CrossMark

Analysis of a Novel Density Matching Criterion Within the ITL Framework for Blind Channel Equalization

Denis G. Fantinato[1] (ID) · **Aline Neves[2]** ·
Romis Attux[1]

Abstract In blind channel equalization, the use of criteria from the field of information theoretic learning (ITL) has already proved to be a promising alternative, since the use of the high-order statistics is mandatory in this task. In view of the several existent ITL propositions, we present in this work a detailed comparison of the main ITL criteria employed for blind channel equalization and also introduce a new ITL criterion based on the notion of distribution matching. The analyses of the ITL framework are held by means of comparison with elements of the classical filtering theory and among the studied ITL criteria themselves, allowing a new understanding of the existing ITL framework. The verified connections provide the basis for a comparative performance analysis in four practical scenarios, which encompasses discrete/continuous sources with statistical independence/dependence, and real/complex-valued modulations, including the presence of Gaussian and non-Gaussian noise. The results indicate the most suitable ITL criteria for a number of scenarios, some of which are favorable to our proposition.

✉ Denis G. Fantinato
    denisgf@dca.fee.unicamp.br

    Aline Neves
    aline.neves@ufabc.edu.br

    Romis Attux
    attux@dca.fee.unicamp.br

[1] School of Electrical and Computer Engineering (FEEC), University of Campinas, Campinas,
    SP 13083-852, Brazil

[2] Engineering, Modeling and Applied Social Science Center (CECS), Federal University of ABC,
    Santo André, SP, Brazil

Birkhäuser

1 Introduction

In the signal processing area, the problem of channel equalization is of paramount significance in view of its vast horizon of applications, like in communication systems, radar, sonar, seismic analysis and image processing [10]. The equalization task can be performed in a supervised or in an unsupervised fashion, which essentially differ by the presence or absence of a desired—or reference—sequence throughout the filter adaptation process. The unsupervised (or blind) operation mode has the advantage of a higher information transmission rate, since the reference sequence (e.g., a header already known by the receiver) does not need to be sent. On the other hand, in order to correctly perform linear blind equalization, the use of equalization criteria able to encompass statistical information beyond second order is necessary [23].

The research branch known as information theoretic learning (ITL) [22] focuses on the study of criteria and methods capable of extracting signal statistical information in a manner as complete as possible, using concepts and measures from information theory [3]. In that sense, its application to the blind channel equalization problem is very attractive. Indeed, there are already a set of criteria successfully employed on this task, like Rényi's entropy [24], correntropy [26] and the matching of probability density functions (PDFs) [13,15].

Although the main features of each ITL criteria have been individually stressed, a "side-by-side" comparison among them has not yet been addressed—with some exceptions, like [13,25], where a few ITL criteria were considered. Thus, this paper includes two main contributions: We develop a joint analysis considering the complete unsupervised ITL framework in the context of channel equalization, and, in addition to that, in order to compose the *corpus* of criteria, we propose a new ITL criterion based on PDF matching. The study will aim at establishing connections between ITL criteria and the classical unsupervised (or even supervised) methods—something very clarifying, since the classical approach can be more intuitive due to its simplicity— as well as the relationship among themselves. In this context, we will show that the proposed criterion presents a good performance, significantly reducing computational cost.

The analysis will be carried out under the assumption of linear channels—in the presence of noise, e.g., Gaussian and impulsive noise—and of linear finite impulse response (FIR) filters as equalizers, for both continuous and discrete sources. Blind criteria are known to present, as a rule, multimodal cost functions [23], which has motivated the use of metaheuristics [2] in their optimization, as they are capable of avoiding, to a certain degree, local optima. For this reason, the analysis of the equalization performance—measured in terms of the residual intersymbol interference (ISI) caused by the channel—will encompass two optimization methods: those based on the stochastic gradient and on the differential evolution (DE) metaheuristic [29].

This work is organized as follows. In Sect. 2, we present the blind channel equalization problem and introduce the main aspects of the ITL research field. In Sect. 3,

we present all the classes of ITL criteria to be analyzed, including the new proposition. The relationship between the ITL framework and the classical theory is shown in Sect. 4 and the connections among them in Sect. 5. In Sect. 6, we present a performance analysis in four distinct scenarios and, finally, the conclusions are exposed in Sect. 7.

2 The Blind Equalization Problem and Information Theoretic Learning

In the equalization problem, an information signal $s(n)$ is transmitted over a channel in order to reach a receiver, as shown in Fig. 1. However, the channel can impose certain distortions to the signal, especially the so-called intersymbol interference (ISI). From the receiver standpoint, the incoming signal can be written as

$$x(n) = \sum_{i=0}^{M_h} h_i^* s(n-i) + \eta(n) = \mathbf{h}^{\mathrm{H}} \mathbf{s}(n) + \eta(n), \tag{1}$$

where it is assumed a linear channel with impulse response $\mathbf{h} = [h_0 \ h_1 \ \ldots \ h_{M_h}]^{\mathrm{T}}$ of length M_h+1 and the presence of additive noise $\eta(n)$, which can be the result of interferences of other signals and/or thermal noise; $\mathbf{s}(n) = [s(n) \ s(n-1) \ \ldots \ s(n - M_h)]^{\mathrm{T}}$ is the vector with the transmitted signal $s(n)$ and its delayed versions, $(\cdot)^*$ denotes complex conjugation, and $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transposition. Both channel and noise are assumed to be stationary.

In order to mitigate or completely remove the channel effects, as well as eventual noise disturbances, a filter known as equalizer is employed. In the case of a linear finite impulse response (FIR) filter with $M+1$ coefficients $\mathbf{w} = [w_0 \ w_1 \ \ldots \ w_M]^{\mathrm{T}}$, the equalizer output signal $y(n)$ is given by

$$y(n) = \mathbf{w}^{\mathrm{H}} \mathbf{x}(n), \tag{2}$$

where $\mathbf{x}(n) = [x(n) \ x(n-1) \ \ldots \ x(n-M)]^{\mathrm{T}}$ is the vector of received signal samples.

Theoretically, the blind (channel) equalization problem is founded on two main pillars: the Benveniste–Goursat–Rouget (BGR) [1] and the Shalvi–Weinstein (SW) [27] theorems. Basically, they state that, in statistical terms, successful equalization requires the extraction not only of second-order moments of underlying signals, but also their high-order statistics (HOS). In view of this, a very promising research branch is that of information theoretic learning (ITL)-based methods [22], which seek an extensive extraction and utilization of the statistical information about the signals, including the HOS. The potential of ITL is usually attributed to its ability in using a

Fig. 1 Block diagram of a communication system
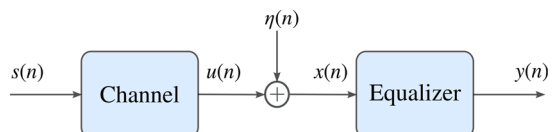
large amount of statistical information about the signal distributions through metrics and concepts derived from information theory [3], like entropy and mutual information. Therefore, based on the BGR and SW theorems and also on the richer statistical content offered by the ITL framework, it becomes a very attractive possibility to employ the ITL methods for solving the problem of blind channel equalization.

In the following, we present, to the best of our knowledge, an original overview of the main ITL criteria that are applied to perform blind channel equalization. Moreover, we also introduce a new ITL criterion, which can be viewed as a variation of the quadratic matching of multivariate distributions [6].

3 Unsupervised ITL Criteria

The unsupervised ITL framework can be roughly divided into three main classes, which are based on different concepts: Rényi's entropy, correntropy and PDF matching.

3.1 Rényi's Entropy

Since Shannon's classical work on information theory [28], the entity named entropy gained remarkable importance in the context of communications [3], which can be explained by its inherent capability of using all the statistical information associated with a random variable (RV) and to represent it as a measure linked to the notion of uncertainty. From the standpoint of equalization, the adoption of the entropy measure proved to be a promising tool for obtaining efficient and sound optimization criteria, given the vast diversity of application scenarios. It is also possible to assert that this entity was primarily responsible for giving rise to the ITL research area [22]. Particularly, the notion of reducing the entropy (or the uncertainty) associated with a filtered signal is in consonance with the idea of mitigating the distortions caused by transmission.

In that sense, in parallel with the development of supervised ITL criteria [24], one of the first blind ITL approaches aimed at bringing together entropy and the p-order dispersion that engenders the Godard's family of cost functions [8]. However, in order to obtain simpler estimators for entropy, instead of Shannon's definition, the α-order Rényi's entropy [22] is considered, being defined as

$$H_\alpha(Y) = \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f_Y^\alpha(v)\, \mathrm{d}v, \tag{3}$$

where $f_Y(v)$ is the probability density function associated with the RV Y. The use of the Rényi's entropy results in the cost function that we call *Rényi's entropy of the p-order Dispersion* (RD) [24]:

$$J_{\mathrm{RD}}(\mathbf{w}) = H_\alpha\left(Y^p - R_p\right) = H_\alpha(Y^p) \tag{4}$$

where the RV Y^p is associated with the modulus signal $|y(n)|^p$, $R_p = E\left[S^{2p}\right]/E\left[S^p\right]$, with S^p associated with $|s(n)|^p$, $p \in \mathbb{Z}$ and $E[\cdot]$ is the statisti-

cal expectation operator. The last equality comes from the fact that the entropy does not depend on the mean of the RV.

As a standard procedure within the ITL framework, the PDF associated with Y^p, $f_{Y^p}(v)$, is estimated according to the Parzen window method [21], a kernel-based approach [22]. Assuming Gaussian kernel functions [22], defined as

$$G_{\sigma^2}(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-|v|^2}{2\sigma^2}\right),$$ (5)

where σ is the kernel size and v is a real- or complex-valued variable, this method is particularly convenient for the $\alpha = 2$ Rényi's entropy (or the quadratic Rényi's entropy), which results in a simplified estimated cost for RD:

$$\hat{J}_{\text{RD}}(\mathbf{w}) = \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{2\sigma^2}(|y(n-j)|^p - |y(n-i)|^p)$$ (6)

where L is the size of the samples window used for estimation. Since Eq. (6) does not consider the negative logarithm present in the Rényi's entropy definition, the objective of the RD criterion is the maximization of the cost $\hat{J}_{\text{RD}}(\mathbf{w})$ [24]. This cost is also referred to as quadratic *Information Potential* (IP) [22].

