WITH THE SOLUTIONS OF MSE, MEEC AND MCC, CONSIDERING DIFFERENT VALUES OF SNR AND σ_{ker}^2 .

Criterion		Solution	SNR	
			13 dB	20 dB
Lowest P_e		1	0.0114	0.0011
Wiener		1	0.0117	0.0011
MEEC		1	0.0166	0.0057
		2	0.0156	0.0055
		3	0.0149	0.0014
MCC	$\sigma_{ker}^2 = 0.1$	1	0.0116	0.0012
		1	0.0166	0.0057
	$\sigma_{ker}^2 = 0.01$	2	0.0114	0.0012
		3	0.0150	0.0014
		3	0.0150	0.0014
		3	0.0150	0.0014

As can be noted, the analytical values of the error probability P_e shown in Table II are considerably higher than that presented in Table I, which is a direct consequence of the difficulty to deal with the impulsive noise. Table III shows that none of the solutions are equivalent and, in most cases, the solutions provided by MEEC and MCC lead to P_e values higher than those of the Wiener solution. Exceptions are only made when the SNR value is of 13 dB and for certain configurations of MCC: for $\sigma_{ker}^2 = 0.1$, the MCC surface has a single maximum

TABLE III
 FILTER \mathbf{w} AND ANGLE θ ASSOCIATED WITH THE SOLUTIONS OF MSE, MEEC AND MCC, CONSIDERING DIFFERENT VALUES OF SNR AND σ_{ker}^2 .

Criterion		Sol.	SNR					
			13 dB			20 dB		
			w_0	w_1	$\theta(\text{rad})$	w_0	w_1	$\theta(\text{rad})$
Lowest P_e		1	–	–	–0.3280	–	–	–0.4502
Wiener		1	0.8503	–0.3572	–0.3977	0.8997	–0.3930	–0.4118
MEEC		1	0.6105	0.0135	0.0221	0.6215	0.0035	0.0056
		2	0.9825	–0.0025	–0.0025	0.9965	0	0
		3	0.7255	–0.4305	–0.5355	0.7325	–0.4390	–0.5399
MCC	$\sigma_{\text{ker}}^2 = 0.1$	1	0.8705	–0.3550	–0.3872	0.8855	–0.3635	–0.3895
		1	0.6120	0.0145	0.0237	0.6210	0.0045	0.0057
	$\sigma_{\text{ker}}^2 = 0.01$	2	0.9695	–0.3640	–0.3592	0.9950	–0.3725	–0.3582
		3	0.7280	–0.4340	–0.5376	0.7340	–0.4390	–0.5390

– Figure 20(b) – very close to Wiener solution (note the angle θ proximity) with associated P_e of 0.0116; for $\sigma_{\text{ker}}^2 = 0.01$, the (second) maximum $\mathbf{w} = [0.9695 \text{ } -0.3640]^T$ leads to $P_e = 0.0114$, which is the solution that best approximates the lowest P_e , but is not equivalent – note that its angle θ is in the middle of the range between the angles of the Wiener solution and the lowest P_e . For the SNR of 20 dB, the solutions of MEEC and MCC lead to higher P_e and angles more distant than those associated with the lowest P_e .

As discussed earlier in the AWGN scenario, the most promising solutions lie in the vicinity of the Wiener solution, and, as shown by Table III this also is the case for the AWIN scenario for the solutions of the MEEC and MCC. However, from the results of both SNR of 13 and 20 dB, we note that there is no specific case or solution for the MEEC and MCC that is more suitable to the problem of channel equalization in place of the Wiener solution as a generic tool to deal with the diverse and variable range of SNR levels present in communication scenarios.

D. Third Specific Scenario: $K > 2$

The previous sections were devoted to the study of the ITL criteria considering cases where the equalizer had only two coefficients ($K = 2$). This restriction allowed us to visualize the surfaces of the theoretical error entropy and correntropy, as well as of their kernel-based estimators, enabling the analysis of their behavior as a function of the SNR

and/or of the kernel size, as well as the identification of the optimum solution of each criterion and, finally, the comparison of the performance offered by these solutions with that obtained by the Wiener filter. Now, we shall consider the design of an equalizer containing a larger number of coefficients according to the MEEC, MCC and MSE with the purpose of assessing their performances.

