Complex Correntropy Applied to a Compressive Sensing Problem in an Impulsive Noise Environment

JOÃO P. F. GUIMARÃES [1,4], **FELIPE B. DA SILVA**[2], **ALUISIO I. R. FONTES** [3], **RICARDO VON BORRIES**[2], **AND ALLAN DE M. MARTINS**[4]

[1]Department of Information, Federal Institute of Rio Grande do Norte, João Câmara 59550-000, Brazil
[2]Department of Electrical and Computer Engineering, The University of Texas at El Paso, El Paso, TX 79968, USA
[3]Department of Information, Federal Institute of Rio Grande do Norte, Pau dos Ferros 59900-000, Brazil
[4]Department of Electrical Engineering, Federal University of Rio Grande do Norte, Natal 59064-741, Brazil

Corresponding author: João P. F. Guimarães (joao.guimaraes@ifrn.edu.br)

ABSTRACT Correntropy is a similarity function capable of extracting high-order statistical information from data. It has been used in different kinds of applications as a cost function to overcome traditional methods in non-Gaussian noise environments. One of the recent applications of correntropy was in the theory of compressive sensing, which takes advantage of sparsity in a transformed domain to reconstruct the signal from a few measurements. Recently, an algorithm called ℓ_0–MCC was introduced. It applies the Maximum Correntropy Criterion (MCC) in order to deal with a non-Gaussian noise environment in a compressive sensing problem. However, because correntropy was only defined for real-valued data, it was not possible to apply the ℓ_0–MCC algorithm in a straightforward way to compressive sensing problems dealing with complex-valued measurements. This paper presents a generalization of the ℓ_0–MCC algorithm to complex-valued measurements. Simulations show that the proposed algorithm can outperform traditional minimization algorithms such as Nesterov's algorithm (NESTA) and the ℓ_0–Least Mean Square (ℓ_0–LMS) in the presence of non-Gaussian noise.

INDEX TERMS Complex correntropy, complex-valued data, compressive sensing, ℓ_0–approximation.

I. INTRODUCTION

In the last years, a great interest has grown in the scientific community around an area called compressive sensing (CS), which takes advantage from the sparsity of signals to reconstruct signals by using fewer measurements than Nyquist-Shannon's sampling theory would predict [1]. Due to this characteristic, CS has been successfully applied to many practical problems e.g. seismic prospecting [2], holographic microwave imaging [3], and health monitoring [4]–[6].

Compressive sensing techniques have been proved able to reconstruct sparse vectors even under certain amounts of noise [7]–[10]. Impulsive noise tends to deteriorate

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan .

the performance of traditional algorithms based on the second-order constraints [1], [11]–[14].

Correntropy is a similarity measure that is capable of extracting high-order statistical information from data [15]. As a nonlinear similarity measure, correntropy has been successfully used as an efficient optimization cost function in signal processing and machine learning, being specially robust in the presence of non-Gaussian noise and successful in different applications such as cognitive radio [16]–[19], adaptive filtering [20]–[22], principal component analysis (PCA) [23], deep learning [24], [25], and state estimation [26].

Correntropy was also used in CS analysis in [27] showing a great potential for nonlinear signal processing and defined a new algorithm called ℓ_0–Maximum Correntropy Criterion (ℓ_0–MCC). It puts together an ℓ_0 gradient approxi-

mation strategy, previously used in the ℓ_0–Least Mean Square (ℓ_0–LMS) [28] and the Maximum Correntropy Criterion (MCC). As the complex-valued data are widely employed in many signal processing applications including compressive sensing problems [29]–[33] and since correntropy is only defined for real-valued random variables, it is not possible to recover a complex-valued sparse vector or deal with complex-valued measurements using correntropy in a straightforward way. Recently, the complex correntropy has been defined [34], [35]. It extends the robustness of correntropy to complex-valued random variables.

This paper introduces a method called ℓ_0–MCCC, which uses Maximum Complex Correntropy Criterion (MCCC) in order to solve compressive sensing problems in the presence of non-Gaussian noise. The proposed algorithm is derived by using Wirtinger Calculus and generalizes both algorithms ℓ_0–MCC and ℓ_0–LMS. Results show that the proposed method could overcome traditional algorithms such as Nesterov's algorithm (NESTA) [11] and ℓ_0–LMS in a non-Gaussian environment. A kernel size dependence analysis is presented as well as the influence of the number of measurements into the method's performance.

