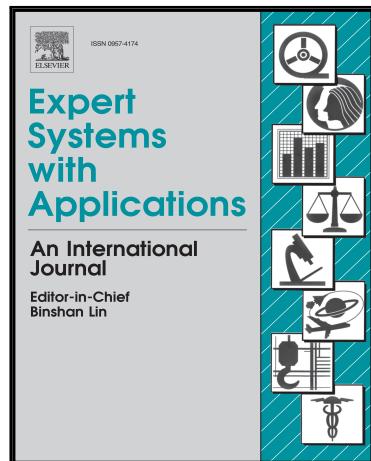


Accepted Manuscript

Complex Correntropy Function: properties, and application to a channel equalization problem

João P.F. Guimarães, Aluisio I.R. Fontes, Joilson B.A. Rego,
Allan de M. Martins, J.C. Principe

PII: S0957-4174(18)30250-1
DOI: [10.1016/j.eswa.2018.04.020](https://doi.org/10.1016/j.eswa.2018.04.020)
Reference: ESWA 11934



To appear in: *Expert Systems With Applications*

Received date: 12 September 2017
Revised date: 19 January 2018
Accepted date: 13 April 2018

Please cite this article as: João P.F. Guimarães, Aluisio I.R. Fontes, Joilson B.A. Rego, Allan de M. Martins, J.C. Principe, Complex Correntropy Function: properties, and application to a channel equalization problem, *Expert Systems With Applications* (2018), doi: [10.1016/j.eswa.2018.04.020](https://doi.org/10.1016/j.eswa.2018.04.020)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- We define a new generalized similarity measure between complex random variables
- A complete deduction is presented and important properties are studied
- Complex correntropy has overcome classical methods in a channel equalization problem
- Now it is possible to use correntropy in problems involving complex-valued data

Complex Correntropy Function: properties, and application to a channel equalization problem

João P. F. Guimarães, Aluisio I. R. Fontes, Joilson B. A. Rego, Allan de M. Martins, J.C. Principe

Federal Institute of Rio Grande do Norte, 59015-300, Natal, Brazil
e-mails: {joao.guimaraes, aluisio.rego}@ifrn.edu.br

Department of Computer Engineering and Automation
Federal University of Rio Grande do Norte, 59078-900, Natal, Brazil
e-mails: {allan, joilson}@dca.ufrn.br

Department of Electrical and Computer Engineering
University of Florida, Gainesville, FL 32611 USA
e-mails: {principe}@cnel.ufl.edu

Abstract

The use of correntropy as a similarity measure has been increasing in different scenarios due to the well-known ability to extract high-order statistic information from data. Recently, a new similarity measure between complex random variables was defined and called complex correntropy. Based on a Gaussian kernel, it extends the benefits of correntropy to complex-valued data. However, its properties have not yet been formalized. This paper studies the properties of this new similarity measure and extends this definition to positive-definite kernels. Complex correntropy is applied to a channel equalization problem as good results are achieved when compared with other algorithms such as the complex least mean square (CLMS), complex recursive least squares (CRLS), and least absolute deviation (LAD).

Keywords:

channel equalization, complex-valued data, correntropy, fixed-point algorithm, maximum complex correntropy criterion, properties;

1. Introduction

Recent studies have demonstrated that correntropy is an efficient tool for analyzing higher-order statistical moments in machine learning and signal processing applications (Liu et al., 2007; Fontes et al., 2014, 2015, 2017). In particular, the maximum correntropy criterion (MCC) has been successfully used in filtering and robust regression (Liu et al., 2017a,b; Wang et al., 2017; Chen et al., 2014; Ma et al., 2017). Correntropy extends the autocorrelation function to the nonlinear systems analysis and nonGaussian sources from real-valued data (Peng et al., 2017; He et al., 2011).

On the other hand, many applications associated to signal processing involve signal sources that employ complex-valued data. However, correntropy has only been defined for real numbers (Santamaria et al., 2006). So, due to the limitation to deal with real-valued data, it is difficult to use correntropy in a straightforward way as applied to problems involving complex-valued data. Inspired by the probabilistic interpretation demonstrated in (Liu et al., 2006), our research group has extended the concept of classic correntropy to include processing from complex-valued data (Guimaraes et al., 2017), as the method has been defined as complex correntropy. This similarity measure is based on the probability density function (PDF) applied to multidimensional spaces using the Parzen estimator with a Gaussian Kernel. Additionally, a new cost function called Maximum Complex Correntropy Criterion (MCCC) is defined, whose performance has proven to be superior to that regarding classical algorithms for processing complex-valued data (Guimaraes et al., 2017). However, the properties of complex correntropy have yet not been properly formalized.

This paper extends the complex correntropy definition to positive-definite kernels (J. Mercer, 1909) and gives a probabilistic interpretation for the case of positive valued kernels. Besides, some important properties are presented and studied, which make complex correntropy structurally similar to the correntropy function. Complex correntropy is also used as a cost function in a channel equalization problem, while the results demonstrate the advantages of this new similarity function in nonGaussian environments when compared to LAD, CLMS, and CRLS algorithms.

The organization of the paper is as follows. Section 2 defines the complex correntropy, as some important properties are presented. In Section 3, the maximum complex correntropy criterion (MCCC) is presented, where Wirtinger derivatives are briefly discussed and used to obtain a fixed-point

recursive algorithm. Section 4 describes a channel equalization problem used to illustrate and verify the theoretical assumptions associated to the properties of MCCC, which provide overall improved performance if compared with the classical solutions. Finally, Section 5 summarizes the main conclusions and potential future work.

2. Definition and Properties of Complex Correntropy

This section presents the complex correntropy definition and properties. Let us always assume that C_1 and C_2 are always complex random variables, where

$$\begin{aligned} C_1 &= X + j Z \\ C_2 &= Y + j S \end{aligned} \tag{1}$$

with $X; Y; Z; S$ being real-valued random variables; and j refers to the imaginary unit. For convenience, let us also define $\beta \in \mathbb{C}$ as the difference between C_1 and C_2

$$\beta = (C_1 - C_2) = (X - Y) + j(Z - S) \tag{2}$$

Consequently, it is also true to say that

$$\begin{aligned} \beta_{re} &= (X - Y) \\ \beta_{im} &= (Z - S) \end{aligned} \tag{3}$$

where $\beta_{re}, \beta_{im} \in \mathbb{R}$, being the real and the imaginary part of β .

2.1. Definition

Complex Correntropy is a generalized similarity measure between two arbitrary scalar complex random variables C_1 and C_2 defined as

$$V_\sigma^C(C_1, C_2) = E_{C_1 C_2}[K(C_1, C_2)] \tag{4}$$

where K is a positive-definite kernel.

2.2. Properties:

- Important properties of the complex correntropy are listed and proved in this section. Since complex correntropy was developed to keep important characteristics from the classical correntropy, some properties such as
- 60 symmetry, high-order statistical measure, probabilistic interpretation, consistent and asymptotically unbiased estimator, are the same for both cases (i.e. properties 1, 3, 4 and 5). Other properties must be properly adapted and modified e.g. boundedness (property 2). Besides, two new properties exist in the case of complex correntropy: one of them connects two similarity
 - 65 measures numerically (property 6), while the other one is adequate for the polar representation of complex numbers (property 7).

