Analysis of the Correntropy-Based Criterion for Blind Equalization with Precoded Sources

Vinícius A. de Oliveira, Denis G. Fantinato, Rafael Ferrari, Romis Attux, Levy Boccato

***Abstract*— Temporally structured sources may arise in blind equalization as the result of coding schemes. In this context, a promising approach involves the use of correntropy, which is capable of exploring both the statistical information and the temporal structure of the signals. In this work, we perform a detailed analysis of the correntropy-based criterion for equalization, giving a special attention to the effect of the adjustable parameters as well as to the comparison with the analytical correntropy measure, whose formal derivation is another contribution of this work. The experimental results indicate the attainable performance and the influence of the main parameters.**

***Keywords*— Adaptive filtering, Blind channel equalization, Information-theoretic learning, Correntropy.**

I. INTRODUCTION

The field of unsupervised signal processing deals with the challenge of extracting signal(s) of interest from a collection of measurements that usually contain distortions, such as those resulting from interferences between the signals and/or from the presence of noise, having at disposal a minimum amount of information regarding the original signal(s) and the specific characteristics of the process that generates the distortions.

An example of such challenge arises in the context of the blind equalization problem, in which the objective is to design a filter at the receiver, named equalizer, that attempts to cancel the noxious effects of the channel used for the transmission, especially the so-called *intersymbol interference* (ISI), thus enabling the recovery of the transmitted information.

Two fundamental results establish sufficient conditions for perfect equalization: the Benveniste-Goursat-Ruget (BGR) and the Shalvi-Weinstein (SW) theorems [1]. The BGR theorem proves that if the probability density function (PDF) of the signal at the equalizer output is equal to the PDF of the source signal, then the channel has been equalized. On the other hand, the SW theorem demonstrates that it is not necessary to resort to the PDFs of the involved signals: only a few higher-order statistical moments need to match in order to achieve the equalization. Based on both theorems, several criteria and algorithms have been proposed for blind equalization [1], [2].

However, when the source signal presents a temporal structure of dependency, i.e., it does not correspond to a process with independent and identically distributed (i.i.d.) samples, the aforementioned conditions are not directly applicable, and many of the existing blind techniques experience a significant degradation in performance [3]. This type of scenario may emerge due to the natural characteristics of the involved signal, e.g., in the context of speech, audio and video signals, or even due to the use of coding schemes at the transmitter, which purposely introduce redundancies in the signal prior to its transmission.

Interestingly, the area of information-theoretic learning (ITL) provides a measure, known as correntropy, which serves as the basis for a blind equalization criterion that represents the state-of-the-art for the treatment and analysis of temporally dependent signals [4], [5]. The correntropy can be seen as a generalized correlation function, since it computes a similarity measure in a kernel feature space between a signal and its delayed version as a function of the delay, but with the advantage that several higher-order moments of the considered random variable are implicitly exploited due to the use of a kernel function [5].

The criterion proposed by [5] corresponds to minimizing the squared error between the values of the correntropy of the source signal, assumed to be known, and the correntropy observed at the equalizer output considering a limited number of delays. Hence, the temporal structure of the source is effectively incorporated in the design process of the equalizer, and the statistical information of the involved signals is implicitly forced to be equal due to the properties of the correntropy.

Despite of these attractive features, the effective use of correntropy in blind equalization requires the definition of some parameters, such as the kernel size and the number of lags (time delays) that must match, whose influence on the attainable performance of the equalizer is not completely clear [6]. Additionally, the correntropy at the equalizer output needs to be estimated and, thus, the number of samples used in such estimation process represents another important factor.

In view of these facts, in this work we perform an analysis of the correntropy-based criterion for blind equalization considering the presence of temporally dependent sources, which are generated through the use of a linear precoding stage. In this context, we formally derive the analytical expression of the correntropy associated with the signal generated at the output of a finite impulse response (FIR) equalizer, thus enabling the investigation of the corresponding criterion in its ideal form. A second contribution of this work refers to an analysis of the role played by each parameter of the correntropy-based criterion, considering both its theoretical and estimated versions, and, finally, to the subsequent comparative study between the performances associated with each equalizer.