For the cases in which the transmitted signals are continuous, it is possible to consider as a criterion the minimization of $H_\alpha(Y)$ [4], which will be called here *Rényi's entropy for Continuous sources* (RC). Again, the attractive simplicity of choosing $\alpha = 2$ and the use of Gaussian kernels in the Parzen window method lead to the following estimate for the cost function:

$$\hat{J}_{\text{RC}}(\mathbf{w}) = \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{2\sigma^2}(y(n-j) - y(n-i)),$$ (7)

which should be maximized, as the negative logarithm of entropy is not considered. The RV Y can be either real- or complex-valued.

The RD and RC criteria make use only of the statistical information brought by the filter output signal $y(n)$, and as a consequence, it is necessary to avoid convergence toward the trivial solution, which can be done through the application of a constraint over the equalizer, such as keeping one of the taps equal to one [24].

3.2 Correntropy

Other emblematic entity in the context of unsupervised ITL is the measure called correntropy, which can be viewed as a generalized correlation function [22,26]. It takes into account not only the statistical distribution of signals but also their time structure, which is particularly useful when treating signals with statistical dependence. It may be defined as

$$v_Y(m) = E\left[\kappa(Y_n, Y_{n-m})\right], \tag{8}$$

where $\kappa(\cdot, \cdot)$ denotes a positive definite kernel function and m is the time delay between samples. As in the Rényi's entropy class, the Gaussian kernel is usually considered, and a sample mean approximates the statistical expectation in Eq. (8), resulting in

$$\hat{v}_Y(m) = \frac{1}{L-m+1} \sum_{n=m}^{L} G_{\sigma^2}(y(n) - y(n-m)). \tag{9}$$

In this case, an interesting criterion for blind equalization is the minimization of the following cost function [26]:

$$J_{\text{corr}}(\mathbf{w}) = \sum_{m=1}^{P} (v_S(m) - v_Y(m))^2, \tag{10}$$

where v_S is the correntropy of the source, v_Y is the correntropy of the equalizer output, and P is the number of lags being considered. Equation (10) can be seen as a matching of correntropies. It is usually assumed that $v_S(m)$ is analytically obtained, while $v_Y(m)$ is replaced by its estimate $\hat{v}_Y(m)$.

Other approaches considering correntropy can be found in the literature, like [9, 16, 17]; however, as we are considering here the specific problem of blind channel equalization, the only criterion to be studied in this class is that of minimizing the cost given by Eq. (10).

3.3 PDF Matching

The third class within the unsupervised ITL framework is associated with the BGR theorem [1], which basically states that a channel can be successfully equalized if the distribution associated with $s(n)$ is identical to that of $y(n)$. In that sense, the set of criteria belonging to this class aims at the matching of distributions, which is done mainly under the perspective of a quadratic divergence (QD) measure [13, 25]. Interestingly, this idea also establishes connections with the notion of mutual information [22].

In the sequence, we will present the set of criteria belonging to this ITL class, which is much more diversified than the previous ones. We start with the classical QD criterion, then discuss its extensions/variations, and finally, we present the criteria which have the peculiarity of not considering the modulus of the signals—including a variation that is one of the contributions of this work.

3.3.1 The Quadratic Divergence Between Distributions

The most generic notion of a PDF matching by means of the QD measure is the criterion proposed in [13]:

$$J_{\text{QD}}(\mathbf{w}) = \int (f_{Y^p}(v) - f_{S^p}(v))^2 \, dv$$
$$= \int f_{Y^p}^2(v)dv + \int f_{S^p}^2(v)dv - 2 \int f_{Y^p}(v)f_{S^p}(v)dv, \tag{11}$$

where $f_{Y^p}(v)$ and $f_{S^p}(v)$ are the PDFs associated with the RVs $Y^p = \{|y(n)|^p\}$ and $S^p = \{|s(n)|^p\}$, respectively. Note that this method considers the modulus of the symbols to the power p, similarly to the p-order dispersion [8] discussed in Sect. 3.1. Also, it is important to mention that the second term after the last equality of Eq. (11) is generally disregarded, since it is assumed that $f_{S^p}(v)$ is a target PDF and remains fixed during the filter adaptation process; we will follow this assumption here.

In the equalization problem, the PDF $f_{Y^p}(v)$ is unknown and must be estimated. Traditionally, this is done through the Parzen window method considering a Gaussian kernel, which is particularly useful when computing the product between PDFs [22]—this same idea is employed in the RD and RC criteria. As for $f_{S^p}(v)$, since S^p is generally a discrete RV in communication problems, this PDF is considered as the convolution between the Gaussian kernel and the discrete distribution of the source $p_S(v)$ [13]:

$$\hat{f}_{S^p}(v) = \sum_{i \in \mathcal{A}} p_S(s_i) G_\sigma(v - |s_i|^p) \tag{12}$$

where $\mathcal{A}$ is the alphabet of all possible occurrences of the RV S—associated with $s(n)$—and s_i is the ith symbol $\in \mathcal{A}$. Based on this, the QD cost function can be estimated as

$$\hat{J}_{\text{QD}}(\mathbf{w}) = \frac{1}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} G_{2\sigma^2}(|y(n-j)|^p - |y(n-i)|^p)$$
$$- \frac{2}{LN_{\mathcal{A}}} \sum_{i \in \mathcal{A}} \sum_{j=0}^{L-1} G_{2\sigma^2}(|y(n-j)|^p - |s_i|^p), \tag{13}$$

where $N_{\mathcal{A}}$ is the number of elements in $\mathcal{A}$. Note that we have not considered the second term of Eq. (11) for obtaining (13).

The extension of this idea allowed the emergence of some extensions/variations, such as the matching by sampling points (SPs) [14] and the reduced PDF matching [13,15], which will differ in the computational cost, algorithmic robustness and other features that we will discuss ahead.

3.3.2 Extended Formulations of the QD Cost Function

Other PDF matching-based criteria make additional simplifications over the QD criterion given by Eq. (11). Basically, there are three main variations of this concept, as shown in the following.

In the work of Santamaría et al. [25], instead of using the PDF $f_{S^p}(v)$, the authors suggest the use of an approximated distribution $\hat{f}_{S^p}(v) = G_\sigma(v - \hat{R}_p)$ in Eq. (11), which basically results in a change in the last term of the cost estimation (13), being

$|s_i|^p$ replaced by a constant named $\hat{R}_p$. Since this method encompasses a constant parameter to establish the matching of PDFs, we refer to it as QD-C. It is important to mention that the QD-C will only differ from QD if the modulus to the power p of the $s(n)$ symbols are not constant; otherwise, these methods will coincide.

The second alternative formulation of the QD cost can be achieved by considering a sampled version of the distributions $f_{Y^p}(v)$ and $f_{S^p}(v)$ [14]. In this case, the PDF $f_{Y^p}(v)$ is estimated at sampling points (SPs) v_i and they should match the corresponding values of $f_{S^p}(v)$ at each SP. Mathematically, the cost function can be written as

$$J_{\text{QD−SP}}(\mathbf{w}) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left(f_{Y^p}(v_i) - f_{S^p}(v_i) \right)^2, \qquad (14)$$

where N_s is the number of sampling points. It is important to observe that since there is no integration, the use of other kernels different from the Gaussian for PDF estimation is also suitable for gradient computation [14]. Hence, the estimated cost function becomes

$$\hat{J}_{\text{QD−SP}}(\mathbf{w}) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left(\frac{1}{L} \sum_{j=0}^{L-1} \kappa_\sigma \left(v_i - |y(n-j)|^p \right) - \frac{1}{N_{\mathcal{A}}} \sum_{k \in \mathcal{A}} \kappa_\sigma \left(v_i - |s_k|^p \right) \right)^2 \qquad (15)$$

Finally, the third variation of the QD measure is based on the idea that the quadratic distance between the two PDFs is in fact measured by the last term of Eq. (11), which is referred to as *cross-IP*; the other terms work just as normalization factors and can be disregarded in the cost [13,15,22]. Thus, it is possible to consider a reduced cost function of the form:

$$J_{\text{QD−R}}(\mathbf{w}) = \int f_{Y^p}(v) f_{S^p}(v) \mathrm{d}v. \qquad (16)$$

Following the canonical approach, i.e., considering the Parzen window method and Gaussian kernels, the estimated cost function may be written as:

$$\hat{J}_{\text{QD−R}}(\mathbf{w}) = \frac{1}{L N_{\mathcal{A}}} \sum_{i \in \mathcal{A}} \sum_{j=0}^{L-1} G_{2\sigma^2} \left(|y(n-j)|^p - |s_i|^p \right) \qquad (17)$$

which is basically the last term of (13). Note, however, that the cost function (17) should be maximized, while (13) should be minimized. As a curiosity, the cost function (17) can also be intuitively formulated within a digital communication scenario under the assumption of Gaussian noise disturbances, as pointed out in [18].

The QD-C, QD-SP and QD-R criteria presented here form a set of alternative QD-based methods, whose main objective is to achieve greater mathematical simplicity than that of the original QD criterion, defined in Eq. (11). However, the variations of the QD is not limited by them, as will be described next.

3.3.3 The Direct Matching of PDFs

The common feature of this PDF matching subclass is the non-application of the modulus (to the power p) to the considered RVs, which, as will be discussed, allows the adoption of the criteria in a broad set of scenarios. Under this perspective, there are two possible approaches concerning the distribution of the source $s(n)$: (i) it is assumed that the distribution associated with the RV S can be approximated by a combination of Gaussian kernels, as is similarly done in Eq. (12), allowing the direct application of the Parzen window method for PDF estimation [11,12], and (ii) it is assumed that the RV S is associated with a probability mass function (PMF)—i.e., a discrete source—and a matching between PDF and PMF must be considered [6], which leads to a reduction in the complexity and reveals itself to be a very important link to other unsupervised ITL criteria. Although the authors assume multivariate distributions in [6] (this approach is very suitable when dealing with statistically dependent sources), in this work, we consider a slight modification of the criterion proposed in [6] by assuming only univariate distributions, as presented in the following. This new criterion, along with other density matching criteria, will be discussed in the sequel.

3.3.4 The Matching of Two PDFs

As previously mentioned, the notion of PDF matching can be directly employed in the variables Y and S—associated with the signals $y(n)$ and $s(n)$, respectively, without resorting to the application of the modulus. Following the same concept of measuring (dis)similarity via the QD, we obtain the resulting cost function [11,12]:

$$J_{\text{QD-D}}(\mathbf{w}) = \int f_Y^2(v)\mathrm{d}v + \int f_S^2(v)\mathrm{d}v - 2\int f_Y(v)f_S(v)\mathrm{d}v, \qquad (18)$$

where, again, we will not consider the second term. Very interestingly, by adopting the RVs Y and S instead of Y^p and S^p, we are not privileging any type of distribution, hence either continuous or discrete distributions can be considered.