The organization of the paper is as follows. Section II explains the compressive sensing problem, while Section III shows the complex correntropy and highlights important properties. A brief review of Wirtinger Calculus and the proposed algorithm ℓ_0–MCCC are presented in section IV. Section V describes the impulsive noise environment used in the tests. Section VI presents the results to verify the theoretical assumptions associated with the properties of ℓ_0–MCCC, which provide overall improved performance if compared with the classical solutions. Finally, Section VII summarizes the main conclusions and potential future work.

II. COMPRESSIVE SENSING

The theory of compressive sensing is based on the reconstruction of an S-sparse long-length signal $\mathbf{x}$ from a short-length signal $\mathbf{y}$, which is obtained as linear measurements mapped by a wide matrix $\mathbf{A}$ of dimension $M \times L$, $M < L$. Mathematically, this can be expressed by the underdetermined linear system $\mathbf{y} = \mathbf{Ax}$. However, in compressive sensing, it is required that the sparsity level $S = \|\mathbf{x}\|_0$ of the signal $\mathbf{x}$ be very small compared to the signal length, therefore $S \ll L$. As the linear system in a compressive sensing problem is underdetermined, it has no solution or infinite many solutions. By assuming the system has at least one solution, one can look for the sparsest solution and then find a unique solution under a certain tolerance. The unique solution can be found if the matrix $\mathbf{A}$ holds the restricted isometry property (RIP) [9], which is a high probability sufficient condition for reconstruction of sparse signals.

In most practical applications, the noise has to be taken into account. Therefore, the linear system that models the compressive sensing problem is slightly modified to $\mathbf{b} = \mathbf{y} + \gamma$, where γ is the additive noise. In addition, if the signal $\mathbf{x}$ is not sparse, a sparsifying bijective transformation $\mathbf{T}$ can be applied

so that $\mathbf{Tx} = \hat{\mathbf{x}}$ is sparse. Hence the problem can be written as $\mathbf{b} = \mathbf{AT}^{-1}\hat{\mathbf{x}} + \gamma$. Without loss of generality, someone can write $\mathbf{b} = \mathbf{A}\hat{\mathbf{x}} + \gamma$ just by redefining the sensing matrix $\mathbf{A}$.

In order to solve the system $\mathbf{b} = \mathbf{Ax} + \gamma$, [36] highlights three main approaches in the literature: Greedy, Bayesian algorithms, and convex optimization based. Greedy algorithms such as the iterative hard-thresholding (IHT) [37] have a lower computational cost but its performance is highly affected by noise. Reference [38] proposes incorporate prior information to the problem in order to deal with this. Bayesian algorithms such as the sparse Bayesian learning (SBL) [39] also have high computational complexity but the approximate message passing (AMP) algorithm [40] avoids a matrix inversion in order to solve this. This kind of algorithm has been recently improved in [36] and in [41]. Reference [42] proposes a unified Bayesian inference framework for generalized linear model (GLM), and [28] solves the sparse signal reconstruction problem employing a stochastic gradient-based adaptive filtering framework, which is commonly used in system identification problems.

Regarding the convex optimization approach, the algorithms usually implement either an approximated solution of the minimization of $\|\mathbf{x}\|_0$, given some constraints, or an approximation of $\|\mathbf{x}\|_0$ itself by means of relaxation strategies like basis pursuit [11] and correntropy induced metric [43]. For instance, in a quadratically constrained problem, the compressive sensing problem can be formulated as

$$\underset{\mathbf{x}}{\text{minimize}} \; \|\mathbf{x}\|_0$$
$$\text{subject to} \; \|\mathbf{Ax} - \mathbf{b}\|_2^2 \leq \epsilon, \qquad (1)$$

where ϵ is a tolerance factor.