Vinícius A. de Oliveira, Denis G. Fantinato, Rafael Ferrari, Romis Attux, Levy Boccato, School of Electrical and Computer Engineering, University of Campinas (UNICAMP), Campinas, SP, Brazil, E-mails: {valvesd,denisgf,rferrari,attux,lboccato}@dca.fee.unicamp. This work was partially supported by FAPESP (2013/14185-2), CAPES and CNPq.

II. Problem Statement and Correntropy

The scenario of the channel equalization problem that will be treated in this work is depicted in Fig. 1.

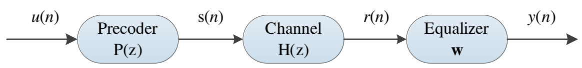


Fig. 1. Block diagram of the channel equalization problem.

Differently from the classical approach [1], we consider that there exists a precoder $P(z)$, which receives an i.i.d. signal $u(n)$, which belongs to the BPSK modulation, and produces a source signal $s(n)$ with a temporal structure, i.e., a sequence of statistically dependent samples [5]. The effect of the precoder can be related to symbolic coding shemes, such as channel coding with linear blocks or convolutional codes [7].

The channel is represented by its coefficient vector $\mathbf{h} = [h_0 \; \dots \; h_D]^T$, where $[\cdot]^T$ stands for matrix transposition, and is responsible for introducing the ISI. The equalizer is also a FIR filter with coefficients given by $\mathbf{w} = [w_0 \; w_1 \; \dots \; w_M]^T$, and generates an output signal $y(n)$ according to the following expression:

$$y(n) = \mathbf{w}^T \mathbf{r}(n), \tag{1}$$

where $\mathbf{r}(n) = [r(n) \; \dots \; r(n{-}M)]^T$ is the received signal vector, with $r(n) = h(n) * s(n)$ and $*$ represents the discrete-time convolution. The objective is to adapt $\mathbf{w}$ in order that $y(n)$ becomes a correct estimate of a transmitted symbol $s(n-d)$, being allowed an equalization delay (d) [1]. The criterion used in the adaptation process shall be based on the concept of correntropy, which is defined in the following section.

A. Definition of Correntropy

An emblematic entity in the context of unsupervised ITL is the measure called *correntropy*, which can be seen as a generalized correlation function [4], [5]. Besides taking into account the statistical distribution of signals, the correntropy is able to encompass their temporal structure, which is particularly useful when dealing with signals with statistical dependence. It may be defined as

$$\begin{aligned} v_Y(m) &= \int_D \kappa_\sigma(\mathbf{v})\, f_{Y_n,Y_{n-m}}(\mathbf{v})\, d\mathbf{v} \\ &= E_{Y_n,Y_{n-m}}\left[\kappa_\sigma(\mathbf{v})\right], \end{aligned} \tag{2}$$

where $\kappa(\cdot)$ denotes a positive definite kernel function, $f_{Y_n,Y_{n-m}}(\cdot)$ represents the joint probability density function (PDF) of Y_n and Y_{n-m}, $E[\cdot]$ is the expectation operator, σ is the kernel size and m is the time delay between samples. As usual in the ITL field, a Gaussian kernel function is considered, and a sample mean approximates the statistical expectation in Eq. (2), resulting in

$$\hat{v}_Y(m) = \frac{1}{L-m+1} \sum_{n=m}^{L} G_{\sigma^2}(y(n) - y(n{-}m)), \tag{3}$$

being L the window length or the number of samples and

$$G_{\sigma^2}(y(i) - y(j)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-|y(i) - y(j)|^2}{2\sigma^2}\right), \tag{4}$$

the Gaussian kernel with kernel size σ.

B. The Correntropy-Based Criterion for Equalization

An interesting criterion for blind equalization is the matching between the correntropies of the source and of the equalized signal, through the minimization of the following cost function [5]:

$$J_{corr}(\mathbf{w}) = \sum_{m=1}^{B} \left(v_S(m) - v_Y(m)\right)^2, \tag{5}$$

where v_S is the correntropy of the source, v_Y is the correntropy of the equalizer output and B is the number of lags being considered. In this case, the temporal structure of the source is extracted by the correntropy function $v_S(m)$, which contains the *higher order statistics* (HOS) information that can be compared with the correntropy associated with the equalized signal, $v_Y(m)$.