In order to estimate the cost in (18), the first approach points toward the use of the Parzen window method for estimating both the distributions, and the result is very similar to Eq. (13) (but with the direct samples of $y(n)$ and $s(n)$—Eq. (12)—and not their modulus to the power p) [11,12].

3.3.5 The Matching Between PDF and PMF

An alternative view bearing the very notion of the matching between a PDF and a PMF [6] is assumed in the second approach, being able to provide a simpler estimation method when dealing with discrete sources (in a certain sense, the idea is closely related to the sampled PDF given by Eq. (14)). Here, we modified the criterion proposed in [6] to univariate distributions. The estimated cost becomes

$$\hat{J}_{\text{QD--D}}(\mathbf{w}) = \int \left(\hat{f}_Y(v) - p_S(v) \right)^2 \, dv$$

$$= \frac{1}{L^2} \sum_{i=1}^{L} \sum_{j=1}^{L} G_{2\sigma} \left(y(n-i) - y(n-j) \right) \tag{19}$$

$$- \frac{2}{L} \sum_{i \in \mathcal{A}} \left[p_S(s_i) \left(\sum_{j=1}^{L} G_\sigma \left(s_i - y(n-j) \right) \right) \right],$$

where $p_S(v)$ is the PMF associated with the RV S—note that we have not considered the term involving the square of $p_S(v)$.

The source PMF can be estimated through samples—e.g., via histogram-based methods—or assumed to be known. This criterion will be adopted here when the source is discrete. For continuous sources, the previous approach—the matching of two PDFs—will be used. Both of them will be referred to as QD-D and will be distinguished according to the type of source (continuous/discrete).

3.3.6 Novel Simplified Matching Criterion

Similarly to the relation between the QD and QD-R, the term that really measures the difference between the two PDFs (or the PDF and the PMF in the case of discrete sources) in the QD-D criterion is given by the last term of Eq. (18). Thus, we propose to use, as a criterion, only this last term—the term correspondent to the cross-IP, as in Eq. (16). The resulting reduced cost function will be referred to as QD-DR. Just like the QD-R, we discard the constant factor -2 so that the QD-DR cost must be maximized. Similarly to the QD-D criterion, we will also consider two QD-DR versions, one for continuous sources and the other for discrete ones. For the continuous case, the cost is similar to Eq. (17)—but with the direct use of the signals $y(n)$ and $s(n)$ and not their modulus—and, for the discrete case, the estimated cost function will be given by:

$$\hat{J}_{\text{QD--}DR}(\mathbf{w}) = \frac{1}{L} \sum_{i \in \mathcal{A}} \left[p_S(s_i) \left(\sum_{j=1}^{L} G_\sigma \left(s_i - y(n-j) \right) \right) \right]. \tag{20}$$

The continuous or discrete cost will be distinguished according to the type of source.

In general terms, since the criteria presented in this topic share the common notion of the matching of distributions, it is expected that their solutions occur at the same filter coefficients values (or, at least, within a small neighborhood), as will be investigated later in the next sections.

3.4 Summary of the ITL Criteria

In summary, there are three main classes of the ITL criteria applied to the blind channel equalization problem: the entropy-, the correntropy- and the density matching-based criteria. Due to the number of criteria encompassed in this study, we provide a brief

Table 1 Summary of the ITL criteria for blind channel equalization

Class	Criteria	Acronym	Equation
Entropy	Rényi's entropy of the p-order dispersion	RD	(6)
	Rényi's entropy for continuous sources	RC	(7)
Correntropy	Matching of correntropies	corr	(10)
Distribution matching	The quadratic div. between distributions	QD	(13)
	The QD-constant approximation	QD-C	–
	The QD at samples points	QD-SP	(15)
	The QD-reduced cost	QD-R	(17)
	The QD-direct matching	QD-D	(19)
	The QD-direct matching Reduced	QD-DR	(20)

list in Table 1 with all of them, including their acronyms and estimated cost equation number.

In order to investigate these classes of methods, we will analyze in the following the relationship between the ITL criteria and the classical approaches and also among themselves, including a cost surface analysis. The aim of such analysis is to help improve the understanding of the ITL framework as a whole, something that still is lacking in the literature. Then, a performance comparison will be held in distinct scenarios for equalization.

4 Connections with the Classical Approaches

The unsupervised ITL criteria raise interesting possibilities for extracting the statistical information about the variables of interest in a more extensive manner. Nevertheless, as we intend to show, they also preserve strong connections with the classical equalization theory, specially the constant modulus (CM) criterion. In that sense, we present in this section some of these relations, which shall be of use to the improvement in our understanding of these ITL criteria.

We start by the RD criterion, defined by Eq. (4), which presents a straightforward connection to the p-order dispersion criterion in its strict formulation. The adoption of Rényi's entropy, in this case, is the key factor that contributes to the use of a richer statistical content. For the continuous counterpart, the relationship between the RC criterion and the classical framework is not so evident, but it shows itself clearer under certain considerations. Specifically, from Eq. (7), if we assume expectation operators instead of the sample mean and through a second-order Taylor series expansion around zero of the Gaussian kernel, we obtain:

$$\hat{J}_{\text{RC}}(\mathbf{w}) \approx \frac{1}{\sqrt{4\pi\sigma^2}} \left(1 - \frac{1}{2\sigma^2} \left(\sigma_Y^2 - r_Y \right) \right) \tag{21}$$

where σ_Y^2 is the variance of Y and r_Y is the correlation for all time differences, i.e., $E\left[Y_{n-i}Y_{n-j}^*\right]$ for $i, j \in \mathbb{Z}$ (for the sake of simplicity, we dropped the notation of the cost dependence on $\mathbf{w}$). Since a scale factor and an additive constant do not impact on the optimal points, it is possible to state that $\hat{J}_{RC}(\mathbf{w})$ is proportional to the correlation function r_Y, and, as the RC cost should be maximized, it results that r_Y should be maximized as well. In that sense, from the classical standpoint, the RC criterion can produce a signal $y(n)$ as similar as possible to itself, but also causes the samples to be colored.

For the correntropy, similarly to the RC criterion, it can be shown that its second-order Taylor series expansion around zero results in a term proportional to the negative classical correlation [5,22], i.e., $v_Y(m) \approx -r_Y(m)$; however, it is able to consider arbitrary time differences m. Additionally, considering the specific case in which $m = 0$, $\mu_Y = E[Y] = 0$ and the instantaneous estimate $r_Y(0) \approx y(n)y^*(n) = |y(n)|^2$, the correntropy-based criterion given by (10) becomes $(|y(n)|^2 - r_S(0))^2$, which is very similar to the stochastic approach used in the CM algorithm (CMA) [5]—except for the constant term, which is now the source correlation $r_S(0)$ (under certain circumstances, they can be equal). It is true that the correntropy-based criterion does not even consider the delay $m = 0$ in its formulation, even though this standpoint reveals certain degree of similarity to the CM cost and how the signal time structure is used.

In the third class of the blind ITL criteria, the most generic criterion based on the QD measure can be represented by Eq. (13), which is composed of two terms. The first one is basically the estimate of the RD criterion, Eq. (6), which, as we mentioned, refers to the p-order dispersion criterion in its strict definition. For the second term, the connection with the classical framework can be made more explicit through the following assumptions: in Eq. (13), $L = 1$ and $|s_i|^p = \hat{R}_s$, for all $i \in \mathcal{A}$; in this case, the second term of (13) reduces to

$$-2G_{2\sigma^2}(|y(n)|^p - \hat{R}_s) \approx \frac{1}{\sigma\sqrt{\pi}}\left(\frac{(|y(n)|^p - \hat{R}_s)^2}{4\sigma^2} - 1\right), \qquad (22)$$

where the approximation is the result of the second-order Taylor series expansion around zero. Since neither a scale factor nor a constant subtraction affects the minima, if $\hat{R}_s = R_p$, then the second-order approximation leads exactly to the same possible solutions of the p-order dispersion algorithm—the same assertion can be held for the CM criterion when $p = 2$.

Regarding the extended formulations of the QD measure, it is also possible to make similar comparisons: for the QD-C criterion [25], the analysis is even more straightforward, since the assumption of $|s(n)|^p = \hat{R}_s$ is already considered in the cost formulation; for the QD-SP criterion [14], which also has the two mentioned terms, the only difference remains at the SPs v_i; however, considering that v_i is the representation of a chosen sample of the RV Y^p, the comparison remains the same, and finally, the QD-R criterion [13,15] is basically the last term of the estimate of QD criterion, and no further considerations are necessary.

The last blind ITL case to be analyzed is the QD-D criterion, which, differently from the previous approaches, does not raise points of contact with the p-order dis-

person framework, at least from the adopted perspective. However, a closer look at the two terms of Eq. (19) can also reveal very attractive relationships with the classical approach. The first term is identical to the estimate of the RC criterion, which is in turn related to the correlation function. The second term, which gives rise to the proposed QD-DR criterion, very curiously, can be linked to the mean-squared error (MSE) method, a supervised approach. To understand this, we will assume a single sample of $y(n)$ for PDF estimation, and, again, Taylor series expansion, which results in:

$$ -2E\left[G_\sigma\left(s_i - y(n)\right)\right] \approx \frac{-2}{\sqrt{2\pi}\,\sigma}\left(1 - \frac{1}{2\sigma^2}E\left[|s_i - y(n)|^2\right]\right), \qquad (23) $$

for $i \in \mathcal{A}$; this is similar to the mean-squared error signal, although it encompasses all possible occurrences of $s(n)$ in the expectation operator—represented by the symbols s_i. Indeed, the efficacy of this approach lies in the proximity between samples $y(n)$ and $s(n)$—notice that the transmitted sequence $s(n)$ is not known, only the probability structure of its symbols.

So far, the search for connections with the classical equalization theory contributed to better elucidate the modus operandi of each blind ITL criterion—since the classical approach, which was extensively studied along decades, provides more intuitive tools—however, as we intend to show in the next section, a deeper understanding can be achieved by studying the relationship among these criteria.

5 Blind ITL Criteria Interconnections

Although the classes of blind ITL criteria employ distinct concepts to deal with the statistical information, they still present some common features—as also suggested by the comparison to the classical equalization theory—that can be explored to find analytical relations among them.