Let $H(z) = 0.2144 + 0.7505z^{-1} - 0.5361z^{-2} + 0.3216z^{-3}$ be the transfer function of the channel, which receives sequences of BPSK symbols at the input. A FIR filter composed of $K = 5$ coefficients shall be employed as the equalizer considering a SNR of 17 dB, an equalization delay of $d = 1$ and the AWGN model of noise. We can safely restrict the analysis to the AWGN case having in view the central limit theorem [21]: since the filtered noise, which is the basis for deriving the expression of the error PDF, consists of the linear combination of K noise samples (recall that the term related to $g(n)$ is not present when the source is discrete), as shown in Section III-B, its PDF can be approximated by a Gaussian function as K is increased. In other words, as the length of the equalizer increases, the less important is the actual PDF of the noise for the signal model, since the filtered noise can be represented in the form of a Gaussian random variable. Therefore, the analysis of the AWGN case encompasses, to a certain extent, the scenarios considering other noise models, such as the AWIN.

Since it is no longer possible to see the surfaces of the error entropy and correntropy and there is

the menace of the trivial solution, particularly in the context of the error entropy, we adopted a more elaborate strategy for determining the optimal solutions of these criteria: (i) an immune-inspired search algorithm, known as CLONALG [27], was employed to spread a large repertoire of candidate solutions throughout the search space and progressively improve such solutions towards the optimal regions; (ii) having the final set of candidate solutions, we compute the norm of the gradient of the error entropy or correntropy at each solution; (iii) finally, among the solutions with the smallest values of the gradient norm, the solution which yields the best value for the error entropy or correntropy is considered to be the optimum, as long as it does not correspond to the origin.

After determining the optimal solutions of the ITL criteria, we evaluated the performances of the corresponding equalizers in terms of the analytical probability of error, whose values are displayed in Table IV along with those associated with the Wiener solution. Additionally, we show in Table V the optimum filter coefficients obtained by each criterion.

TABLE IV
PROBABILITY OF ERROR (P_e) ASSOCIATED WITH THE OPTIMAL SOLUTIONS OF MSE, MEEC AND MCC CONSIDERING DIFFERENT VALUES OF σ_{ker}^2 .

Criterion		SNR
		17 dB
Wiener		4.23×10^{-5}
MEEC		4.26×10^{-5}
MCC	$\sigma_{ker}^2 = 0.05$	4.48×10^{-5}
	$\sigma_{ker}^2 = 0.01$	4.18×10^{-5}

TABLE V
OPTIMUM COEFFICIENTS ASSOCIATED WITH THE MSE, MEEC AND MCC CONSIDERING THE THIRD SCENARIO.

Coefficients	Criterion			
	Wiener	MEEC	MCC	
			$\sigma_{ker}^2 = 0.05$	$\sigma_{ker}^2 = 0.01$
w_0	1.015	1.015	0.9770	0.9199
w_1	0.6933	0.6951	0.6776	0.6249
w_2	0.1049	0.1057	0.1054	0.0937
w_3	-0.1631	-0.1629	-0.1564	-0.1472
w_4	-0.1260	-0.1261	-0.1222	-0.1133

The results shown in Table IV indicate that the performances associated with each criterion are relatively similar. For instance, the optimum

solution offered by the MEEC is quite similar to the Wiener solution, as indicated in Table V, and the corresponding probabilities of error are almost equivalent, being the performance of the Wiener solution slightly superior to that obtained by the MEEC solution.

Interestingly, the best and the worst performances are related to the MCC: when $\sigma_{ker}^2 = 0.01$, the probability of error obtained by the MCC solution is slightly smaller than that associated with the Wiener solution; on the other hand, when $\sigma_{ker}^2 = 0.05$, the MCC solution reaches a slightly worse value for the probability of error. The reason behind these results can be observed in Figure 23, which presents the shape of the PDF of the error signal considering the Wiener and the MCC solutions.

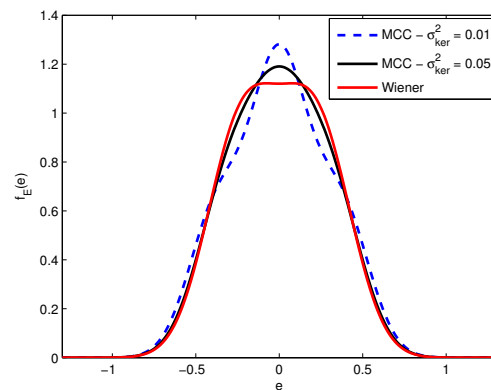


Fig. 23. PDFs of the error signal associated with the Wiener filter and the MCC optimum solution considering $\sigma_{ker}^2 = 0.05$ and $\sigma_{ker}^2 = 0.01$.