Property 1: For symmetric kernels, complex correntropy is symmetric.

$$V_\sigma^C(C_1, C_2) = E_{C_1 C_2}[K(C_1, C_2)] = E_{C_1 C_2}[K(C_2, C_1)] = V_\sigma^C(C_2, C_1) \quad (5)$$

Proof. This property follows from the concepts of positive definiteness and symmetry. ■

70

Property 2: Complex correntropy is positive and bounded. For the complex Gaussian kernel, its estimated value \hat{V}_σ^C is always real and between zero, the minimum value, and $1/2\pi\sigma^2$, which is achieved when $C_1 = C_2$.

$$0 \leq \hat{V}_\sigma^C(C_1, C_2) \leq \frac{1}{2\pi\sigma^2} \quad (6)$$

Proof. Applying the complex Gaussian kernel G_σ^C , defined in (7), to the

75 definition established by (4) gives (8).

$$G_\sigma^C(C_1 - C_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (7)$$

$$\hat{V}_\sigma^C(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (8)$$

If $C_1 = C_2$, then $\hat{V}_\sigma^C(C_1, C_2) = 1/2\pi\sigma^2$. Otherwise, if C_1 is much different from C_2 , the negative exponential term of equation (8) increases and $\hat{V}_\sigma^C(C_1, C_2)$ tends to zero, thus validating the property. ■

Property 3: For the Gaussian kernel, the complex correntropy is a weighted sum of all the even moments of the random variable $\beta = C_1 - C_2$. Furthermore, increasing the kernel size makes correntropy tends to the correlation of C_1 and C_2 .

Proof. Considering equations (1)–(3), let us analyze $V_\sigma^C(C_1, C_2) = V_\sigma^C(\beta_{re}, \beta_{im}) = E[G_\sigma^C(\beta)]$ according to its respective Taylor series expansion over the variables β_{re} and β_{im} , which gives:

$$\begin{aligned} V_\sigma^C(\beta_{re}, \beta_{im}) = & \frac{1}{2\pi\sigma} E[1 - \frac{\beta_{re}^2}{2\sigma^2} + \frac{\beta_{re}^4}{8\sigma^4} - \frac{\beta_{re}^6}{48\sigma^6} + \frac{\beta_{re}^8}{384\sigma^8} - \frac{\beta_{im}^2}{2\sigma^2} + \frac{\beta_{im}^4}{8\sigma^4} + \\ & - \frac{\beta_{im}^6}{48\sigma^6} + \frac{\beta_{im}^8}{384\sigma^8} + \frac{\beta_{re}^2\beta_{im}^2}{4\sigma^4} + \frac{\beta_{re}^2\beta_{im}^4}{16\sigma^6} - \frac{\beta_{re}^4\beta_{im}^2}{16\sigma^6} + \\ & + \frac{\beta_{re}^2\beta_{im}^6}{96\sigma^8} + \frac{\beta_{re}^4\beta_{im}^4}{64\sigma^8} + \frac{\beta_{re}^6\beta_{im}^2}{96\sigma^8} + \dots] \end{aligned} \quad (9)$$

One could group the terms with σ^2 in the denominator and define h_{σ^4} as a variable containing the high-order terms of the summation. Then, it is possible to write

$$V_\sigma^C(\beta_{re}, \beta_{im}) = \frac{1}{2\pi\sigma} - \frac{1}{2\pi\sigma} E \left[\frac{\beta_{re}^2 + \beta_{im}^2}{2\sigma^2} \right] + h_{\sigma^4} \quad (10)$$

$$V_\sigma^C(X, Y, Z, S) = \frac{1}{2\pi\sigma} - \frac{1}{2\pi\sigma} \frac{1}{2\sigma^2} E[(X - Y)^2 + (Z - S)^2] + h_{\sigma^4} \quad (11)$$

$$V_\sigma^C(C_1, C_2) = \frac{1}{2\pi\sigma} - \frac{1}{4\pi\sigma^3} E[(C_1 - C_2)(C_1 - C_2)^*] + h_{\sigma^4} \quad (12)$$

$$V_\sigma^C(C_1, C_2) = \frac{1}{2\pi\sigma} - \frac{1}{4\pi\sigma^3} R[C_1, C_2] + h_{\sigma^4} \quad (13)$$

where $R[C_1, C_2] = E[C_1 C_2^*]$ is the correlation between C_1 and C_2 . One can notice in equation (13) that the higher-order terms represented by h_{σ^4} tend to zero faster than the second one as σ increases, what corresponds to the correlation involving two complex variables C_1 and C_2 , which completes the proof. ■

95 *Property 4:* When using the Gaussian Kernel with kernel size σ , decreasing σ to zero causes complex correntropy to approach the value associated to the probability density of the event $C_1 = C_2$ ($P(C_1 = C_2)$).

$$\lim_{\sigma \rightarrow 0} V_\sigma^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(u_1, u_1, u_2, u_2) du_1 du_2 = P(C_1 = C_2) \quad (14)$$

where $f_{XYZS}(u_1, u_1, u_2, u_2)$ is the joint PDF.

100 *Proof:* Let us recall the complex correntropy definition presented in equation (4) i.e. $V^C(C_1, C_2) = E_{C_1 C_2}[K(C_1, C_2)]$. Two complex numbers are equal when both their real and imaginary parts are also equal to each other. Then, using the Gaussian kernel and expanding the terms associated C_1 and C_2 , one can obtain

$$V_\sigma^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(x, y, z, s) G_\sigma(x - y) G_\sigma(z - s) dx dy dz ds \quad (15)$$

105 From the theory of distributions, the following statement can be demonstrated

$$\lim_{\sigma \rightarrow 0} G_\sigma(x) \equiv \delta(x) \quad (16)$$

Then, it gives

$$\lim_{\sigma \rightarrow 0} V_\sigma^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(x, y, z, s) \delta(x - y) \delta(z - s) dx dy dz ds \quad (17)$$

Making $x = y = u_1$ and $z = s = u_2$ results in

$$\lim_{\sigma \rightarrow 0} V_\sigma^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XYZS}(u_1, u_1, u_2, u_2) du_1 du_2 = P(C_1 = C_2) \quad (18)$$

■

Complex correntropy was defined in order to maintain the probabilistic meaning of correntropy. Equation (18) represents a variable transformation of the joint PDF, f_{XYZS} , which is evaluated with equal arguments u_1 and u_2 . By using a Parzen method in (Parzen, 1962) to estimate f_{XYZS} , when σ goes to zero and the product $N\sigma$ tends to infinity, respectively, the estimative of the joint PDF, $\hat{f}_{XYZS}(x, y, z, s)$, approaches $f_{XYZS}(x, y, z, s)$ asymptotically in the mean square sense. Therefore, using the Gaussian Kernel with a small kernel size in the Parzen estimation of f_{xyzs} causes equation (18) to provide a scalar value that approaches the probability density of the event ($C_1 = C_2$). A thorough discussion on this probabilistic meaning is presented in (Guimaraes et al., 2017).