In some cases, we will provide illustrations of the surface contours of the cost functions, which will be result of the simulations held in a scenario with a binary phase shift keying (BPSK) source signal, a channel with transfer function $H(z) = 1 + 0.6z^{-1}$ and an equalizer with two coefficients, w_0 and w_1—for the sake of visualization. The parameters assumed for all ITL criteria were $\sigma = 0.65$ and $L = 100$. Observe that this simple scenario is not sufficient to obtain general conclusions, but it might give some important intuition on the interconnections. The comparison in more complex scenario will be carried in the next section.

A straightforward relationship is one among the costs associated with the estimated versions of the RD, QD and QD-R criteria, represented by Eqs. (6), (13) and (17), respectively, which can be expressed as

$$ \hat{J}_{\text{QD}} = \hat{J}_{\text{RD}} - 2\hat{J}_{\text{QD-R}}. \qquad (24) $$

A similar expression was presented in [25], but the QD-R cost was not considered. In that sense, Eq. (24) allows us to make additional remarks. Firstly, we have that $\hat{J}_{\text{QD}}$ must be minimized, which, taking as a reference Eq. (24), implies in the minimization of $\hat{J}_{\text{RD}}$ and in the maximization of $\hat{J}_{\text{QD-R}}$ (due to the negative sign). However, in
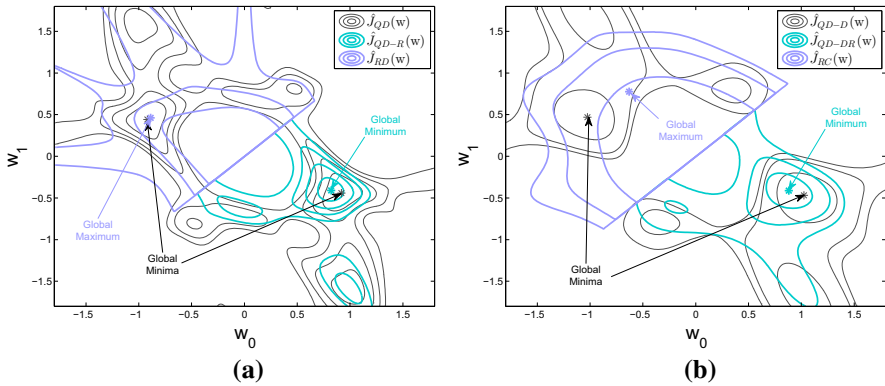
Fig. 2 Surface contours. **a** Surface contours of the QD, QD-R and RD costs. **b** Surface contours of the QD-D, QD-DR and RC costs

the original context, the cost $\hat{J}_{RD}$—Eq. (6)—must be maximized, which indicates that this cost plays a different role in (24): it points toward decorrelated samples. Hence, regarding the relationship between the criteria QD and QD-R, the cost $\hat{J}_{RD}$ will contribute, to a greater or lesser extent, to modifications in the surface, so that the optimum points of QD and QD-R are not coincident. Notwithstanding, it is expected that their solutions be as close as possible, since both of them aim at the matching of PDFs. In Fig. 2a, we illustrate the contours and solutions[1] of the three mentioned cost functions (for the sake of visualization, we exhibited only half of the QD-R and RD costs—due to their symmetry along the axis $w_0 = w_1$, there is no loss of information). It is possible to observe that the costs are multimodal and, indeed, their global solutions are close. However, the same does not hold for local solutions. For this reason, the surface $\hat{J}_{QD}$—that is the combination of both $\hat{J}_{QD-R}$ and $\hat{J}_{RD}$—presents several local minima. Hence, additional care is necessary when using gradient-based methods for optimization. We remark that the global solutions of these criteria will only coincide when the zero-forcing (ZF) solution is achieved, i.e., when the channel is completely equalized. Finally, Eq. (24) also highlights that the computational complexity of the QD cost is higher than the RD or the QD-R costs, since it includes both of them.

A situation similar to Eq. (24) can be outlined, but involving the QD-D estimated cost function. Basically, this cost can be related to the RC and the proposed QD-DR estimated costs in the following manner:

$$\hat{J}_{QD-D} = \hat{J}_{RC} - 2\hat{J}_{QD-DR}, \tag{25}$$

and the comparison is analogous to that presented for Eq. (24). As illustrated in Fig. 2b, the QD-D and QD-DR present surfaces similar to that of the QD and QD-R, with points of maxima and minima close to each other. However, in contrast, the RC surface is clearly different from that of QD-D, with global solutions relatively distant from each

---

[1] The global maxima of the RD and RC criteria were obtained considering a unity norm restriction over the filter coefficients.
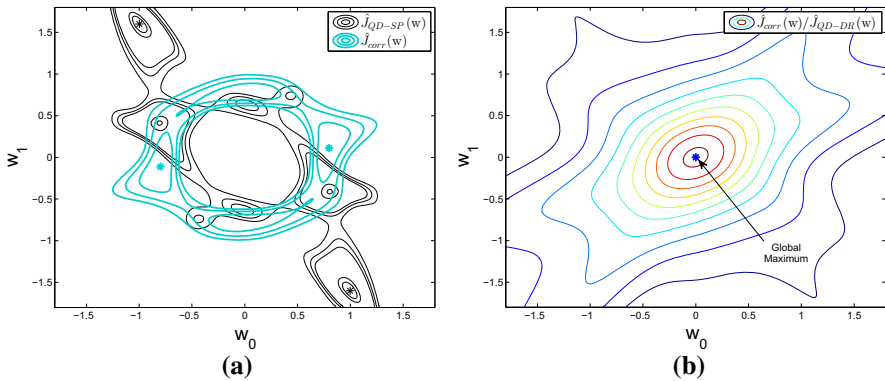
Fig. 3 Comparison of surface contours. **a** Surface contours of the QD-SP and the correntropy-based costs. **b** Surface contour of correntropy cost or (the specific case) of QD-DR cost

other. Indeed, from Eq. (25), we know that the QD-D is much more similar to QD-DR due to its greater weight (by a factor of two).

The resemblance between the correntropy-based criterion, Eq. (10), and the estimated QD-SP, Eq. (15), is also worth mentioning, as it can clarify the points of contact with the notions of matching of PDFs and correntropy. To illustrate, we will consider the hypothetical case in which only one SP is used in the QD-SP, assumed to be $v_i = |y(n)|^p = |s(n)|^p$ (being used $v_i = |y(n)|^p$ in the first term and $v_i = |s(n)|^p$ in the second term inside parentheses in Eq. (15)). In such case, the estimated QD-SP criterion can be expressed as

$$\hat{J}_{\text{QD-SP}} = \left(\frac{1}{L} \sum_{j=0}^{L-1} \kappa_\sigma (|y(n)|^p - |y(n-j)|^p) - \frac{1}{N_\mathcal{A}} \sum_{k\in\mathcal{A}} \kappa_\sigma (|s(n)|^p - |s_k|^p) \right)^2$$

$$= \left(\frac{1}{L} \sum_{j=0}^{L-1} \hat{v}_{Y^p}(j) - \frac{1}{N_\mathcal{A}} \sum_{k\in\mathcal{A}} \hat{v}_{S^p}(k) \right)^2, \tag{26}$$

which can be interpreted as the matching of the average correntropy with respect to the RVs Y^p and S^p. Although the correntropy-based criterion does not consider variables to the power p and computes differences in each delay m, in a certain sense, both criteria make use of a common framework. In Fig. 3a, we illustrate the surface contours of both criteria (we assumed $P = 3$ for correntropy). It is possible to note that the use of the modulus of the signal and the different delays at the same time are determinant to obtain completely different surfaces. In addition to that, we can say from Fig. 3a that the QD-SP, as a PDF matching-based criterion, is much more alike to the QD than the correntropy-based criterion.

Finally, the proposed criterion QD-DR can also be related to the correntropy. As explained in [6], the second term of (19) can correspond to a more generic version of correntropy: by assuming, for example, the matching of a PDF associated with the signal $u(n) = y(n) - y(n-m)$ and a PMF associated with the constant signal

$s(n) = 0$ (i.e., a null signal with $p_S(0) = 1$), the QD-DR will reduce to the correntropy. Mathematically,

$$\hat{J}_{\text{QD-DR}} = \frac{1}{L} \sum_{j=1}^{L} G_\sigma \left(y(j) - y(j - m) \right)$$
$$= \hat{v}_Y(m), \tag{27}$$

which means that correntropy can also be viewed, from the standpoint of the QD-DR criterion, as a measure of the similarity between the density associated with $u(n) = y(n) - y(n - m)$ and the discrete distribution associated with a null signal, similarly as occurs with an error signal. In that sense, increasing the correntropy will make the signal $y(n)$ be constant as $u(n)$ approaches zero. In Fig. 3b, we illustrate the surface contours of the estimated correntropy for the equalizer output (and $m = 1$, i.e., $\hat{v}_Y(1)$), which is equivalent to the $\hat{J}_{\text{QD-DR}}$ cost for the conditions established in Eq. (27). Note that the global maximum occurs for null weights of $\mathbf{w}$, corresponding to the case in which the signal $y(n)$ is null and, consequently, $u(n) = y(n) - y(n - 1)$ is null as well. Hence, the distributions $p_U(v)$ and $p_S(v)$ are equivalent and the correntropy is maximum, agreeing with the new presented perspective.

6 Performance Analysis

In this section, a performance comparison among the presented criteria will be carried out considering four equalization scenarios and two optimization methods, viz., a stochastic gradient-based approach and an evolutionary search. Traditionally, the gradient method is employed along with ITL criteria due to its lower computational cost. However, as indicated in the previous section, the potential existence of local optima can possibly lead this adaptation method to converge to suboptimal solutions, depending on the parameter initialization. With this in mind, the second option—the evolutionary search—shows to be more robust against local convergence, but is computationally more costly, depending on the size of the search space, which grows in function of the number of coefficients in the equalizer. Undoubtedly, a less costly approach would be the adoption of several different initialization points for the gradient-based algorithms, but, given the great computational power of nowadays, the evolutionary search shows to be suitable for the analysis. The evolutionary algorithm to be employed in the simulations is the differential evolution (DE) [29], a metaheuristic whose population adaptation operators strictly use the information available in the current solution candidates, instead of the classical random perturbations.