As we can notice, the error PDF associated with the MCC using $\sigma_{ker}^2 = 0.01$ achieves a significantly higher peak at the origin and is also narrower than the error PDF related to the Wiener solution in the region where the magnitude of the error values are below 0.5. Having in view the performances verified in Table IV, it is possible to say that these characteristics counterbalance the larger tail of the error PDF associated with the MCC and, ultimately, lead to the slightly smaller value of the probability of error.

However, if $\sigma_{ker}^2 = 0.05$, even though the peak at the origin is also higher than that associated with the Wiener solution, it is not sufficient to compensate the slower decay of the error PDF in the regions close to the tails, which may be the factor causing the small deterioration in the probability of error

when compared with the value obtained by the Wiener solution.

V. CONCLUSIONS

In this work, we performed an analysis - in the context of linear channel equalization - of two criteria central to the theory of supervised information theoretic learning, those of minimum error entropy (MEEC) and maximum correntropy (MCC). The work brought a number of original contributions that allowed a clearer view of the potential and of the drawbacks of both approaches to be reached:

- A derivation of the exact cost functions in the case of interest. This allowed the question of statistical estimation (e.g. via kernel methods) to be separated from the inherent characteristics of each criterion. This derivation is explicitly shown for both Gaussian and impulsive additive white noises for linear FIR equalizers.
- A detailed discussion between the points of contact between the aforementioned exact derivations and the data-driven canonical cost functions based on Parzen window estimation.
- An analysis of the cost functions of both criteria in the combined channel + equalizer space. This space, in which the solutions of interest are known a priori (they are zero-forcing receivers), is important in revealing the consistency of any adaptive criterion. For instance, the existence of spurious minima in this space (which occurs for some Bussgang methods [20]) indicates that the algorithm might not have a satisfactory degree of robustness. With respect to the MEEC, it is shown that it may present undesirable solutions as global optima for the case of discrete sources, whereas it is more robust when continuous sources are employed. The MCC, albeit also more suited to the case of continuous sources, is more robust than the MEEC in terms of the quality of the verified optima.
- A similar analysis in the equalizer parameter space, including different noise models, that raises points of contact between the use of Gaussian-based Parzen window estimators and the effect of additive noise and reveals drawbacks associated with the MEEC, for discrete sources, insofar as the null solution is considered. More robustness is obtained, once more, if continuous sources are adopted.

- Finally, a careful study including Wiener and minimum probability of error solutions is carried out for distinct noise levels, which indicates that, in some cases, the use of ITL leads to solutions with a better symbol error rate than those obtained with the classical mean squared error paradigm.

As perspectives for future works, we may point out (i) a theoretical analysis regarding the use of Shannon's entropy and (ii) the extension of the study to the unsupervised scenario, starting with a prediction-based approach [20], considering both the MEEC and the MCC, followed by an analysis of blind ITL criteria, such as the blind counterparts of entropy and correntropy, as well as those based on the notion of PDF matching [28].

APPENDIX

DERIVATION OF THE THEORETICAL PROBABILITY OF ERROR

Considering the 2-PAM modulation, the probability of error, denoted by P_e , can be expressed as follows:

$$P_e = P(e|s(n-d) = +1)P(s(n-d) = +1) + P(e|s(n-d) = -1)P(s(n-d) = -1). \quad (42)$$

The decision rule of the equalizer is given by $\hat{s}(n-d) = \text{sign}(y(n))$. Thus, assuming that the transmitted symbol we wish to recover is $s(n-d) = +1$, a decision error occurs if $y(n) < 0$. Analogously, if $s(n-d) = -1$, the decision is incorrect if $y(n) > 0$. Hence, by expressing the conditional probabilities of error as a function of the equalizer output $y(n)$, we obtain:

$$P_e = \frac{1}{2}P(y(n) < 0|s(n-d) = +1) + \frac{1}{2}P(y(n) > 0|s(n-d) = -1). \quad (43)$$

In Section III-B, we have derived the expressions of the conditional PDFs $f_Y(Y = y|s(n-d) = +1)$ and $f_Y(Y = y|s(n-d) = -1)$. Having these PDFs at hand, the probability of error can be determined by the following equation:

$$P_e = \frac{1}{2} \int_{-\infty}^0 f_Y(y|s(n-d) = +1)dy + \frac{1}{2} \int_0^{\infty} f_Y(y|s(n-d) = -1)dy. \quad (44)$$

By substituting (14) and (15) in (44), we obtain:

$$P_e = \frac{1}{2^{K+D-1}} \left(\sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \int_{-\infty}^0 f_{\eta_f}(y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)) dy + \sum_{i \in \mathcal{S}^{\mathcal{D}^-}} \int_0^{\infty} f_{\eta_f}(y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)) dy \right), \quad (45)$$

where $\eta'(n)$ is given by (11). However, in the context of the 2-PAM, the first term in (11) does not exist.