Since the precoder is assumed to be known, it is usually considered that $v_S(m)$ can be analytically obtained, while $v_Y(m)$ is replaced by its estimate $\hat{v}_Y(m)$, given by Eq. (3). However, we will also address the case in which $v_Y(m)$ is analytically computed, which certainly allows a deeper understanding of the criterion behavior, and its dependence on the adjustable parameters. Additionally, we may also establish a comparison between the analytical and estimated versions of the criterion, as well as between the corresponding optimum equalizers in terms of the capability of reducing the ISI. In the following, we present the formal derivation of the correntropy associated with the equalizer output signal.

C. Analytical Correntropy

When the source is discrete, the correntropy can be analytically computed from a probability mass function (PMF) associated with a filtered signal according to the methodology proposed in the following. Consider that the precoder $P(z)$ is combined with a time differentiating system with transfer function $Q(z) = 1 - 1z^{-m}$, which results in $P'(z) = P(z)Q(z) = P(z)(1 - 1z^{-m})$. The system $P'(z)$, besides applying the source precoding, also considers the difference between m-delayed time samples, which is a crucial comparison term in correntropy when Gaussian kernels are considered, as can be seen in Eq. (3). The filtering process then yields:

$$s'(n) = \mathbf{p'}^T \mathbf{u}(n). \tag{6}$$

If the PMF associated with $s'(n)$ is

$$p_{S'}(v) = \sum_{i \in \mathcal{A}_{S'}} P(v = a_{S'}(i))\delta(v - a_{S'}(i)), \tag{7}$$

where $\mathcal{A}_{S'}$ is the alphabet of all possible occurrences of S', $a_{S'}(i)$ is the i-th symbol $\in \mathcal{A}_{S'}$ and $P(v = a_{S'}(i))$ is the probability of $v = a_{S'}(i)$, then the correntropy can be simply calculated as

$$v_S(m) = \sum_{i \in \mathcal{A}_{S'}} p_{S'}(a_{S'}(i)) G_{\sigma^2}\left(a_{S'}(i)\right). \tag{8}$$

We consider, for example, the case in which $u(n)$ is a BPSK modulated signal. If the vector $\mathbf{p}'$ is of length $M_{p'} + 1$, then there are $2^{M_{p'}+1}$ possible occurrences for $\mathbf{u}(n)$ in Eq. (6)

(all possible permutations of $\{+1, -1\}$ in a vector of length $2^{M_{p'}+1}$), each of them with probability $1/2^{M_{p'}+1}$. In this case, the i-th state of S', $a_{S'}(i)$, is simply the output given by Eq. (6) for the i-th possible occurrence of $\mathbf{u}(n)$, with probability $P(v = a_{S'}(i)) = 1/2^{M_{p'}+1}$.

Interestingly, assuming a noiseless scenario, the analytical computation of the correntropy associated with the equalizer output can be easily obtained: by defining the combined system $G'(z) = P(z)H(z)W(z)(1-1z^{-m})$, the produced signal is given by $y'(n) = \mathbf{g'}^T\mathbf{u}(n)$, whose PMF will be $p_{Y'}(v)$, which can be written in a very similar manner to Eq. (7), with $\mathcal{A}_{\mathcal{Y}'}$ being the alphabet of all possible occurrences of Y'. So, having $p_{Y'}(v)$ in hand, the correntropy $v_Y(m)$ can be exactly determined with the aid of Eq. (8).

In view of this, there emerge two possibilities for the present analysis: (*i*) from a theoretical perspective, the case in which both the correntropy of the source and of the equalizer output are analytically computed, as defined on Eq. (5); and (*ii*) from a practical perspective, the case in which the correntropy of the equalizer output is estimated from samples (Eq. (3)), resulting on the estimated cost:

$$\hat{J}_{corr}(\mathbf{w}) = \sum_{m=1}^{B} \left(v_S(m) - \hat{v}_Y(m)\right)^2. \quad (9)$$

In the former case, the parameters are the maximum number of delays B and the kernel size σ, whereas, in the second case, the number of samples L is also a parameter to be adjusted. Naturally, a suitable choice of the parameters will depend on the distribution of the signals, the scenario at hand and other relationships between the parameters, but, as we intend to show, the parameters values that yield the best approximation of the analytical correntropy through the use of the estimator may not be the ones associated with the best performance of the equalizer. In the following, we perform a set of computational simulations that will contribute to a better understanding of the correntropy-based criterion and the adjustment of its parameters.