Performance evaluation, in practical blind equalization, is usually made with measures regarding the equalized samples, e.g., through an eye diagram analysis. However, in our case, to exhibit a detailed profile of the criteria performance, we will count on two ISI-based metrics, which are representative measures of how much ISI remains in the equalization process. Usually, the quadratic ISI (QISI) measure is employed
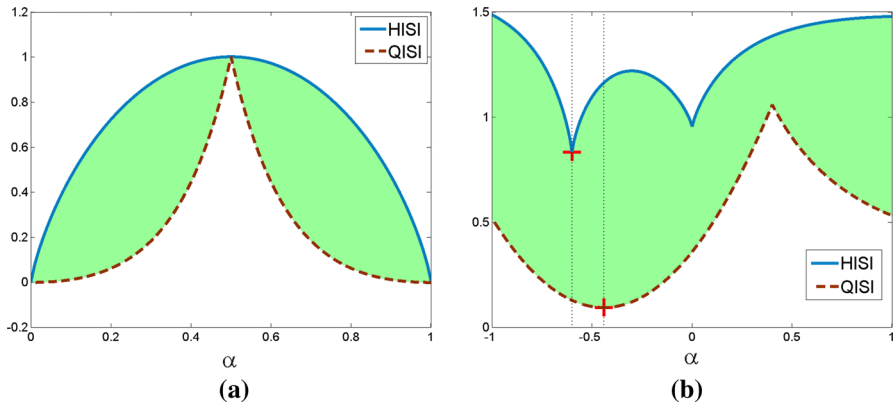
Fig. 4 Measured values for QISI and HISI. **a** Combined channel+equalizer impulse response. **b** Equalizer impulse response

[14,15], defined as

$$\text{QISI}_{\text{dB}} = 10 \log_{10} \frac{\left(\sum_{i=0}^{M_c} |c_i|^2\right) - \max_j |c_j|^2}{\max_j |c_j|^2}, \tag{28}$$

where $\mathbf{c} = [c_0 \ c_1 \ \ldots \ c_{M_c}]^{\text{T}}$ is the combined channel + equalizer impulse response, with length M_c+1, but a promising ISI measure capable of considering the HOS involved in the process is the entropy-based ISI (HISI) [20], defined as

$$\text{HISI}_{\text{dB}} = 10 \log_{10} \left(-\sum_{i=0}^{M_c} \alpha|c_i|\log_2 (\alpha|c_i|)\right), \tag{29}$$

where $\alpha = 1/\sum_i |c_i|$. By considering a richer statistical content, the HISI measure shows to be more adequate for scenarios with non-Gaussian distributed sources and low noise interference. To illustrate the HISI and the QISI features, we consider below two cases: Firstly, the equalizer can achieve the ZF condition, and, secondly, when the ZF condition is not attainable.

When a ZF solution is attainable, both QISI and HISI provide similar performance measures. Considering the channel+equalizer impulse response $\mathbf{c} = [\alpha \ 1-\alpha]^{\text{T}}$, with α varying from 0 to 1, we obtain the measured values for QISI and HISI shown in Fig. 4a, where it is possible to note that the optimal (minimal) values for both QISI and HISI happen at $\alpha = 0$ and $\alpha = 1$ (the ZF solutions). However, for intermediate α values, the HISI measure weights the ISI effect differently from the QISI measure.

When the ZF solutions are not attainable, the difference between HISI and QISI becomes sharper. We consider the case in which the channel has an impulse response $H(z) = 1 + 0.6z^{-1}$, and the equalizer, $W(z) = 1 + \alpha z^{-1}$. By varying α from -1 to $+1$, we obtain the measures of QISI and HISI presented in Fig. 4b. Now, the optimum value for QISI is different from that of HISI (denoted by red crosses in the figure).
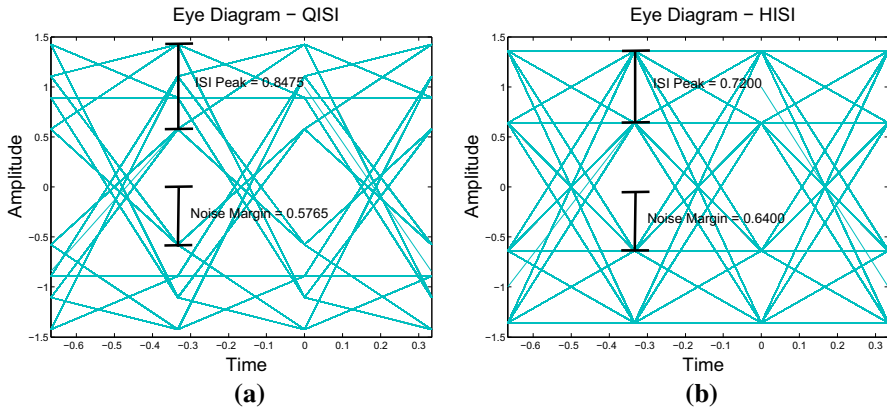
Fig. 5 Eye diagram for the optimal cases of QISI and of HISI. **a** Optimum QISI. **b** Optimum HISI

If it is considered that the source is a BPSK modulated signal, the performance of each optimum in terms of the QISI and the HISI can be evaluated considering the classical eye diagram, as shown in Fig. 5. It is clear that both of them lead to a open-eyed solution for $W(z)$; however, the ISI peak measured in the optimum QISI case (0.8475) is higher than the HISI case (0.7200), and the noise margin is narrower in the QISI case (0.5765 for QISI and 0.6400 for HISI). This indicates that, for this case, the HISI measure can be more adequate as a performance measure instead of the QISI. Indeed, in [20], it is shown that the HISI can be an interesting performance measure when the source is not Gaussian, e.g., for sparse and uniform sources. For the Gaussian case, the classical QISI is sufficient. In view of this, we will make use of both these measures, always considering the scenario at hand.

Other key point to be considered in simulations is the parameter adjustment procedure for the ITL criteria and the optimization methods, as described in the sequence.

6.1 ITL Criteria Parameters

The common parameters of the ITL criteria are the kernel size σ and the window length L. The additional parameters are the number of lags P in the correntropy-based criterion, and the dispersion p for RD, QD, QD-C, QD-SP and QD-R.

The kernel size σ, in special, can significantly change the behavior of the criteria, as it has the capability of modifying the cost surface. Basically, it is known that an increase in the σ value has a smoothing effect on the cost function, which, on the one hand, can contribute to reducing the number of local optima and to a faster convergence, but, on the other hand, reduces the precision of the solutions, causing the ISI to increase [13,22].

In order to obtain the best performance, the adjustment of σ for each criterion and scenario was made through a line search—in the range from 0 to 2 (with resolution of 0.01)—by measuring the HISI associated with the best individual obtained by the DE (it was assumed that there was a set of test samples and DE parameters suitable for each

scenario—as will be described next). The chosen value of σ was that corresponding to the lowest measured HISI.

With respect to the window length, the higher the (integer) value of L, the better the cost estimation [21], but also the higher the computational cost. In that sense, there is a trade-off. To establish a fair comparison among criteria, we adopted a single value of L for all criteria in each scenario. Naturally, the number of samples L depends on the distributions of the source and the equalized signal—continuous distributions, for instance, tend to demand a larger number of samples. In our simulations, we considered a few hundred of samples, depending on the scenario.

The number of lags P and the order of dispersion p were chosen according to the usual practice in the channel equalization task, being P a number about 10 delays [26] and $p = 2$ [13–15,24].

Finally, the source-related variables—like the source correntropy $v_S(m)$, the modulus of symbols $|s(n)|^2$ or the source probability $p_S(v)$—are all assumed to be known and their analytic values are considered in the simulations.

6.2 Parameters of The Optimization Methods

The optimization methods have parameters to be adjusted as well. In the case of the gradient-based search, it is necessary to adjust the step size μ; for the metaheuristic DE, the parameters are: the population size N_p, the population adaptation step size F, the combination rate CR and the number of iterations [29].

The step size μ of each algorithm was adjusted aiming at a fast convergence. However, to keep a fair comparison, the value of μ was restrained so that the mean displacement between iterations (in terms of the Euclidean distance—i.e., $||\mathbf{w}(n+1) - \mathbf{w}(n)||$) after convergence was approximately the same for all algorithms. It is also important to mention that the gradient-based algorithms for the criteria RD, RC, correntropy, QD, QD-C, QD-SP and QD-R can be found in [4,13–15,24–26]. The QD-D and QD-DR algorithms are presented in appendix, since the first is a modification of [6] and the second was developed in this work. In all gradient-based methods, we have not considered constant multiplicative factors on the gradient, as done in the original proposals of the algorithm.

At last, the DE parameters F and CR can assume fixed values for all scenarios, since the search space exploration pattern is basically the same. For all cases, we assumed $F = 0.5$ and $CR = 0.9$ and only changed the population size N_p and the number of iterations, according to the scenario at hand. These values were empirically chosen, observing the HISI performance of the best individual in the population.

6.3 First Scenario—BPSK Modulation

In the first simulation scenario, we consider the classical BPSK modulation, whose symbols belong to the alphabet $\{+1, -1\}$. The independent and identically distributed ($i.i.d.$) source is transmitted through the channel

$$H_1(z) = 0.2258 + 0.5161z^{-1} + 0.6452z^{-2} - 0.5161z^{-3}. \tag{30}$$
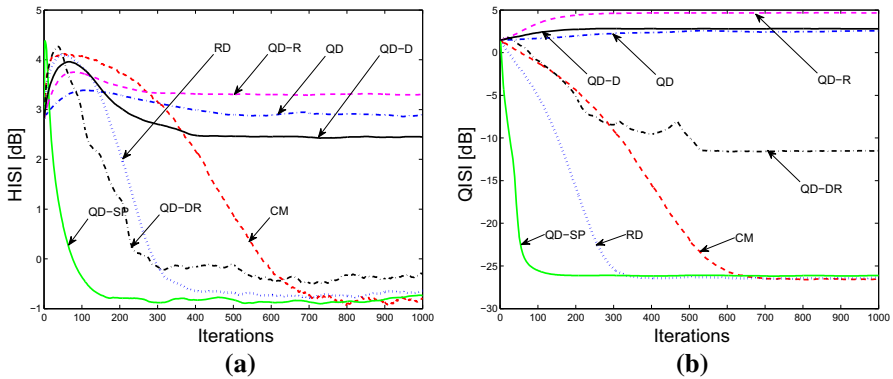
Fig. 6 Scenario 1—Average HISI/QISI performance for the gradient-based algorithms adaptation process. **a** HISI performance. **b** QISI performance

There is also the presence of additive white Gaussian noise (AWGN) with signal-to-noise ratio (SNR) level of 30 dB.