Property 4.1: Assuming i.i.d data $\{(x_i, y_i, z_i, s_i)\}_{i=1}^N$ draw from the joint PDF f_{xyzs} , while $\hat{f}_{\sigma xyzs}$ is its respective Parzen estimate with kernel size σ , the complex correntropy estimated with kernel size $\sigma' = \sigma\sqrt{2}$ is the integral of $\hat{f}_{\sigma xyzs}$ along the plane formed by $x = y$ and $z = s$.

$$\hat{V}_{\sigma'}^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{\sigma XYZS}(x, y, z, s) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2} \quad (19)$$

Proof. For simplicity, the step-by-step proof is presented in detail in Appendix I.

Property 5: Under the condition $N \rightarrow \infty$, $\hat{V}_{N,\sigma}(C_1, C_2)$ is an estimator of $V_{\sigma}(C_1, C_2)$ consistent in mean square. Furthermore, under conditions $N\sigma \rightarrow \infty$ and $\sigma \rightarrow 0$, $\hat{V}_{N,\sigma}(C_1, C_2)$ is an asymptotically unbiased estimator for p_{β} and consistent in mean square.

Proof. Following the properties of the Parzen estimation (Parzen, 1962), one can say that, if the conditions

$$\lim_{n \rightarrow \infty} \sigma = 0; \quad (20)$$

holds, then

$$\lim_{n \rightarrow \infty} \hat{f}_{\sigma XYZS}(x, y, z, s) = f_{\sigma XYZS}(x, y, z, s) \quad (21)$$

Thus,

$$E[\hat{V}_{N,\sigma}^C(C_1, C_2)] = V_{\sigma}^C(C_1, C_2) \quad (22)$$

Recall the random variable $\beta = C_1 - C_2$ with PDF $p_{\beta}(b)$. By using the Gaussian kernel and holding the conditions $N \rightarrow \infty$ and $\sigma \rightarrow 0$, it is possible

to write (see property 4)

$$\lim_{N \rightarrow \infty, \sigma \rightarrow 0} E[\hat{V}_{N,\sigma}^C(C_1, C_2)] = P(C_1 = C_2) = p_\beta(0) \quad (23)$$

Furthermore, to analyze the variance, it is also needed to consider the Parzen estimator properties shown on (Parzen, 1962), then

$$var[\hat{V}_{N,\sigma}^C(C_1, C_2)] = N^{-1} var[G_\sigma(C_1 - C_2)] \quad (24)$$

$$\lim_{N \rightarrow \infty, \sigma \rightarrow 0} N\sigma var[\hat{V}_{N,\sigma}^C(C_1, C_2)] = p_\beta(0) \int G_1^2(u) du \quad (25)$$

where $G_1(u)$ is the Gaussian Kernel with $\sigma = 1$.

140

■

Property 6: For two real random variables, R_1 and $R_2 \in \mathbb{R}$, using the Gaussian Kernel, complex correntropy $\hat{V}_\sigma^C(R_1, R_2)$ differs from conventional correntropy $V_\sigma(R_1, R_2)$ by a factor of $(\sqrt{2\pi}\sigma)$.

$$\hat{V}_\sigma^C(R_1, R_2)\sqrt{2\pi}\sigma = \hat{V}_\sigma(R_1, R_2) \quad (26)$$

Proof. Let us expand the left-hand side of equation (26) as

$$\hat{V}_\sigma^C(R_1, R_2)\sqrt{2\pi}\sigma = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2 + (z_n - s_n)^2}{2\sigma^2}\right) \sqrt{2\pi}\sigma \quad (27)$$

145

If R_1 and R_2 do not have imaginary parts, one can write equation (27) as

$$\hat{V}_\sigma^C(R_1, R_2)\sqrt{2\pi}\sigma = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(x_n - y_n)^2}{2\sigma^2}\right) \quad (28)$$

which is exactly the expression to estimate $V_\sigma(R_1, R_2)$ using the Gaussian Kernel (Principe, 2010).

■

150 It is important to mention that, using the Gaussian kernel, one could expect complex correntropy to generalize the real correntropy also numerically, as to be equal to each other when the imaginary parts of the data be equal to zero. That is not the case because, as an estimate, the error in the estimation causes the Gaussians to “spreads” out differently in the 4 dimensional case (two complex variables) than it does in two.

155 *Property 7:* Using the Gaussian kernel, the complex correntropy can be expressed in polar coordinates as

$$\hat{V}_\sigma^C(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp \left(-\frac{|C_1|^2 + |C_2|^2}{2\sigma^2} - \frac{|C_1||C_2|\cos(\theta - \phi)}{\sigma^2} \right) \quad (29)$$

Proof. Since C_1 and C_2 can be written in the polar form as $C_1 = |C_1|\cos(\phi) + j|C_1|\sin(\phi)$, and $C_2 = |C_2|\cos(\theta) + j|C_2|\sin(\theta)$, equation (8) can now be represented as

$$\begin{aligned} (x - y)^2 &= (|C_1|\cos(\theta) - |C_2|\cos(\phi))^2 = \\ &= |C_1|^2\cos(\theta)^2 - 2|C_1||C_2|\cos(\theta)\cos(\phi) + |C_2|^2\cos(\phi)^2 \end{aligned} \quad (30)$$

160 Then, it is possible to expand term $(z - s)^2$ as

$$(z - s)^2 = (|C_1|\sin(\theta) - |C_2|\sin(\phi))^2 = |C_1|^2\sin(\theta)^2 - 2|C_1||C_2|\sin(\theta)\sin(\phi) + |C_2|^2\sin(\phi)^2 \quad (31)$$

Now, let us make $(x - y)^2 + (z - s)^2$ by adding equation (30) and (31) i.e.

$$|C_1|^2 + |C_2|^2 - 2|C_1||C_2|(\cos(\theta)\cos(\phi) + \sin(\theta)\sin(\phi)) = |C_1|^2 + |C_2|^2 - 2|C_1||C_2|\cos(\theta - \phi) \quad (32)$$

■

3. Maximum Complex Correntropy Criterion

165 Complex correntropy has been defined as a robust similarity measure between two complex random variables (Guim  raes et al., 2017). Let us

consider a linear model and define the error $e = d - y$ as the difference between the desired signal d and the estimated output $y = \mathbf{w}^H \mathbf{X}$, where $y, d, e, \mathbf{X}, \mathbf{w} \in \mathbb{C}$, and the superscript H is the Hermitian (conjugate transpose).

- 170 Let the new criteria MCCC be defined as the maximum complex correntropy between two random complex variables D and Y (Guimãraes et al., 2017).

$$J_{MCCC} = V_\sigma^C(D, Y) = E_{DY}[G_\sigma^C(D - \mathbf{w}^H \mathbf{X})] = E_{DY}[G_\sigma^C(e)] \quad (33)$$

where w is the parameter that controls the error between the estimated and the desired signal, as G_σ^C is the complex Gaussian kernel, defined in equation (7).

- 175 One can obtain a fixed-point solution for the optimal weight by using equation (33) as a cost function. Then, the natural way would be to set the cost function derivative to zero with respect to \mathbf{w}^* . However, even depending on complex-valued parameters (D, Y) , as property 2 shows, using the Gaussian kernel makes complex correntropy a real-valued function. Then, 180 equation (33) is not analytical in the complex domain because the Cauchy-Riemann conditions are violated (Mandic and Goh, 2009). Hence, standard differentiation can not be apply. One possible alternative to overcome this problem is to consider the cost function defined in an Euclidean domain with a double dimensionality (\mathbb{R}^2), although this approach leads to onerous 185 computations (Bouboulis and Theodoridis, 2011). The Wirtinger Calculus, which will be briefly presented in this section later on, provide an elegant way to obtain a gradient of real-valued cost function that are defined in complex domains (Mandic and Goh, 2009; Bouboulis and Theodoridis, 2011).