AWGN

In the case of the AWGN model, the filtered noise $\eta'(n)$ is a Gaussian random variable with zero mean and variance $\sigma_{\eta_f}^2 = \mathbf{w}^T \sigma_{\eta}^2 \mathbf{w}$. This means that the conditional probability $P(y(n) < 0 | s(n-d) = +1)$, denoted as $P_{y|s^+}$, can be calculated by the following expression:

$$P_{y|s^+} = \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{1}{2^{K+D-2}} \times \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma_{\eta_f}^2}} \exp\left(-\frac{(y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n))^2}{2\sigma_{\eta_f}^2}\right) dy.$$

Let $x = \frac{y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sigma_{\eta_f}}$. By expressing the integral in terms of the variable x and exploring the symmetry of the Gaussian function, we obtain:

$$P_{y|s^+} = \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{1}{2^{K+D-2}} \times \int_{\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sigma_{\eta_f}}}{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx. \quad (46)$$

Using the Q -function, defined as

$$Q(x) \triangleq P(X > x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{t^2}{2}\right) dt, \quad (47)$$

the conditional probability $P_{y|s^+}$ is determined as follows:

$$P_{y|s^+} = \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{1}{2^{K+D-2}} Q\left(\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sigma_{\eta_f}}\right). \quad (48)$$

The same procedure can be employed to derive the conditional probability $P(y(n) > 0 | s(n-d) = -1)$, which is given by the following equation:

$$P_{y|s^-} = \sum_{i \in \mathcal{S}^{\mathcal{D}^-}} \frac{1}{2^{K+D-2}} Q\left(\frac{-\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sigma_{\eta_f}}\right). \quad (49)$$

Therefore, the probability of error can be written as:

$$P_e = \frac{1}{2} \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{1}{2^{K+D-2}} Q\left(\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sigma_{\eta_f}}\right) + \frac{1}{2} \sum_{i \in \mathcal{S}^{\mathcal{D}^-}} \frac{1}{2^{K+D-2}} Q\left(\frac{-\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sigma_{\eta_f}}\right). \quad (50)$$

Finally, using the variable β , defined in (23), the theoretical probability of error associated with the equalizer in the AWGN case is given by:

$$P(e) = \frac{1}{2^{K+D-1}} \sum_{i=0}^{2^{K+D-1}-1} Q\left(\beta_i \frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sigma_{\eta_f}}\right). \quad (51)$$

AWIN

For this type of noise, the PDF of the filtered noise is given by Equation (32). Hence, the conditional probability $P(y(n) < 0 | s(n-d) = +1)$ can be calculated as follows:

$$P_{y|s^+} = \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} \frac{1}{2^{K+D-2}} \times \int_{-\infty}^0 \left(p^2 G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2}(y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)) + p(1-p) G_{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2}(y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)) + (1-p)p G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2}(y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)) + (1-p)^2 G_{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2}(y - \mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)) \right) dy.$$

The four integrals involved in the computation of $P(y(n) < 0 | s(n-d) = +1)$ can be represented in terms of the Q -function, so that the conditional probability can be determined by the following equation:

$$P_{y|s^+} = \frac{1}{2^{K+D-2}} \sum_{i \in \mathcal{S}^{\mathcal{D}^+}} p^2 Q\left(\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2}}\right) + p(1-p) Q\left(\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2}}\right) + (1-p)p Q\left(\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2}}\right) + (1-p)^2 Q\left(\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2}}\right). \quad (52)$$

By employing a similar approach, the conditional probability $P(y(n) > 0 | s(n-d) = -1)$ can be

expressed as follows:

$$\begin{aligned}
 P_{y|s^-} = & \frac{1}{2^{K+D-2}} \sum_{i \in \mathcal{S}^{\mathcal{D}^-}} p^2 Q \left(-\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_1^2}} \right) \\
 & + p(1-p) Q \left(-\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2}} \right) \\
 & (1-p)p Q \left(-\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2}} \right) \\
 & (1-p)^2 Q \left(-\frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_2^2}} \right). \quad (53)
 \end{aligned}$$