III. Simulation Results

In order to investigate the relationships between the analytical and the estimated correntropy-based costs, $J_{corr}(\mathbf{w})$ and $\hat{J}_{corr}(\mathbf{w})$, respectively, we first compare the cost surfaces in a scenario with a two-tap filter as the equalizer. Next, by varying the kernel size σ and the number of samples L, the performances of the global solutions are compared in terms of the ISI measure, defined as

$$ISI_{dB} = 10\log_{10}\frac{\left(\sum_{i=0}^{M_c}|c_i|^2\right) - \max_j |c_j|^2}{\max_j |c_j|^2}, \quad (10)$$

where $\mathbf{c} = [c_0\ c_1\ \ldots\ c_{L_c}]^T$ is the combined channel-equalizer impulse response (with transfer function $C(z) = H(z)W(z)$). Hence, the objective will be to minimize the ISI. Finally, the performance in more complex scenarios will be considered, where the influence of the chosen maximum number of delays B will also be investigated.

In all experiments concerning $\hat{J}_{corr}(\mathbf{w})$, the performance analysis – in terms of ISI – considers the average of the results obtained in N_E independent simulations. Additionally, the precoder used in all scenarios has the transfer function $P(z) = 1 - 0.5z^{-1} + 0.3z^{-1}$.

The search for the global optimum of the cost functions shall be performed by the metaheuristic called Differential Evolution (DE), which is an efficient technique to explore the search space and to avoid convergence to local optima [8]. Its main feature is the fact that the candidate solutions are adapted by mechanisms that exploit the information about the search space that is available in the current population, instead of using conventional operators based on random perturbations (for more details, please refer to [8]). The DE parameters are the population size N_P, the step size F, the crossover rate CR and the maximum number of iterations I_T, which will be adjusted according to each scenario at hand.

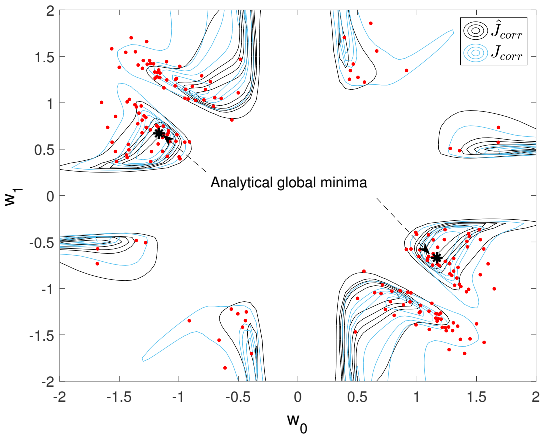
A. Cost Surfaces Analysis

In the first scenario, the channel is a minimum-phase system with transfer function given by $H(z) = 1 + 0.5z^{-1}$. An equalizer with two coefficients $\mathbf{w} = [w_0\ w_1]^T$ is adopted, whose weights will be varied from -2 to 2 to obtain the contours of the costs $J_{corr}(\mathbf{w})$ and $\hat{J}_{corr}(\mathbf{w})$. The DE parameters were chosen to be $N_P = 100$, $F = 0.5$, $CR = 0.9$ and $I_T = 100$ iterations.