3.1. Wirtinger calculus

- 190 The Wirtinger calculus was firstly introduced in (Wirtinger, 1927), based on the duality between spaces \mathbb{C} and \mathbb{R}^2 .

Let $f : \mathbb{C} \rightarrow \mathbb{C}$ be a complex function defined in \mathbb{C} . Such function can also be defined in \mathbb{R}^2 (i.e., $f(x + jy) = f(x, y)$).

- 195 The Wirtinger's derivative of f at a point c is defined as follows (Bouboulis and Theodoridis, 2011)

$$\frac{\partial f}{\partial z}(c) = \frac{1}{2} \left(\frac{\partial f}{\partial x}(c) - j \frac{\partial f}{\partial y}(c) \right) \quad (34)$$

while the Conjugate Wirtinger's derivative of f at c is defined by

$$\frac{\partial f}{\partial z^*}(c) = \frac{1}{2} \left(\frac{\partial f}{\partial x}(c) + j \frac{\partial f}{\partial y}(c) \right) \quad (35)$$

The work developed in (Bouboulis and Theodoridis, 2011) presents a complete overview on the properties associated to Wirtinger Calculus. In summary, in order to compute the Wirtinger derivative of a given function f , which is expressed in terms of z and z^* , one should apply the usual differentiation rules after considering z^* as a constant. Additionally, following the same concept to compute the Conjugate Wirtinger derivative of a function f , also expressed in terms of z and z^* , one should apply the usual differentiation rules considering z as a constant (Bouboulis and Theodoridis, 2011).
 200 For example, considering f as $f(z) = zz^*$ allows concluding that
 205

$$\frac{\partial f}{\partial z} = z^* \quad \text{and} \quad \frac{\partial f}{\partial z^*} = z \quad (36)$$

3.2. Fixed-point solution

The MCCC fixed-point algorithm was firstly introduced and discussed in (Guimaraes et al., 2017). By using Wirtinger calculus, it is possible to set the derivative of the cost function to zero i.e.

$$\frac{\partial J_{MCCC}}{\partial w^*} = \frac{\partial E_{DY}[G_{\sigma\sqrt{2}}^C(e)]}{\partial w^*} = E_{DY} \left[G_{\sigma\sqrt{2}}^C(e) \frac{\partial(ee^*)}{\partial w^*} \right] = \mathbf{0} \quad (37)$$

210 Which gives

$$\frac{\partial(ee^*)}{\partial w^*} = \frac{\partial(d - \mathbf{w}^H \mathbf{X})(d^* - \mathbf{w}^T \mathbf{X}^*)}{\partial \mathbf{w}^*} = (-d^* \mathbf{X} + \mathbf{X} \mathbf{X}^H \mathbf{w}) \quad (38)$$

Then, it is possible to substitute (38) in (37) as

$$\begin{aligned} E_{DY}[G_{\sigma\sqrt{2}}^C(e)(-d^* \mathbf{X} + \mathbf{X} \mathbf{X}^H \mathbf{w})] &= \mathbf{0} \\ E_{DY}[G_{\sigma\sqrt{2}}^C(e) d^* \mathbf{X}] &= E_{DY}[G_{\sigma\sqrt{2}}^C(e) \mathbf{X} \mathbf{X}^H] \mathbf{w} \end{aligned} \quad (39)$$

Finally, the final expression for w is obtained as

$$\mathbf{w} = \left[\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) \mathbf{X}_n \mathbf{X}_n^H \right]^{-1} \left[\sum_{n=1}^N G_{\sigma\sqrt{2}}^C(e_n) d_n^* \mathbf{X}_n \right] \quad (40)$$

A detailed step-by-step solution is presented in Appendix II.

4. Simulations and results

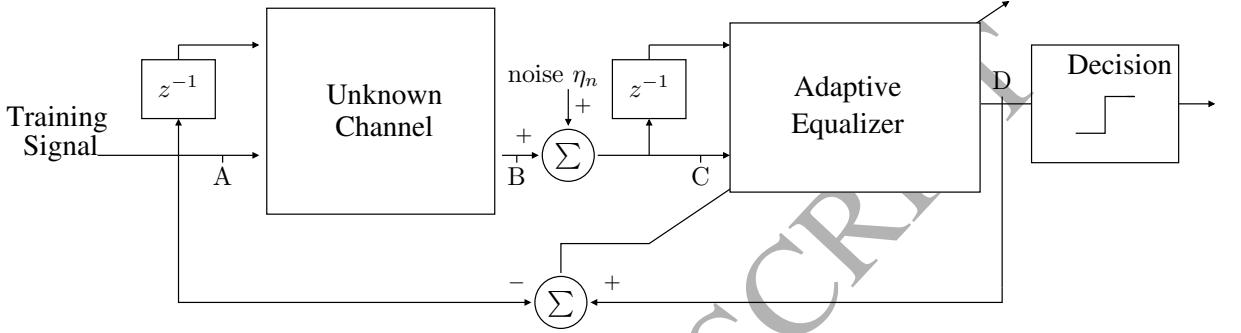


Figure 1: Block diagram representing a channel equalization problem.

215 In order to evaluate the performance of the MCCC, a comparison has been established involving traditional algorithms such as CLMS (Mandic and Goh, 2009), CRLS (Diniz, 2008), and LAD (Alliney and Ruzinsky, 1994) in a channel equalization problem. Figure 1 shows the block diagram of a communication system with adaptive channel equalization.

220 The in-phase component $A_{k,I}$ and quadrature component $A_{k,Q}$ at k th instant are transmitted through a delay line to obtain the current and past signals. The training signal is formed by a sequence of T -spaced complex symbols associated to a 16-QAM (Quadrature Amplitude Modulation) constellation whose values $(\pm 1 \pm j1, \pm 2 \pm j2, \dots, \pm 8 \pm j8)$ are seen in Figure 2-a. This signal is then applied to an unknown channel W modeled as $W = [w_1, w_2]^T$, where w_1 and $w_2 \in \mathbb{C}$. This generates B_k , which is represented as

$$B_K = [A_K \ A_{K-1}] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \quad (41)$$

225 where w_1 and w_2 had their values arbitrary assigned to $(1.1 - j1.1)$ and $(0.9 - j0.2)$, respectively. The distorted signal B_K can be observed in Figure 2-b.