Therefore, by combining the two conditional probabilities according to (43), we obtain the exact expression of the probability of error considering the model:

$$\begin{aligned}
 P(e) = & \frac{1}{2^{K+D-1}} \sum_{i=0}^{N_S-1} p^2 Q \left(\beta_i \frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_1^2}} \right) \\
 & + p(1-p) Q \left(\beta_i \frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_1^2 + w_1^2 \alpha^2 \sigma_2^2}} \right) \\
 & + (1-p)p Q \left(\beta_i \frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_1^2}} \right) \\
 & + (1-p)^2 Q \left(\beta_i \frac{\mathbf{w}^T \mathbf{H} \mathbf{s}_i(n)}{\sqrt{w_0^2 \alpha^2 \sigma_2^2 + w_1^2 \alpha^2 \sigma_2^2}} \right). \quad (54)
 \end{aligned}$$

ACKNOWLEDGMENTS

This work was supported by FAPESP (2013/06322-0 and 2013/14185-2), CNPq and CAPES.

REFERENCES

[1] R. Lucky, "Automatic Equalization for Digital Communication," *Bell Systems Technical Journal*, vol. 44, pp. 547–588, 1965. doi: 10.1002/j.1538-7305.1965.tb01678.x

[2] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Prentice Hall, 2008.

[3] C. C. Cavalcante, J. R. Montalvo, B. Dorizzi, and J. C. M. Mota, "A neural predictor for blind equalization in digital communications," in *Adaptive Systems for Signal Processing, Communications and Control (AS-SPCC' 00)*, 2000. doi: 10.1109/ASSPCC.2000.882498 pp. 347–351.

[4] K. Burse, R. Yadav, and S. Shrivastava, "Channel equalization using neural networks: A review," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 40, no. 3, pp. 352–357, 2010. doi: 10.1109/TSMCC.2009.2038279

[5] L. Boccato, A. Lopes, R. Attux, and F. J. Von Zuben, "An extended echo state network using volterra filtering and principal component analysis," *Neural Networks*, vol. 32, pp. 292–302, 2012. doi: 10.1016/j.neunet.2012.02.028

[6] S. K. Patra and B. Mulgrew, "Fuzzy techniques for adaptive nonlinear equalization," *Signal Processing*, vol. 80, pp. 985–1000, 2000. doi: 10.1016/S0165-1684(00)00015-3

[7] R. Ferrari, R. Suyama, R. R. Lopes, R. R. F. Attux, and J. M. T. Romano, "An optimal MMSE fuzzy predictor for SISO and MIMO blind equalization," in *First IAPR Workshop on Cognitive Information Processing*, 2008, pp. 86–91.

[8] R. Ferrari, C. M. Panazio, R. R. F. Attux, C. C. Cavalcante, L. N. de Castro, F. J. Von Zuben, and J. M. T. Romano, "Unsupervised channel equalization using fuzzy prediction-error filters," in *Neural Networks for Signal Processing, 2003. NNSP'03. 2003 IEEE 13th Workshop on*, 2003. doi: 10.1109/NNSP.2003.1318086 pp. 869–878.

[9] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2918–2927, 1993. doi: 10.1109/78.236513

[10] S. Haykin, *Adaptive filter theory*, 5th ed. Prentice Hall, 2013.

[11] J. C. Principe, *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer, 2010.

[12] L. Boccato, D. G. Silva, D. Fantinato, R. Ferrari, and R. Attux, "A comparative study of non-MSE criteria in nonlinear equalization," in *Telecommunications Symposium (ITS), 2014 International*. IEEE, 2014. doi: 10.1109/ITS.2014.6947953 pp. 1–5.

[13] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. Springer, 2003.

[14] W. Liu, P. Pohkarel, and J. C. Principe, "Correntropy: properties and applications in non-Gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007. doi: 10.1109/TSP.2007.896065

[15] J. C. Principe, D. Xu, and J. Fisher, "Information-Theoretic Learning," in *Unsupervised Adaptive Filtering, Volume 1: Blind Source Separation*, S. Haykin, Ed. Wiley-Interscience, 2000, ch. 7, pp. 1–62.

[16] S. Watanabe, *Pattern Recognition: Human and Mechanical*. Wiley, 1985.