For the equalizer, we assumed an FIR filter with 9 coefficients (i.e., $M = 8$). The DE parameters were chosen to be $N_p = 100$ and 400 iterations (we considered $F = 0.5$ and $CR = 0.9$ fixed). As we deal with an $i.i.d.$ discrete source, the criteria RD, QD, QD-SP, QD-R, QD-D and QD-DR are suitable for this scenario (note that the criteria QD and QD-C are equivalent, since the source modulus is constant). On the other hand, since the RC are aimed at continuous sources and correntropy at dependent sources, they were disregarded. We adopted $L = 100$ for all analyzed criteria and, after performing a linear search for the parameter σ, we obtained the values $\sigma_{QD} = 20$, $\sigma_{QD} = 0.87$, $\sigma_{QD-SP} = 1.48$, $\sigma_{QD-R} = 0.47$, $\sigma_{QD-D} = 0.50$ and $\sigma_{QD-DR} = 0.46$. For the QD-SP criterion, we chose the SP $v_i = \{1\}$; for the RD criterion, we fixed the equalizer center tap at unity in order to avoid the convergence to the trivial solution. Next, the step size μ of each criteria was adjusted as previously described. The resulting values were: $\mu_{RD} = 200$, $\mu_{QD} = 0.01$, $\mu_{QD-SP} = 30$, $\mu_{QD-R} = 0.01$, $\mu_{QD-D} = 0.02$ and $\mu_{QD-DR} = 0.1$ (the mean Euclidean distance between iterations—after convergence—was 7×10^{-4}). We also considered the CM criterion (with $\mu_{CM} = 0.007$) due to its known good performance in this scenario.

After setting the parameters, we carried out 30 independent simulations and measured the performance in terms of HISI and QISI obtained by the gradient-based algorithms—using the center spike initialization [7]—and by the DE solution for each criterion. Figure 6 shows the average performance in the adaptation process and, in Table 2, the average performance of the best individual after applying the DE optimization method for the mentioned criteria. Starting with the analysis of the proposed QD-DR algorithm, Fig. 6a shows that it had a satisfactory performance, attaining an HISI value of -0.3948 dB while QD-SP, RD and CM attained about -0.8 dB. The QD-SP algorithm presented the fastest convergence, followed by QD-DR and RD algorithms. Indeed, one of the main features associated with the QD-SP algorithm is its fast convergence when compared to other algorithms [14]: The convergence occurred after 160 iterations, approximately, while for the QD-DR and RD—that are

Table 2 HISI of the best individual of the DE metaheuristic

Criterion	Scenario 1		Scenario 2	
	HISI [dB]	QISI [dB]	HISI [dB]	QISI [dB]
CM	-8.22×10^{-1}	-24.99	5.10	-0.22
RD	-6.17×10^{-1}	-26.26	4.81	-2.74
QD	-6.68×10^{-1}	-24.96	4.85	-2.35
QD-C	–	–	4.83	-2.10
QD-SP	-8.35×10^{-1}	-25.29	4.80	-2.42
QD-R	-4.42×10^{-1}	-24.08	4.27	-4.28
QD-D	-9.01×10^{-1}	-25.13	5.17	$+1.61$
QD-DR	-8.64×10^{-1}	-25.17	4.16	-0.30

also fast algorithms—needed about 300 and 400 iterations, respectively. The QD, QD-R and QD-D algorithms converged to solutions—probably local optima—associated with higher values of HISI. If analyzed in terms of QISI, Fig. 6b, the comparison, in general lines, is similar to that of HISI, being the QD-SP, RD and CM algorithms responsible for attaining the best performance. The main exception is the QD-DR, which, from the QISI standpoint, indicates a slow convergence and a greater distance from the other criteria in comparison with the HISI case.

Using the DE optimization method for the studied criteria, we observe from Table 2 that, in terms of HISI, not only the QD-DR but also the QD, QD-R and QD-D were able to improve their performances in comparison with their correspondent gradient algorithms, which indicates that they had converged to local optima. On the other hand, the CM, RD and QD-SP criteria obtained a similar performance. The best HISI performance in this case was that associated with the QD-D criterion, followed by the QD-DR, QD-SP and CM criteria. From the QISI perspective, the DE performance associated with the CM, QD, QD-SP, QD-D and QD-DR criteria are close (about -25 dB). Meanwhile, the RD criterion achieved the best QISI performance, and the QD-R the worst. It is important to emphasize that we consider the HISI performance a more suitable evaluation metric when dealing with non-Gaussian signals, as is the case of the source signal considered here [20].

6.4 Second Scenario—16-QAM Modulation and Impulsive Noise

Complex multilevel modulation schemes—such as 8-PAM, 16-QAM and 32-QAM—increase the information rate, but cause the equalization process to become more difficult. In this scenario, we will investigate the ITL criteria performance for an $i.i.d.$ 16-QAM modulated source; furthermore, we also consider the presence of impulsive noise, for which the ITL criteria are known to exhibit a certain robustness.

The channel is the same considered in the previous case, except for introducing a phase rotation of 45 degrees, i.e., $H_2(z) = H_1(z)e^{\frac{j\pi}{4}}$. The impulsive noise was mod-
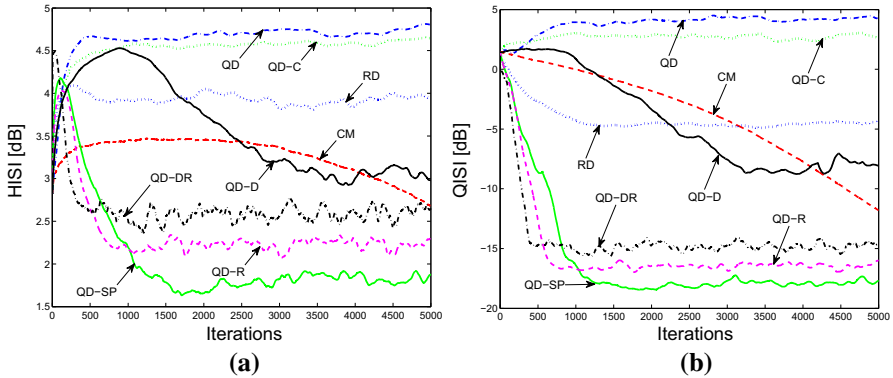
Fig. 7 Scenario 2—Average HISI/QISI performance for the gradient-based algorithms adaptation process. **a** HISI performance. **b** QISI performance

eled as the combination of two Gaussian distributions [26] with different variances. The probability of occurrence of the small variance Gaussian was set to 0.85, while the large variance Gaussian was 0.15. The variances were adjusted so that the resulting SNR was 22 dB.

The equalizer was assumed to have 7 complex coefficients, which, considering the real and complex parts, resulted in 14 coefficients to be adjusted. In view of this large search space, the used DE parameters were $N_p = 200$ and 500 iterations. The criteria considered for this scenario were the same of the previous one plus the QD-C. We adopted $L = 150$ for all, and from the linear search for σ, we obtained $\sigma_{RD} = 20$, $\sigma_{QD} = 0.77$, $\sigma_{QD-C} = 0.79$, $\sigma_{QD-SP} = 1.03$, $\sigma_{QD-R} = 1.02$, $\sigma_{QD-D} = 0.72$ and $\sigma_{QD-DR} = 0.36$. For the filter to be trained with the RD criterion, the center tap is fixed at 1; for the QD-SP criterion, we adopted the SPs {0.2, 1, 1.8} (since the transmitted symbols were $(\pm\{1, 3\} \pm \{1, 3\}j)/\sqrt{10}$), and for the QD-C, we set $\hat{R}_p = 1$. The step sizes for the gradient-based algorithms were adjusted so that the coefficients mean Euclidean distance between iterations was of 10^{-3}, resulting in $\mu_{CM} = 0.001$, $\mu_{RD} = 130$, $\mu_{QD} = 0.03$, $\mu_{QD-C} = 0.03$, $\mu_{QD-SP} = 2.5$, $\mu_{QD-R} = 0.15$, $\mu_{QD-D} = 0.32$ and $\mu_{QD-DR} = 4.2$.

In Fig. 7, we illustrate the mean performance of the gradient-based algorithms adaptation process in terms of HISI and QISI for 30 independent simulations. Comparing to Fig. 6, we note that the QD-SP algorithm is not the fastest anymore, being replaced by the QD-DR algorithm; however, it attains the lowest HISI/QISI level. The CM becomes even slower, but still converges to a low level of HISI/QISI. The QD-R algorithm now converges to a better optima, but with HISI/QISI level higher than the QD-SP; the RD algorithm, on the other hand, is affected by the multilevel modulation and loses performance (which possibly also affects the QD and QD-C algorithm, due to their connection, as Eq. (24) shows); the QD-D algorithm maintains the HISI level achieved in the previous case. Finally, the QD-DR algorithm presents an improvement in terms of the QISI measure, being close to the QD-R performance.

Interestingly, the analysis of the signal scatter plots can also be clarifying in this case, since the ISI measures do not consider phase rotation. In that sense, we displayed
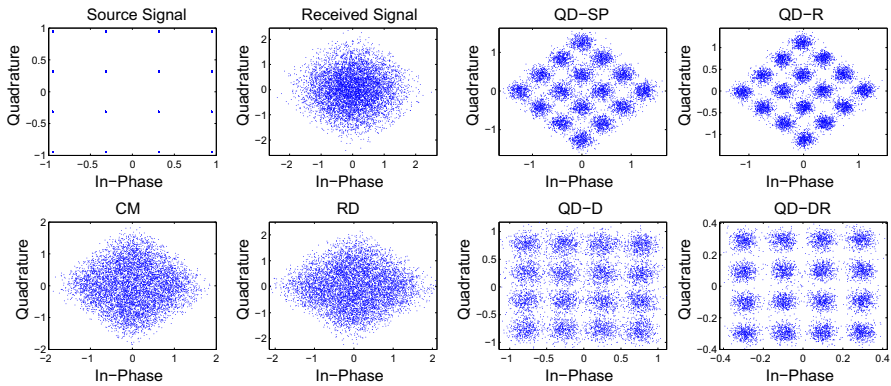
Fig. 8 Scenario 2—Scatter plot of the signals involved in the equalization process

in Fig. 8 these plots of the transmitted, $s(n)$, and the received signals, $x(n)$, as well as the filter output, $y(n)$, trained by the CM, RD, QD-SP, QD-R, QD-D and QD-DR algorithms (considering a single simulation). It is possible to observe that the resulting CM- and RD-trained filters generated signals with rotated constellation and non-easily identifiable symbols; with respect to the QD-SP and QD-R trained filters, the symbols can be identified, but the phase rotation persists; for the QD-D and QD-DR, the symbols are identifiable and there is no phase rotation, which is a very interesting property inherent to these criteria, since they are not based on the modulus of the signals.