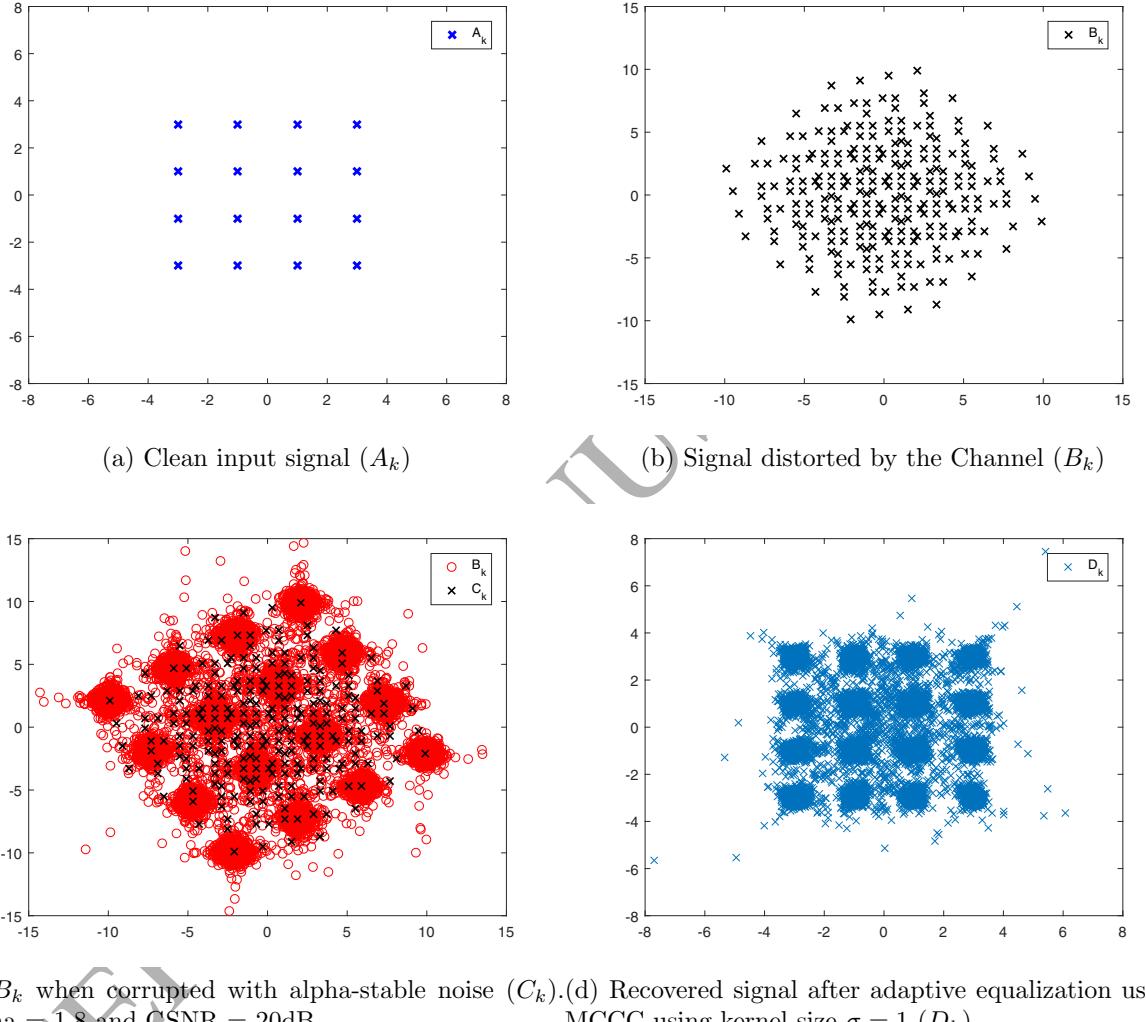


Figure 2: Data flow analysis, where each letter represents the corresponding stage in Figure 1.

230 Thus, in order to assess the robustness of the methods to outliers, B_K is corrupted with an additive noise η_k producing signal $C_K = B_K + \eta_K$ as shown in Figure 2-c. Due to the capability from the alpha-stable distribution to model impulsive noise (Shao and Nikias, 1993), this distribution was chosen to model η_k . The impulsive noise strength is measured by the generalized

235 signal to noise ratio (GSNR) (Tsakalides and Nikias, 1995) defined as

$$GSNR = 10 \log_{10} \left(\frac{1}{\gamma M} \sum_{t=1}^M |s(t)|^2 \right) \quad (42)$$

where γ is the dispersion parameter of α -stable noise. Different GSNR values varying from 10 dB to 20 dB have been used in the simulation tests. It is worth to mention that the result in Figure 2-c was obtained when GSNR=20 dB and $\alpha = 1.8$.

240 Signal C_k and its respective delay are used as inputs of the adaptive equalizer, as well as the error between the training signal and the equalizer output D_K . In the simulations, W is always initialized with zeros. While MCCC and also CRLS are fixed-point solutions, both CLMS and CMOD techniques employ an ascendant gradient with learning rate of 0.01. After 245 the equalization process with the training signal, the parameters found by each algorithm at every iteration are used to generate a Bit Error Rate (BER) for all tested GSNR levels. In order to compare the results in Figure 3-a with the ones provided by classical methods e.g. CLMS, CRLS, and LAD, three different values of σ are chosen i.e. 1, 10, and 100. The performances of 250 algorithms have been evaluated for an average of 10^5 Monte Carlo trials in an environment with $\alpha = 1.5$ and GSNR varying from 10 to 20 db.

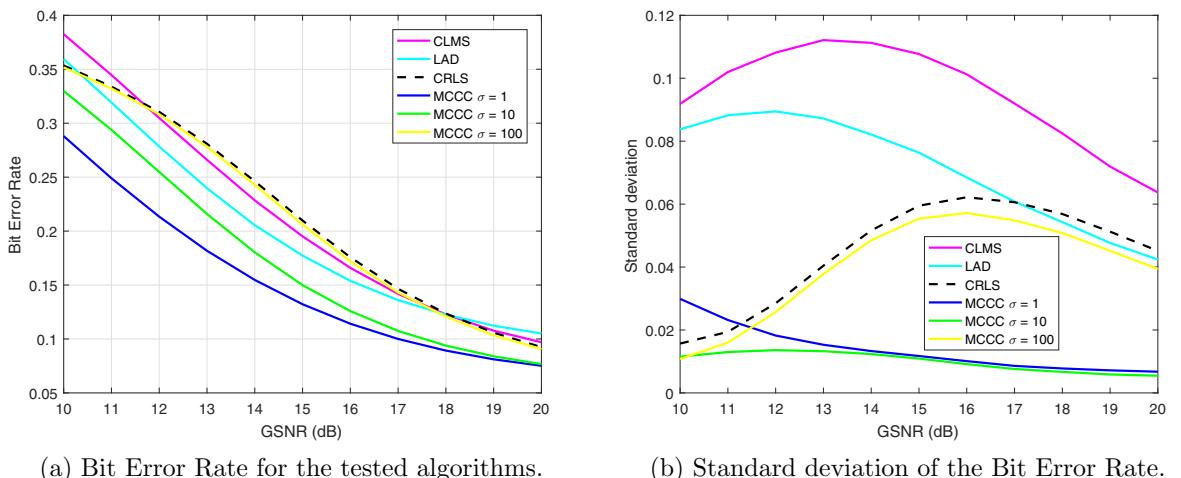


Figure 3: BER (left) and respective standard deviation (right) resulting from 10^5 iterations of the tested algorithms for different GSNR scenarios.