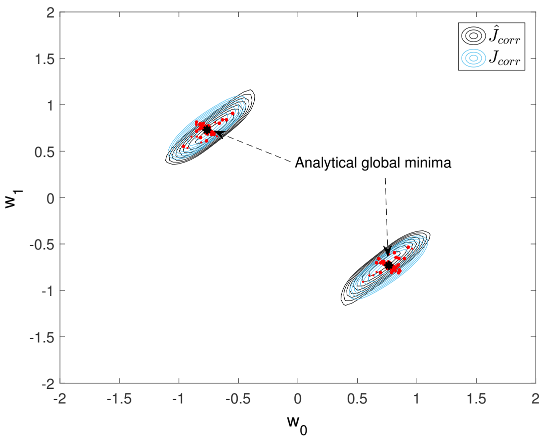
Firstly, we analyze the effect of the kernel size σ on the surfaces. We fixed $B = 2$ for both analytical and estimated costs, using $L = 500$ samples for the latter. Then, by varying the equalizer weights, we obtained the contours of $J_{corr}(\mathbf{w})$ and $\hat{J}_{corr}(\mathbf{w})$ for the kernel sizes of $\sigma = 0.2$ and $\sigma = 0.9$, as shown in Figs. 2(a) and 2(b), respectively. Additionally, we also exhibit the solutions found by the DE, which are represented by asterisks (*) for the analytical cost and by solid dots (.) for the estimated cost function, considering $N_E = 100$ independent experiments. It is possible to observe that, in both cases, the contours of the analytical and the estimated costs are, to a certain extent, similar, but the modification in the kernel size causes a significant difference on the surface shape: for $\sigma = 0.2$, there are several local minima, while for $\sigma = 0.9$, there exist only global minima. Additionally, in Fig. (2(a)), we can notice that $\hat{J}_{corr}$ offers solutions that are actually close local minima, instead of the global minima, of the analytical cost function, which suggests that low kernel size values may cause larger estimation errors. Indeed, when the kernel size is increased to 0.9, Fig. 2(b), the solutions become more concentrated at the vicinity of the analytical global minima.

However, increasing the kernel size leads to a degradation in the equalizer performance, as we can see in Tab. I, which presents the ISI values attained by the solutions associated with the analytical and estimated cost functions. So, even though $\hat{J}_{corr}$ provides solutions more similar to the analytical one for $\sigma = 0.9$, the equalizer is not the best we could obtain with this type of criterion, as a better performance is achieved for a smaller kernel with J_{corr}.

Proceeding with the same scenario, we now analyze the influence of the kernel size and of the number of samples in the performance associated with each version of the correntropy-based criterion. Firstly, we vary σ from 0.05 to 1 in steps of



(a) $\sigma = 0.2$



(b) $\sigma = 0.9$

Fig. 2. Contours of the analytical and estimated cost function.

TABLE I

ISI (IN DB) ASSOCIATED WITH J_{corr} AND $\hat{J}_{corr}$ OPTIMUM SOLUTIONS.

Cost Function	$\sigma = 0.2$	$\sigma = 0.9$
Estimated	-3.2789	-3.4204
Analytical	-10.5570	-3.6109

0.05, considering for the estimated cost a total of $L = 5000$ samples. For each value of σ, we show in Fig. 3 the ISI value associated with the best solution offered by the analytical cost function, which was found by the DE algorithm, as well as the average ISI obtained by the solutions of the estimated cost function. The shaded area is the ISI range taking the standard deviation obtained by all the solutions found for $\hat{J}_{corr}$. For comparison purposes, we also exhibit the ISI performance of the Wiener solution considering the best equalization delay [2].

It is possible to notice in Fig. 3 that the ISI values associated with $J_{corr}(\mathbf{w})$ may significantly differ from those of $\hat{J}_{corr}(\mathbf{w})$ for $\sigma \leq 0.6$. For $\sigma > 0.6$, the performances are similar, suggesting that the solutions of the analytical and estimated cases are close, as we verified in Fig. 2(b), but their ISI levels are higher. It is also worth mentioning that the best ISI performance is attained for $\sigma = 0.2$ in both cases. However, as depicted in Fig. 2(a), the solutions associated with the estimated correntropy-based cost may vary significantly, as can also be seen in Fig. 3.

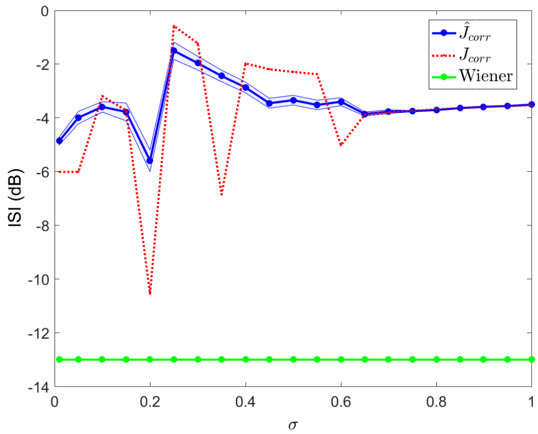


Fig. 3. ISI as a function of the kernel size σ.