[17] I. Santamaria, D. Erdogmus, and J. C. Principe, "Entropy minimization for supervised digital communications channel equalization," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1184–1192, 2002. doi: 10.1109/78.995074

[18] D. Erdogmus and J. Principe, "Generalized information potential for adaptive systems training," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1035–1044, 2002. doi: 10.1109/TNN.2002.1031936

[19] I. Santamaria, P. Pokharel, and J. Principe, "Generalized Correlation Function: Definition, Properties, and Application to Blind Equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006. doi: 10.1109/TSP.2006.872524

[20] J. M. T. Romano, R. R. d. F. Attux, C. C. Cavalcante, and R. Suyama, *Unsupervised Signal Processing: Channel Equalization and Source Separation*. CRC Press, 2010.

[21] A. Leon-Garcia, *Probability, Statistics and Random Processes for Electrical Engineering*, 3rd ed. Prentice Hall, 2008.

[22] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a non-minimum phase system: blind adjustment of a linear equalizer in data communications," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 385–399, 1980. doi: 10.1109/TAC.1980.1102343

[23] G. J. Foschini, "Equalizing without altering or detecting data," *Bell Labs Technical Journal*, vol. 64, no. 8, pp. 1885–1911, 1985. doi: 10.1002/j.1538-7305.1985.tb00040.x

[24] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of non-minimum phase systems (channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 312–321, 1990. doi: 10.1109/18.52478

[25] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems,"

Proceedings of the IEEE, vol. 86, no. 11, pp. 2210–2239, 1998. doi: 10.1109/5.726788

- [26] M. Styblinski and T. Tang, “Experiments in nonconvex optimization: Stochastic approximation with function smoothing and simulated annealing,” *Neural Networks*, vol. 3, pp. 467–483, 1990. doi: 10.1016/0893-6080(90)90029-K
- [27] L. N. de Castro and F. J. Von Zuben, “Learning and optimization using the clonal selection principle,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239–251, 2002. doi: 10.1109/TEVC.2002.1011539
- [28] M. Lazaro, I. Santamaria, D. Erdogmus, K. Hild, C. Pantaleon, and J. Principe, “Stochastic blind equalization based on PDF fitting using Parzen estimator,” *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 696–704, 2005. doi: 10.1109/TSP.2004.840767



Rafael Ferrari received his B.S. (2001), M.Sc. (2005), and Ph.D. (2011) degrees in Electrical Engineering from the University of Campinas (UNICAMP). From 2011 to 2015, he has been a researcher at the Center for Petroleum Studies (CEPETRO-UNICAMP). His main research interest is in digital signal processing, especially its application to communication systems and to seismic data analysis.



Levy Boccato was born in Piracicaba, São Paulo, Brazil, in 1986. He received the titles of Computer Engineer (2008), Master in Electrical Engineering (2010) and Doctor in Electrical Engineering (2013), all from the University of Campinas (UNICAMP), São Paulo, Brazil. Currently, he is an Assistant Professor at the same university. His main research interests include computational intelligence, adaptive filtering, machine learning and signal processing.

filtering, machine learning and signal processing.



Aline Neves received the B.S. and M.S. degree in Electrical Engineering from the State University of Campinas (UNICAMP), Brazil, in 1999 and 2001 respectively. She received her Ph.D. degree in 2005, also in Electrical Engineering, from the University René Descartes (Paris V), Paris, France. Recently, she is an associate professor at the Engineering, Modeling and Applied Social Science Center of

the Federal University of ABC, Santo André, Brazil. Her research interests consist of equalization, channel estimation, source separation and information-theoretic learning.



Denis G. Fantinato was born in Americana, Brazil, in 1985. He received the B.S. and M.Sc. degrees in Electrical Engineering from the University of Campinas (UNICAMP) in 2011 and 2013, respectively. Currently, he is an Ph.D. student at the same institution. His main research interests are blind signal processing, adaptive filtering and information-theoretic learning.



Daniel G. Silva was born in Botucatu, Brazil, in 1983. He received the B.S. degree in Computer Engineering and the M.S. and Ph.D. degrees in Electrical Engineering, all from the University of Campinas (Unicamp), São Paulo, Brazil, in 2006, 2009, and 2013, respectively. Currently, he is a Professor at the Department of Electrical Engineering (ENE) of the University of Brasília (UnB). His main

research interests are information-theoretic learning, adaptive signal processing and computational intelligence.



Romis Attux was born in Goiânia, Brazil, in 1978. He obtained the degrees of Electrical Engineer (1999), Master in Electrical Engineering (2001) and Doctor in Electrical Engineering from the University of Campinas (UNICAMP), Brazil. He is currently an Associate Professor at the same university. His main research areas are unsupervised signal processing, computational intelligence, dynamical systems / chaos and brain-computer interfaces.

systems / chaos and brain-computer interfaces.