As it can be seen in Figure 3-a, all methods are able to achieve equivalent good results for high values of GSNR. However, as this parameter decreases associated to the nature of the selected type of noise with α -stable distribution, the probability of existence of outliers increases. This type of environment degrades the performance level of second-order methods such as CRLS and CLMS. Figure 4 shows the BER performance regarding MCCC for a kernel size $\sigma = 1$ and different values of alpha, which vary from 1 to 2. According to property 3, when using the Gaussian kernel, the complex correntropy is a weighted sum of all the even moments of the probability density estimation from event $C_1 = C_2$, what makes MCCC robust to outliers similarly to LAD.

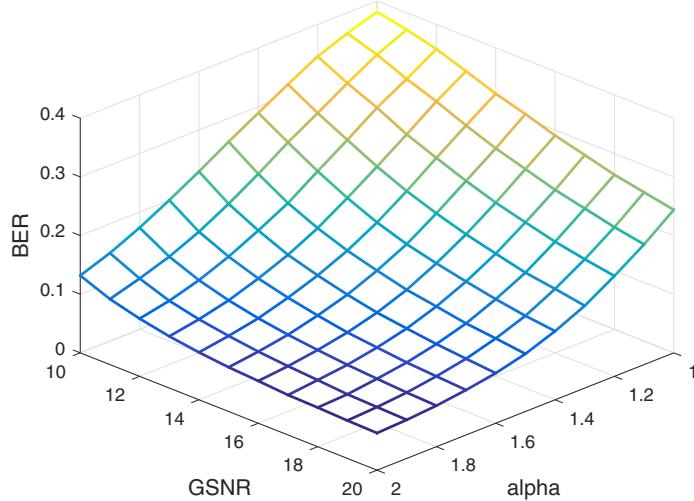


Figure 4: BER for the MCCC algorithm with fixed kernel size $\sigma = 1$ for distinct GSNR scenarios and values of alpha.

Additionally, the standard deviation of the Monte Carlo trials in Figure 3-a is shown Figure 3-b. One can notice that, for all tested kernel sizes and GSNR values, MCCC has the smallest standard deviation among all tested algorithms thus demonstrating its robustness. It is also worth to mention that this method has fast convergence rates since it is a fixed-point method (Guimaraes et al., 2017). According to Figure 3-a, the MCCC performance regarding the channel equalization problem is strictly related to the selection

270 of a proper kernel size, which is a free parameter and must be wisely chosen by
 the user. The best results could be obtained with a kernel size $\sigma = 1$, where
 Figure 2-d shows the undistorted signal D_K produced by the MCCC method
 with $\sigma = 1$ when $GSNR = 20dB$. Therefore, one can notice, for example,
 that the LAD algorithm outperforms MCCC when $\sigma = 100$ for some values of
 275 GSNR. According to property 4, when the kernel size is increased, complex
 correntropy tends to behave as the correlation analogously to correntropy
 in the real-valued case. Then, as expected, one can notice how similar the
 MCCC performance is to that regarding CRLS when $\sigma = 100$. In order to
 highlight the importance of the kernel size adjustment, Figure 5 is presented
 280 to show a surface formed by each BER curve created with different kernel
 sizes within the range from 5 to 50.

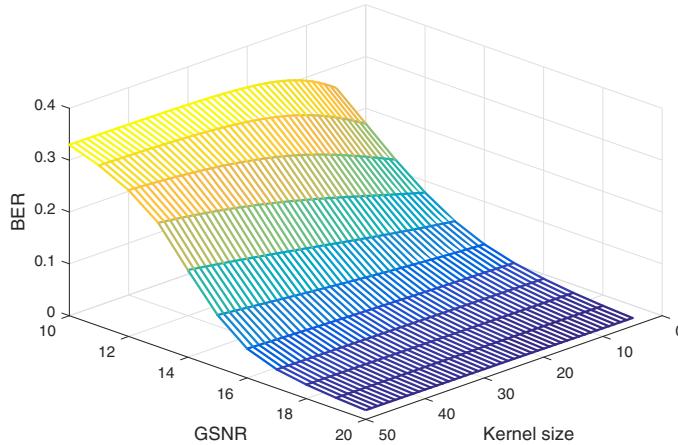


Figure 5: Influence of different kernel sizes on the BER behavior for the MCCC algorithm with different values of GSNR, where $\alpha = 1.5$ is fixed.

Once the kernel size is adjusted, the MCCC method is able to achieve better BER levels than the classic algorithms for all tested values of GSNR, being an effective tool to deal with outlier environments.

285 **5. Conclusions**

This paper has expanded the definition of complex correntropy to positive-definite kernels. Furthermore, a complete derivation of this similarity mea-

sure is presented as well as important properties such as symmetry, boundedness, high-order statistical measure, probabilistic meaning, mathematical relationship with classical correntropy, and polar coordinate representation. This expanded definition together with the properties makes complex correntropy structurally similar to the correntropy function. For positive valued kernels with complex arguments, complex correntropy could be used as a generalization of conventional real-valued arguments correntropy.

A brief review on Wirtinger calculus has been discussed in order to use the complex correntropy as a cost function in a channel equalization problem, thus resulting in the MCCC algorithm, which was analyzed in detail. Therefore, the advantages associated to the use of complex correntropy in nonGaussian signal processing have been clearly demonstrated both theoretically and experimentally.

The results obtained from the channel equalization problem have shown the improved performance of the complex correntropy function in nonGaussian environments when compared to classic algorithms such as LAD, CLMS, and CRLS. After setting the proper kernel size, the MCCC method has been able to overcome the performance of other similar algorithms, thus providing better values of BER for all tested scenarios. Being a fixed-point solution, it also provides MCCC with a fast convergence rate. There are several potential applications that may explore the benefits from this similarity measure. For instance, spectrum sensing and automatic modulation classification are interesting lines for further research. Another approach lies in understanding complex correntropy when using the theory of reproducing kernel Hilbert space.

References

References

- Alliney, S., Ruzinsky, S.A., 1994. An algorithm for the minimization of mixed l1 and l2 norms with application to bayesian estimation. *IEEE Transactions on Signal Processing* 42, 618–627. doi:10.1109/78.277854.
- Bouboulis, P., Theodoridis, S., 2011. Extension of wirtinger's calculus to reproducing kernel hilbert spaces and the complex kernel lms. *IEEE Transactions on Signal Processing* 59, 964–978. doi:10.1109/TSP.2010.2096420.