In Table 2, we also show the performance obtained by the DE optimization method. It is possible to see that the mean HISI/QISI performance in 30 simulations are equal or higher than that obtained from the gradient-based method: Although the DE metaheuristic presents a wider search potential, there is no guarantee of optimal convergence. Indeed, since we have a large search space, the DE method has generally converged to suboptimal solutions. Undoubtedly, the DE parameter N_p and the number of iterations could have been incremented in order to expand the search potential, but the computational burden would become extremely costly. Nonetheless, the criteria QD-R and QD-DR performed better with the DE in terms of HISI. From the QISI perspective, the QD-R followed by RD, QD-SP and QD criteria were better, but we emphasize that, as the source is discrete (and uniformly distributed) and the noise is non-Gaussian, the HISI measure is more adequate.

6.5 Third Scenario—Continuous Source

The ITL criteria also admit application in scenarios where the sources are continuous. To illustrate, we consider a Laplacian source transmitted over the recursive noiseless channel:

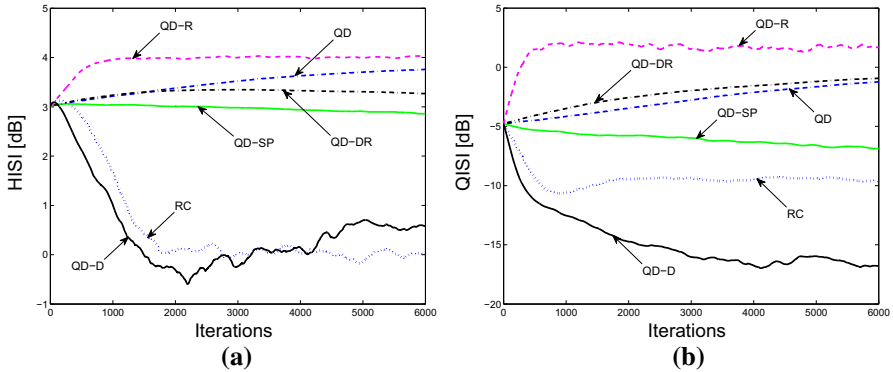$$H_3(z) = \frac{1}{1 - 0.5z^{-1}}. \tag{31}$$

Fig. 9 Scenario 3—Average HISI/QISI performance for the gradient-based algorithms adaptation process. **a** HISI performance. **b** QISI performance

The DE parameters were set as $N_p = 100$ and 300 iterations. The equalizer is an FIR filter with $M = 2$ (3 coefficients), whose weights were adapted by the criteria RC, QD, QD-SP, QD-R, QD-D and QD-DR (being the continuous version for the QD-D and QD-DR)—RD and correntropy were not addressed as they consider discrete and colored sources, respectively; the QD-C does not perform well in this case and was also disregarded. For the RC criterion, the central tap of the filter was set fixed to unity. For the family of criteria based on PDF matching, a fixed set of 150 samples randomly generated from the Laplacian distribution was considered—as it is assumed that the source distribution is known—to be used as the reference signal $s(n)$ in the cost functions. Naturally, this set differs from the actual transmitted source, which is unknown by the receiver. For the QD-SP, we adopted the SPs {0, 1, 2, 3}. The window length was chosen to be $L = 200$ for all analyzed criteria.

To determine the kernel sizes, we made the linear search and found: $\sigma_{RC} = 0.49$, $\sigma_{QD} = 0.18$, $\sigma_{QD-SP} = 0.87$, $\sigma_{QD-R} = 0.05$, $\sigma_{QD-D} = 0.70$ and $\sigma_{QD-DR} = 0.96$. The step sizes were adjusted to $\mu_{RC} = 0.004$, $\mu_{QD} = 0.0008$, $\mu_{QD-C} = 0.001$, $\mu_{QD-SP} = 0.08$, $\mu_{QD-R} = 0.0008$, $\mu_{QD-D} = 0.025$ and $\mu_{QD-DR} = 0.004$, with coefficients mean Euclidean distance between iterations of 0.0001.

The mean HISI/QISI performance after 30 independent simulations for the gradient-based algorithms is illustrated in Fig. 9. It is possible to note that the RC and QD-D algorithms achieved the lowest levels of HISI/QISI. Note that both algorithms present a common term, as shown in Eq. (25). The difference, given by the term correspondent to QD-DR, on the other hand, causes the QD-D algorithm to behave differently from the RC. From the HISI point of view, the QD-D algorithm performs similarly to the RC algorithm until 4000 iterations, when it converges to a minima with HISI level of 0.4 dB, while the other remains at about -0.03 dB. Most probably, the presence of a local optimum in the QD-DR algorithm—which causes him to lose performance—also influences the QD-D. The QISI measure reveals an inverted perspective for the RC and QD-D performance, but again, the use of the HISI is more suitable. All the other criteria converge to local optima.

Table 3 HISI of the best individual of the DE metaheuristic

Criterion	Scenario 3		Criterion	Scenario 4	
	HISI [dB]	QISI [dB]		HISI [dB]	QISI [dB]
RC	0.52	−09.16	Corr.	+3.40	−01.28
QD	2.23	−05.68	QD	−2.44	−12.92
QD-SP	2.15	−07.49	QD-SP	−3.81	−29.95
QD-R	3.94	+01.47	QD-R	+2.61	−04.96
QD-D	1.08	−12.30	QD-D	−5.86	−36.72
QD-DR	3.95	−00.44	QD-DR	+2.48	−05.11

By applying the DE optimization method, we measured the average HISI/QISI performance presented in Table 3. The result is very similar to the gradient-based optimization method in this case, except for the QD and QD-SP criteria, which have now improved their performance, but still remain worse than the RC and QD-D.

Therefore, this illustrative case reveals that the term involving the entropy of $y(n)$— i.e., the term relative to the RC criterion—is essential when dealing with continuous sources, being the RC and QD-D criteria the most suitable for this scenario. Although the QD and QD-SP present a term similar to that of the RC, they are based on the modulus of $y(n)$, which reduces the performance. The QD-R and QD-DR criteria do not have neither of these terms and perform poorer.

6.6 Fourth Scenario—Dependent Source

In the last scenario, we analyze the performance of the ITL criteria when the source is characterized by statistical dependence. In this case, it is admitted that an $i.i.d.$ BPSK signal is submitted through a codification process, which can be modeled by a linear FIR filter, named pre-coder [6,26], with transfer function $P(z) = 1 + 0.5z^{-1}$. The resulting signal of this process—named $s(n)$, with symbols $\{-0.5, -1.5, 0.5, 1.5\}$—is then transmitted through the channel, here assumed to be

$$H_4(z) = 0.1856 - 0.9650z^{-1} + 0.1856z^{-2}. \tag{32}$$

There is also the presence of AWGN with SNR level of 30 dB. It is important to remark that the pre-coder is responsible for introducing the temporal dependence in the source, being not required any additional structure at the receiver, since the objective here is to recover at the equalizer output the same temporally dependent source.

The equalizer is a 5-coefficient filter, being the DE parameters $N_p = 300$ and 500 iterations enough for an extensive exploration of the search space. The criteria considered in this simulation case were the correntropy-based, QD, QD-SP, QD-R, QD-D and QD-DR—the RD, RC and QD-C were not considered in the analysis as they are aimed at independent sources or are unable to approach the desired PDF. For the correntropy-based criterion, it was chosen $P = 5$; for the QD-SP, the SPs
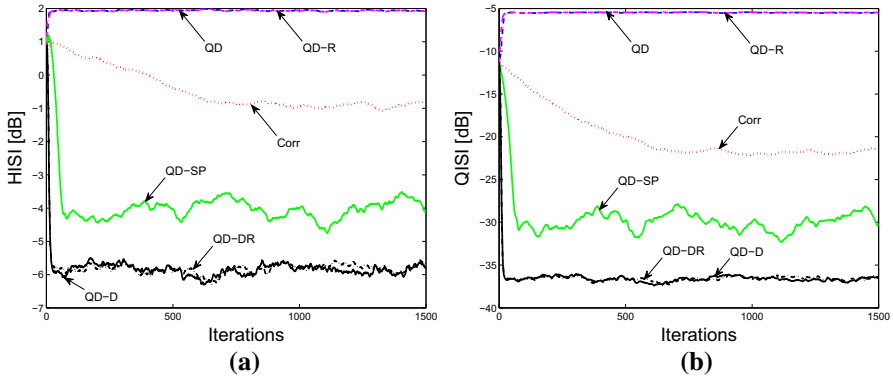
Fig. 10 Scenario 4—Average HISI/QISI performance for the gradient-based algorithms adaptation process. **a** HISI performance. **b** QISI performance

{0.25, 2.25}, and the filter initialization for the gradient-based algorithms was $\mathbf{w} = [0\ 0\ 1\ 0\ 0]$. We adopted $L = 100$ for all analyzed criteria; the kernel size values were $\sigma_{\mathrm{Corr}} = 1.74$, $\sigma_{\mathrm{QD}} = 0.37$, $\sigma_{\mathrm{QD-SP}} = 0.18$, $\sigma_{\mathrm{QD-R}} = 1.78$, $\sigma_{\mathrm{QD-D}} = 0.33$ and $\sigma_{\mathrm{QD-DR}} = 0.26$, and the step sizes $\mu_{\mathrm{Corr}} = 0.5$, $\mu_{\mathrm{QD}} = 0.015$, $\mu_{\mathrm{QD-SP}} = 0.01$, $\mu_{\mathrm{QD-R}} = 3$, $\mu_{\mathrm{QD-D}} = 0.1$ and $\mu_{\mathrm{QD-DR}} = 0.1$ (with the mean Euclidean distance of the coefficients between iterations of 0.001).

Figure 10 shows the average HISI/QISI performance for 30 independent simulations. It is possible to see that the QD-D and QD-DR algorithms have similar performances, with the fastest convergence and lowest HISI/QISI. The QD-SP and the correntropy-based algorithms, although slower, also converge to good solutions, even though the QD-SP algorithm attains lower HISI/QISI level with faster convergence in comparison with correntropy. The QD and QD-R algorithms converge to local optima.