In particular, the fact that the ISI variance associated with $\hat{J}_{corr}$ is higher for low σ values suggests that the estimation may not be accurate enough with the chosen number of samples L. Hence, we now analyze the effect of L in the estimation by varying it from 1000 to 40000 in steps of 1000. During the analysis, we kept $B = 2$ and $\sigma = 0.2$ fixed for J_{corr} and $\hat{J}_{corr}$. The resulting ISI performances are illustrated in Fig. 4, where we also displayed the ISI range of the estimated case and the ISI obtained by the Wiener solution. It is possible to notice that as L is increased, the average ISI

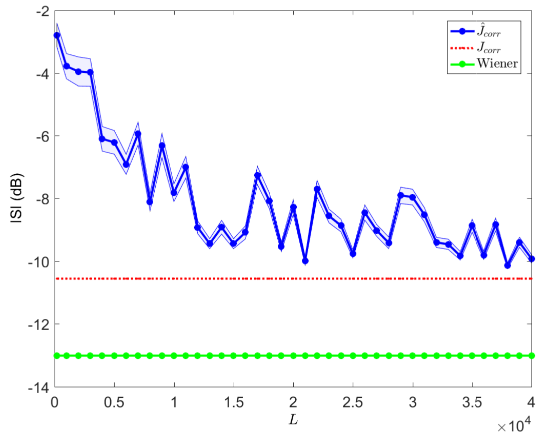


Fig. 4. ISI values as a function of the number of samples L.

obtained with $\hat{J}_{corr}$ gets closer to the ISI value associated with the analytical cost function. Moreover, the ISI range tends to decrease for larger L, which means that the solutions are varying less. Notwithstanding, the enormous amount of data necessary to obtain more accurate correntropy estimates can be of major concern, depending on the specific application. On the other hand, the use of larger kernel sizes would require a smaller number of samples for an adequate estimation, but it would also lead to poorer solutions, as observed in Fig. 3.

Finally, we briefly verify the impact of the number of lags B for J_{corr} and $\hat{J}_{corr}$ when $\sigma = 0.2$ and $L = 20000$ samples. Figure 5 displays the source correntropy function (the correntropy target values) for time delays m from 1 to 10, along with the equalizer output correntropy function for both the analytical and estimated versions (mean of 100 Monte Carlo simulations), considering $B = 2$ (upper) and $B = 10$ (bottom). The ISI performances for both delays are displayed in Tab. II. As can be observed in Fig. 5 for $B = 2$ (upper),

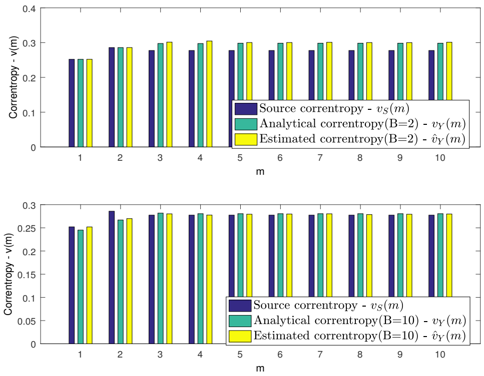


Fig. 5. Correntropy profile as a function of the delay m.

TABLE II

ISI (IN DB) ASSOCIATED WITH $\hat{J}_{corr}$ AND J_{corr} OPTIMUM SOLUTIONS.