- Chen, B., Xing, L., Liang, J., Zheng, N., Prncipe, J.C., 2014. Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion. *IEEE Signal Processing Letters* 21, 880–884. doi:10.1109/LSP.2014.2319308.
- 325 Diniz, P., 2008. Adaptive Filtering: Algorithms and Practical Implementation. Kluwer international series in engineering and computer science, Springer.
- Fontes, A.I., de M. Martins, A., Silveira, L.F., Neto, A.D., Souza, P.T., Príncipe, J., 2014. Classification system of pathological voices using correntropy. *Mathematical Problems in Engineering* 42, 1–8.
- 330
- Fontes, A.I., de M. Martins, A., Silveira, L.F., Príncipe, J., 2015. Performance evaluation of the correntropy coefficient in automatic modulation classification. *Expert Systems with Applications* 42, 1–8.
- Fontes, A.I., Rego, J.B., Martins, A.d.M., Silveira, L.F., Príncipe, J.C., 2017.
- 335 Cyclostationary correntropy: Definition and applications. *Expert Systems with Applications* 69, 110–117.
- Guimãraes, J.P.F., Fontes, A.I.R., Rego, J.B.A., de M. Martins, A., Príncipe, J.C., 2017. Complex correntropy: Probabilistic interpretation and application to complex-valued data. *IEEE Signal Processing Letters* 24, 42–45. doi:10.1109/LSP.2016.2634534.
- 340
- He, R., Zheng, W.S., Hu, B.G., 2011. Maximum correntropy criterion for robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 1561–1576.
- J. Mercer, B.A., 1909. Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 209, 415–446. URL: <http://rsta.royalsocietypublishing.org/content/209/441-458/415>, doi:10.1098/rsta.1909.0016, arXiv:<http://rsta.royalsocietypublishing.org/content/209/441-458/415.full.pdf>.
- 345
- 350 Liu, W., Pokharel, P.P., Príncipe, J.C., 2006. Correntropy: A localized similarity measure, in: The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 4919–4924. doi:10.1109/IJCNN.2006.247192.

- 355 Liu, W., Pokharel, P.P., Príncipe, J.C., 2007. Correntropy: Properties and
applications in non-gaussian signal processing. *IEEE Transactions on Signal Processing* 55, 5286–5298.
- 360 Liu, X., Chen, B., Xu, B., Wu, Z., Honeine, P., 2017a. Maximum correntropy unscented filter. *International Journal of Systems Science* 48, 1607–1615. URL: <https://doi.org/10.1080/00207721.2016.1277407>, doi:10.1080/00207721.2016.1277407, arXiv:<https://doi.org/10.1080/00207721.2016.1277407>.
- 365 Liu, X., Chen, B., Zhao, H., Qin, J., Cao, J., 2017b. Maximum correntropy kalman filter with state constraints. *IEEE Access* 5, 25846–25853. doi:10.1109/ACCESS.2017.2769965.
- 370 Ma, W., Zheng, D., Zhang, Z., Duan, J., Chen, B., 2017. Robust proportionate adaptive filter based on maximum correntropy criterion for sparse system identification in impulsive noise environments. *Signal, Image and Video Processing* URL: <https://doi.org/10.1007/s11760-017-1137-0>, doi:10.1007/s11760-017-1137-0.
- 375 Mandic, D., Goh, V., 2009. Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models. *Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control*, Wiley.
- 380 Parzen, E., 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33, pp. 1065–1076. URL: <http://www.jstor.org/stable/2237880>.
- Peng, S., Chen, B., Sun, L., Ser, W., Lin, Z., 2017. Constrained maximum correntropy adaptive filtering. *Signal Processing* .
- Príncipe, J., 2010. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. *Information Science and Statistics*, Springer New York.
- 385 Santamaria, I., Pokharel, P., Príncipe, J., 2006. Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Transactions on Signal Processing* 54, 2187–2197.

- 385 Shao, M., Nikias, C.L., 1993. Signal processing with fractional lower order moments: stable processes and their applications. *Proceedings of the IEEE* 81, 986–1010. doi:10.1109/5.231338.
- 390 Tsakalides, P., Nikias, C.L., 1995. Maximum likelihood localization of sources in noise modeled as a stable process. *IEEE Transactions on Signal Processing* 43, 2700–2713. doi:10.1109/78.482119.
- 395 Wang, F., He, Y., Wang, S., Chen, B., 2017. Maximum total correntropy adaptive filtering against heavy-tailed noises. *Signal Processing* 141, 84 – 95. URL: <http://www.sciencedirect.com/science/article/pii/S0165168417302001>, doi:<https://doi.org/10.1016/j.sigpro.2017.05.029>.
- Wirtinger, W., 1927. Zur formalen theorie der funktionen von mehr komplexen vernderlichen. *Mathematische Annalen* 97, 357–376. URL: <http://eudml.org/doc/182642>.

Appendix I

400 *Property 4.1:* Assuming i.i.d data $\{(x_i, y_i, z_i, s_i)\}_{i=1}^N$ are drawn from the joint PDF f_{xyzs} , and $\hat{f}_{\sigma xyzs}$ is its Parzen estimate with kernel size σ , the complex correntropy estimated with kernel size $\sigma' = \sigma\sqrt{2}$ is the integral of $\hat{f}_{\sigma xyzs}$ along the plane formed by $x = y$ and $z = s$.

$$\hat{V}_{\sigma'}^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{\sigma XYZS}(x, y, z, s) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2} \quad (43)$$

Proof.

405 Let us assume two complex random variables $C_1 = X + jZ$ and $C_2 = Y + jS$, where $C_1, C_2 \in \mathbb{C}$, and X, Y, Z, S are real-valued random variables. Complex correntropy was developed to keep the probabilistic meaning from correntropy. So, estimate the complex correntropy is also estimate the probability density of the event $C_1 = C_2$, which implies in estimating the probability density of the event $X = Y$ and $Z = S$. The reasoning is that two complex numbers are equal when their real parts are equal and their imaginary parts are also equal to each other, then:

$$\hat{P}(C_1 = C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, y, z, s) \delta(x - y) \delta(z - s) dx dy dz ds \quad (44)$$

When $x = y$ and $z = s$, equation (44) can be rewritten as:

$$\hat{P}(C_1 = C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(x, y, z, s) du_1 du_2 \Big|_{x=y=u_1, z=s=u_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{XYZS}(u_1, u_1, u_2, u_2) du_1 du_2 \quad (45)$$

which is the right side of the equation (43).

415 Let us recall equation (46), which estimates correntropy from data using a Gaussian Kernel.

$$\hat{V}_{\sigma}^C(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp \left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2} \right) \quad (46)$$

Then, to complete the proof, one should obtain equation (46) by solving the double integral from equation (43).

First, let us replace \hat{f}_{XYZS} for the Parzen estimator defined as:

$$\hat{f}_{X^1, X^2, \dots, X^L}(x^1, x^2, \dots, x^L) = \frac{1}{N} \sum_{n=1}^N \prod_{l=1}^L G_\sigma(x^l - x_n^l) \quad (47)$$

420 where

$$G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Then, using $L = 4$ in equation (47) and then replacing \hat{f}_{XYZS} in equation (45):

$$\hat{V}_\sigma^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^N G_\sigma(u_1 - x_n) G_\sigma(u_1 - y_n) G_\sigma(u_2 - z_n) G_\sigma(u_2 - s_n) du_1 du_2 \quad (48)$$

The best way to solve this double integral is rewrite equation (48) is make:

$$\hat{V}_\sigma^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(4\pi^2\sigma^4)} \exp\left(-\frac{1}{2\sigma^2}(a(u_1 - b)^2 + a'(u_2 - b')^2 + c)\right) du_1 du_2$$

and then

$$\hat{V}_\sigma^C(C_1, C_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(4\pi^2\sigma^4)} \exp\left(-\frac{1}{2\sigma^2}a(u_1 - b)^2\right) \exp\left(-\frac{1}{2\sigma^2}a'(u_2 - b')^2\right) \exp\left(-\frac{c}{2\sigma^2}\right) du_1 du_2 \quad (49)$$