Table 3 exhibits the mean HISI/QISI performance obtained with the DE optimization method. From both HISI and QISI measures, the QD-D, QD-SP and QD criteria showed the best performances (which reveals that the QD algorithm converged to a local solution). Indeed, these criteria are able to extract the temporal dependencies included in the PDFs—as pointed out in Sect. 4, where the second-order approximation revealed correlation-like functions—and are able to provide good solutions even for dependent sources. The correntropy-based and the QD-DR criteria, curiously, had their performance reduced in comparison with the correspondent gradient-based algorithms. Considering that the DE offers greater probability of global convergence, this means that there are local solutions (like the ones in which the algorithms converged) related to these criteria capable of reducing the HISI/QISI to levels lower than that associated with the global solutions, which is not desired. Note that this similar behavior between correntropy and the QD-DR criteria is in accordance with the comparison held in Sect. 5.

In general, statistically dependent sources can provoke changes in the criteria cost surfaces [19], but the QD-D (principally, its multivariated version [6]) and the correntropy-based criteria are known to be more robust in this type of scenarios [5].

Correntropy needs additional care, since the source can lead to similar values of correntropy for different delays and the performance can be reduced. If an evolutionary algorithm is employed, the QD-SP and QD are also possible candidates.

7 Conclusion

In this work, the objectives are twofold: the proposition of a new PDF matching criterion and the presentation of a general overview of the main ITL criteria for blind channel equalization. In the former, the notion of the matching was considered by means of the cross-term product between the two PDFs (or between a PDF and a PMF), reducing the computational cost. The theoretical analysis of the QD-DR showed that such criterion can be linked to the well-known mean-squared error criterion, when compared to the classical approaches, and can be viewed as an extension of the correntropy measure, when analyzed in the ITL framework. In terms of performance, the algorithm presented good results in the scenarios with discrete independent sources and, together with the QD-D criterion, had one of the best performances in the case of dependent sources. The only situation in which it was not able to converge to a good solution was in the presence of continuous sources, case in which the majority of the algorithms also failed.

In the second objective, the ITL criteria overview revealed the connections with the classical framework, where it was possible to associate: terms of the CM criterion to the RD, correntropy and some of the PDF matching-based criteria, and terms of the supervised MSE in the QD-D and, as already mentioned, the QD-DR criteria. We also showed the relationships among some of the ITL criteria, like the QD, RD and QD-R, and the QD-D, RC and QD-DR, whose surface analysis indicated similar global optima solutions. Also, the QD-SP encompasses a correntropy-like comparison.

In the performance analysis considering both the gradient-based algorithm and the DE optimization method, the criteria QD-SP, QD-D, QD-R and also the QD-DR demonstrated better performance for discrete sources in the chosen scenarios (from this set, the QD-DR presents the lower computational cost, followed by the QD-R and QD-SP). For constant modulus modulations, the RD criterion can also be included in this group. We highlight the QD-D and QD-DR criteria performance for complex-valued modulations, since these methods are able to recover phase distortions. When dealing with continuous sources, the simulation case revealed that the QD-D and RC are the most promising criteria, being the QD-D more complex but more robust. For statistically dependent sources, the best performances in simulations were achieved by the QD-D, QD and QD-SP (the QD-SP algorithm presents lower complexity but is slower)—the correntropy-based criterion, although not performing well in this case, is also considered to be a suitable criterion for dealing with dependent sources.

In general, the set of ITL criteria studied represents efficient tools for dealing with the task of blind equalization; however, each criterion has its particularity that can be better explored according to the scenario at hand, as we have shown. Indeed, the ITL framework provides powerful resources capable of dealing with a large content of information; notwithstanding, the manner in which the statistical information is employed in each criterion can privilege certain types of data. In that sense, in view

of frontier research topics that deal with unusual distributions, we expect that the ITL framework will continually grow in order to encompass different signal characteristics.

Appendix: QD-D and QD-DR Gradient-Based Algorithms

The stochastic gradient-based algorithms rely on adaptive filters whose weights are updated in every iteration according to

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla J(\mathbf{w}(n)), \tag{33}$$

where $\nabla J(\mathbf{w}(n))$ is the gradient vector of $J(\mathbf{w}(n))$ and μ is the step size.

Hence, for the QD-D cost, given by Eq. (19), the gradient is

$$
\begin{aligned}
\nabla_{\mathbf{w}} \hat{J}_{\mathrm{QD-D}} = &\frac{1}{2\sigma^2 L^2} \Bigg(\sum_{i=1}^{L} \sum_{j=1}^{L} G_{2\sigma} \left(y(n-i) - y(n-j) \right) \\
&\cdot \left(y(n-j) - y(n-i) \right) \left(\mathbf{x}(n-i) - \mathbf{x}(n-j) \right) \Bigg) \\
&- \frac{2}{\sigma^2 L} \sum_{i \in \mathcal{A}} \left[p_S(s_i) \left(\sum_{j=1}^{L} G_\sigma \left(s_i - y(n-j) \right) \left(s_i - y(n-j) \right) \mathbf{x}(n-j) \right) \right]
\end{aligned}
\tag{34}
$$

which can be directly replaced in Eq. (33) to form the QD-D gradient-based algorithm. The term $1/\sigma^2$ is a common factor that usually is disregarded.

The QD-DR gradient is basically the second term of Eq. (34)—and the factor $2/\sigma^2$ can be disregarded.

For the continuous counterpart, the samples of $s(n)$ are used instead of the source PMF, and the QD-D (and QD-DR) gradient only differs on the second term, as shown in [12].

References

1. A. Benveniste, M. Goursat, G. Ruget, Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications. IEEE Trans. Autom. Control **AC–25**(3), 385–399 (1980)
2. I. Boussaïd, J. Lepagnot, P. Siarry, A survey on optimization metaheuristics. Inf. Sci. **237**, 82–117 (2013)
3. T.M. Cover, J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991)
4. D. Erdogmus, *Information Theoretic Learning: Rényi's Entropy and Its Applications to Adaptive System Training*. Ph.D. thesis (University of Florida, 2002)
5. D.G. Fantinato, R.R.F. Attux, A. Neves, R. Suyama, J.M.T. Romano, Blind Deconvolution of Correlated Sources Based on Second-Order Statistics. XXXI Simpósio Brasileiro de Telecomunicações (2013)
6. D.G. Fantinato, L. Boccato, A. Neves, R.R.F. Attux, Multivariate PDF Matching via Kernel Density Estimation, in *IEEE Symposium Series on Computational Intelligence* (*SSCI*) (2014)
7. G. Foschini, Equalization without altering or detecting data. AT&T Tech. J. **64**(8), 1885–1911 (1985)
8. D. Godard, Self-recovering equalization and carrier tracking in two-dimensional data communication systems. IEEE Trans. Commun. **COM–28**(11), 1867–1875 (1980)

9. A. Gunduz, J. Principe, Correntropy as a novel measure for nonlinearity tests. Signal Process. **89**, 14–23 (2009)
10. S. Haykin, *Adaptive Filter Theory*, 3rd edn. (Prentice Hall, Englewood Cliff, 1996)
11. N. Kim, A study on the complex-channel blind equalization using ITL algorithms. J. Korean Inf. Commun. Soc. **35**(8), 760–767 (2010)
12. N. Kim, L. Yang, A new criterion of information theoretic optimization and application to blind channel equalization. J. Internet Comput. Serv. **10**(1), 11–17 (2009)
13. M. Lázaro, I. Santamaría, D. Erdogmus, K. Hild, C. Pantaleón, J.C. Principe, Stochastic blind equalization based on PDF fitting using parzen estimator. IEEE Trans. Signal Process. **53**(2), 696–704 (2005)
14. M. Lázaro, I. Santamaría, C. Pantaleón, D. Erdogmus, K. Hild, J.C. Principe, Blind equalization by sampled PDF fitting, in *Proceedings of the Fourth International Symposium Independent Component Analysis Blind Equalization, Nara, Japan, 2003*, pp. 1041–1046
15. M. Lázaro, I. Santamaría, C. Pantaleón, D. Erdogmus, J.C. Principe, Matched PDF-based blind equalization, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. IV-297–300 (2003)
16. R. Li, W. Liu, J.C. Principe, A unifying criterion for instantaneous blind source separation based on correntropy. Signal Process. **87**, 1872–1881 (2007)
17. W. Liu, P.P. Pokharel, J.C. Principe, Correntropy: a localized similarity measure, in *International Joint Conference on Neural Networks, Vancouver, Canada*, pp. 4919–4924 (2006)
18. J. Montalvão, C. Cavalcante, B. Dorizzi, J. Mota, A Simple PDF Fitting Approach for Blind Equalization. XVIII Simpósio Brasileiro de Telecomunicações (2000)
19. A. Neves, C. Wada, R. Suyama, R. Attux, J.M. Romano, An analysis of unsupervised signal processing methods in the context of correlated sources. Lect. Notes Comput. Sci. **5441**, 82–89 (2009)
20. K. Nose-Filho, D.G. Fantinato, R. Attux, A. Neves, J. Romano, A Novel Entropy-based Equalization Performance Measure and Relations to Lp-Norm Deconvolution. XXXI Simpósio Brasileiro de Telecomunicações (2013)
21. E. Parzen, On the estimation of a probability density function and the mode. Ann. Math. Stat. **33**, 1065–1067 (1962)
22. J.C. Principe, *Information Theoretic Learning Rényi's Entropy and Kernel Perspectives* (Springer, Berlin, 2010)
23. J.M.T. Romano, R.R.F. Attux, C.C. Cavalcante, R. Suyama, *Unsupervised Signal Processing: Channel Equalization and Source Separation* (CRC Press, Boca Raton, 2010)
24. I. Santamaría, C. Pantaleón, L. Vielva, J.C. Principe, A fast algorithm for adaptive blind equalization using order-α Renyi's entropy. IEEE Trans. Acoust. Speech Signal Process. **2**(1), 2657–2660 (2002)
25. I. Santamaría, C. Pantaleón, L. Vielva, J.C. Principe, Adaptive blind equalization through quadratic PDF matching. Proc. Eur. Signal Process. Conf. **II**, 289–292 (2002)
26. I. Santamaría, P. Pokharel, J.C. Principe, Generalized correlation function: definition, properties and application to blind equalization. IEEE Trans. Signal Process. **54**(6), 2187–2197 (2006)
27. O. Shalvi, E. Weinstein, New criteria for blind deconvolution of nonminimum phase systems (channels). IEEE Trans. Inf. Theory **36**(2), 312–321 (1990)
28. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**(379–423), 623–656 (1948)
29. R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. **11**, 341–359 (1997)