Delays	Estimated	Analytical
$B = 2$	-7.2661	-10.5571
$B = 10$	-5.5405	-13.0102

both the analytical and the estimated correntropy functions at the equalizer output match that of the source for $m = 1$ and $m = 2$, which are precisely the delays the criterion effectively attempts to match. However, for the subsequent correntropy lags m, this does not hold and the solutions of estimated cost performed poorer due to the estimation errors, as exhibited in Tab. II. When B is increased to 10 (Fig. 5 bottom), the correntropy profiles are not perfectly matched with that of the source for both analytical and estimated cases, but, even so, the analytical cost solution points toward a lower ISI level, while the estimated correntropy loses performance, most probably due to the cumulative estimation errors along the several delays. In this sense, a small number of delays B seems to be more suitable to minimize the estimation errors.

B. More Complex Channel/Equalizer

In the following, we consider a higher dimensional scenario with a 5-tap equalizer as the filtering structure in order to support the generality of our observations. Now, the source signal is distorted by a three-tap channel with transfer function $H(z) = 1 + 0.8z^{-1} - 0.25z^{-2}$. The kernel size for both J_{corr} and $\hat{J}_{corr}$ was $\sigma = 0.6$. In the case of $\hat{J}_{corr}$, $N = 10^5$ samples were used to obtain the output correntropy. In all experiments, the DE algorithm was used with the following parameters: $N_P = 500$, $I_T = 500$, $F = 0.5$ and $CR = 0.9$. For this analysis, we considered $B = 4$ lags in the computation of both criteria. The performances in terms of ISI are shown in Tab. III for an average of $N_E = 20$ simulations.

TABLE III

ISI ASSOCIATED WITH J_{corr} AND $\hat{J}_{corr}$ OPTIMUM SOLUTIONS.

	Estimated	Analytical	Wiener
ISI (dB)	-8.8961	-14.6835	-30.6349

Two important remarks can be drawn from Tab. III: (*i*) the average performance associated with $\hat{J}_{corr}$ is inferior to that of J_{corr}, which, in view of the previous results, suggests that other minima have been found instead of a solution close to the analytical; (*ii*) the blind correntropy solution for the equalizer does not reach the same performance level of the Wiener (supervised) solution.

IV. CONCLUSIONS

In this work, we performed an analysis of the correntropy-based criterion for blind equalization with temporally dependent sources. For this purpose, we derived the analytical expression of the correntropy at the equalizer output, which can be considered one of the contributions of this work, and we investigated the influence of the main parameters on the correntropy-based criterion, considering both its analytical and estimated versions. From the simulation results, we verified that correntropy is more sensitive to small kernel sizes, which is also the situation that may lead to better results and to higher discrepancies between the analytical and estimated correntropy-based criterion. In this case, the number of samples can be (considerably) increased to improve the estimation quality. Using more delays, additional information is provided, as observed in the performance of the analytical correntropy, but, for the estimated correntropy, a cumulative estimation error is perceived in this case, causing a performance loss.

For future works, we intend to analyze the precoder adjustment aiming at reducing the ISI, and also to study the behavior of correntropy-based criterion in scenarios with the presence of noise and with other types of source signals.

REFERENCES

[1] J. M. T. Romano, R. Attux, C. Cavalcante and R. Suyama, *Unsupervised Signal Processing: Channel Equalization and Source Separation*. CRC Press, 2010.
[2] S. Haykin, *Adaptive Filter Theory*. 5th Edition. Pearson, 2013.
[3] D. G. Fantinato, A. Neves, and R. Attux, “Analysis of a Novel Density Matching Criterion Within the ITL Framework for Blind Channel Equalization,” *Circ., Syst., and Sig. Proc.*, v. 34(543), pp. 1–29, 2017.
[4] J. C. Principe, *Information Theoretic Learning Renyi’s Entropy and Kernel Perspectives*. Springer, 2010.
[5] I. Santamaría, P. Pokharel and J. Principe, “Generalized Correlation Function: Definition, Properties and Application to Blind Equalization,” *IEEE Trans. on Signal Processing*, v. 54, pp. 2187–2197, 2006.
[6] D. G. Silva, D. G. Fantinato, J. Canuto, L. Duarte, A. Neves, R. Suyama, J. Montalvão Filho, R. Attux, “An Introduction to Information Theoretic Learning, Part II: Applications,” *Journal of Communication and Information Systems*, v. 31(1), 2016.
[7] J. R. Barry, E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Springer, 3rd. Edition, 2003.
[8] K. Price, R. Storn and J. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Springer, 2005.