425 The coefficients a, b, a', b' and c are obtained by the completing square technique, detailed as follows:

$$(u_1 - x_i)^2 + (u_1 - y_i)^2 + (u_2 - z_i)^2 + (u_2 - s_i)^2 = a(u_1 - b)^2 + a'(u_2 - b')^2 + c \quad (50)$$

Developing the right side of the equation (50):

$$au_1^2 + a'u_2^2 - 2u_1ab - 2u_2a'b' + ab^2 + a'b'^2 + c$$

And then, the left side:

$$2u_1^2 + 2u_2^2 + x_i^2 + y_i^2 + z_i^2 + s_i^2 - 2u_1(x_i + y_i) - 2u_2(z_i + s_i)$$

One can say that:

$$a = a' = 2; b = \frac{(x_i + y_i)}{2}; b' = \frac{(z_i + s_i)}{2}; \text{ and } c = \frac{(x_i - y_i)^2 + (z_i - s_i)^2}{2}$$

430 Using these values in the equation (49) we have:

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^N \frac{1}{(4\pi^2\sigma^4)} \exp\left(-\frac{(u_1 - b)^2}{\sigma^2}\right) \exp\left(-\frac{(u_2 - b')^2}{\sigma^2}\right) \exp\left(-\frac{c}{2\sigma^2}\right) du_1 du_2 \quad (51)$$

Making $\sigma^2 = 2\theta^2$, one can rewrite equation (51) as:

$$= \int \int \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2\pi\theta^2}\right) \left(\frac{1}{2\pi 4\theta^2}\right) / \exp\left(-\frac{(u_1 - b)^2}{2\theta^2}\right) \exp\left(-\frac{(u_2 - b')^2}{2\theta^2}\right) \exp\left(-\frac{c}{4\theta^2}\right) du_1 du_2 \quad (52)$$

But

$$\left(\frac{1}{\sqrt{2\pi}\theta}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(u_1 - b)^2}{2\theta^2}\right) du_1 \left(\frac{1}{\sqrt{2\pi}\theta}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(u_2 - b')^2}{2\theta^2}\right) du_2 = 1 \quad (53)$$

Then, equation (52) can be rewrite as:

$$\hat{V}_{\sigma}^C(C_1, C_2) = \left(\frac{1}{2\pi 4\theta^2}\right) \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(x_i - y_i)^2 + (z_i - s_i)^2}{8\theta^2}\right)$$

435 Or, making $8\theta^2 = (2\sigma')^2$:

$$\hat{V}_{\sigma}^C(C_1, C_2) = \left(\frac{1}{2\pi(\sigma')^2}\right) \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(x_i - y_i)^2 + (z_i - s_i)^2}{2(\sigma')^2}\right) \quad (54)$$

Where $\sigma' = \sigma\sqrt{2}$, which implies

$$V_\sigma^C(C_1, C_2) = \frac{1}{N} \sum_{n=1}^N G_{\sigma\sqrt{2}}(x_n - y_n) G_{\sigma\sqrt{2}}(z_n - s_n)$$

The non parametric estimator of complex correntropy with Parzen windows can be written as

$$\hat{V}_\sigma^C(C_1, C_2) = E[G_\sigma^C(C_1 - C_2)] \quad (55)$$

where

$$G_\sigma^C(C_1 - C_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_n - y_n)^2 + (z_n - s_n)^2}{2\sigma^2}\right) \quad (56)$$

alternatively,

$$\hat{V}_\sigma^C(C_1, C_2) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (57)$$

which completes the proof. ■

Appendix II

MCCC Fixed Point Solution:

Considering a linear model and define the error, $e = d - y$, as the difference between the desired signal d and the filter output $y = \mathbf{w}^H \mathbf{X}$, with $y, d, e, \mathbf{X}, \mathbf{w} \in \mathbb{C}$, let the new criteria MCCC be defined as the maximum complex correntropy between two random complex variables D and Y :

$$J_{MCCC} = V_\sigma^C(D, Y) = E_{DY}[G_{\sigma\sqrt{2}}^C(D - \mathbf{w}^H \mathbf{X})] = E_{DY}[G_{\sigma\sqrt{2}}^C(e)]$$

For archiving the FP solution for the optimal weights, one can set the cost function derivative to zero in respect to \mathbf{w}^*

$$\frac{\partial J_{MCCC}}{\partial \mathbf{w}^*} = \frac{\partial E_{DY}[G_{\sigma\sqrt{2}}^C(e)]}{\partial \mathbf{w}^*} = E_{DY} \left[G_{\sigma\sqrt{2}}^C(e) \frac{\partial(ee^*)}{\partial \mathbf{w}^*} \right] = \mathbf{0}$$

$$\frac{\partial(ee^*)}{\partial w^*} = \frac{\partial(D - \mathbf{w}^H \mathbf{X})(D - \mathbf{w}^H \mathbf{X})^*}{\partial w^*}$$

$$\frac{\partial(ee^*)}{\partial w^*} = \frac{\partial(D - \mathbf{w}^H \mathbf{X})(D^* - \mathbf{w}^T \mathbf{X}^*)}{\partial w^*}$$

Then, let us find the derivative of ee^* in respect to w^* using Wirtinger Calculus:

$$\frac{\partial(ee^*)}{\partial w^*} = \frac{\partial(DD^* - D\mathbf{X}^H w - \mathbf{w}^H \mathbf{X} \mathbf{X}^* + \mathbf{w}^H \mathbf{X} \mathbf{X}^H \mathbf{w})}{\partial w^*}$$

$$\frac{\partial DD^*}{\partial \mathbf{w}^*} = 0$$

$$\frac{\partial D\mathbf{X}^H w}{\partial \mathbf{w}^*} = 0$$

$$\frac{\partial \mathbf{w}^H \mathbf{X} D^*}{\partial \mathbf{w}^*} = \frac{\partial \mathbf{X}^T \mathbf{w}^* D^*}{\partial \mathbf{w}^*} = \mathbf{X} D^*$$

$$\frac{\partial \mathbf{w}^H \mathbf{X} \mathbf{X}^H \mathbf{w}}{\partial \mathbf{w}^*} = \frac{\partial \mathbf{w}^T \mathbf{X}^* \mathbf{X}^T \mathbf{w}^*}{\partial \mathbf{w}^*} = \mathbf{X} \mathbf{X}^H \mathbf{w}$$

always using the denominator layout (result is column vector)

$$\frac{\partial(ee^*)}{\partial w^*} = (\mathbf{X} D^* - \mathbf{X} \mathbf{X}^H \mathbf{w})$$

$$E_{DY}[G_\sigma^C(e)(\mathbf{X} D^* - \mathbf{X} \mathbf{X}^H \mathbf{w})] = \mathbf{0}$$

$$E_{DY}[G_\sigma^C(e)\mathbf{X} D^*] = E_{DY}[G_\sigma^C(e)\mathbf{X} \mathbf{X}^H] \mathbf{w}$$

$$\mathbf{w} = \left[\sum_{n=1}^N G_\sigma^C(e_n) \mathbf{X}_n \mathbf{X}_n^H \right]^{-1} \left[\sum_{n=1}^N G_\sigma^C(e_n) D_n^* \mathbf{X}_n \right] \quad (58)$$

A full discussion on how to implement the above equation is presented in (Guimaraes et al., 2017).