The restricted isometric property measures how close $\|\mathbf{Ax}\|_2$ is to $\|\mathbf{x}\|_2$. We say that the matrix $\mathbf{A}$ holds the restricted isometry property if there exists $\delta_S \in (0, 1)$ such that for all S-sparse signals $\mathbf{x}$ the condition below holds

$$(1 - \delta_S)\|\mathbf{x}\|_2^2 \leq \|\mathbf{Ax}\|_2^2 \leq (1 + \delta_S)\|\mathbf{x}\|_2^2. \qquad (2)$$

Some matrices are known to hold the RIP like submatrices of the following: discrete Fourier transform, discrete cosine transform and random matrices. Those matrices hold the RIP with high probability, which prevents the system to be solved from being ill-posed. In this paper, the sensing matrix $\mathbf{A}$ is a submatrix of a Gaussian random matrix.

III. COMPLEX CORRENTROPY
Since this paper deals with complex-valued data, it is necessary to use the complex correntropy, which is defined in [35]:

$$V_\sigma^c(Q, B) = E[K_\sigma(Q, B)], \qquad (3)$$

where $K_\sigma(.)$ is any positive-definite kernel with size σ; $E[.]$ is the expected value operator; *Q and B* are complex random variables. Some important properties of complex correntropy are presented next.

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

Property 2: Complex correntropy is positive and bounded. For the complex Gaussian kernel shown in (4), its estimated value $\hat{V}_\sigma^c$ is a non-negative real number and between zero, the minimum value, and $1/2\pi\sigma^2$, the value corresponding when $Q = B$.

$$G_\sigma^c(x - y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-y)(x-y)^*}{2\sigma^2}\right), \quad (4)$$

where x, y are complex-valued scalars and $[.]^*$ denotes the complex-conjugate operator.

Property 3: For the Gaussian kernel, the complex correntropy is a weighted sum of all the even moments of the random variable $Q - B$. Furthermore, as the kernel size σ increases, the correntropy tends to the correlation between Q and B.

For a complete overview and mathematical proofs of complex correntropy properties, see [35].

One can estimate complex correntropy using the complex Gaussian kernel as [34]

$$\hat{V}_\sigma^c(Q, B) = \frac{1}{2\pi\sigma^2} \frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{(q_i - b_i)(q_i - b_i)^*}{2\sigma^2}\right), \quad (5)$$

where $Q = \{q_i\}_{i=1}^N$ and $B = \{b_i\}_{i=1}^N$ are samples from the random variables Q and B, respectively. For now on, this papers makes use of an abuse of notation calling $V_\sigma^c(\mathbf{q}, \mathbf{b})$ instead of $V_\sigma^c(Q, B)$ when referring to the estimation of the complex correntropy, where $\mathbf{b}, \mathbf{q} \in \mathbb{C}^{N \times 1}$ are vectors with N samples of the measured random variables Q and B.

IV. ℓ_0-MAXIMUM COMPLEX CORRENTROPY CRITERION

The ℓ_0–MCC algorithm was first introduced in [44]. It uses the MCC to filter outliers in a compressive sensing problem, but although the good performance in the presence of impulsive noise, its application is restricted for real-valued data only. Since many compressive sensing problems deal with complex-valued data, instead of doing the minimization showed in (1), let us define a new cost function J inspired in [44] as

$$J = \min_{\mathbf{w}} \left\{ -V_\sigma^c(\mathbf{b}, \mathbf{y}) + \lambda \|\mathbf{w}\|_0 \right\}$$
$$= \min_{\mathbf{w}} \left\{ -V_\sigma^c(\mathbf{b}, \mathbf{Aw}) + \lambda \|\mathbf{w}\|_0 \right\}, \quad (6)$$

where $V_\sigma^c(\mathbf{b}, \mathbf{y})$ is the complex correntropy between the desired signal vector $\mathbf{b}$, which is the noisy measurements, while the estimated output vector $\mathbf{y} = \mathbf{Aw}$. Recall that $\mathbf{A} \in \mathbb{C}^{M \times L}$ is the measurement matrix and $\mathbf{w} \in \mathbb{C}^L$ is the estimated sparse vector, and λ is a simple regularization parameter that weighs the ℓ_0–approximation of $\mathbf{w}$.

By minimizing (6), one will maximize the complex correntropy $V_\sigma^c(\mathbf{b}, \mathbf{y})$ by taking into account $\lambda \|\mathbf{w}\|_0$, which will ensure the sparsest solution [45].

This paper obtains a stochastic gradient solution in order to minimize (6). This is achieved by calculating the derivative of $V_\sigma^c(\mathbf{b}, \mathbf{y})$ with respect to $\mathbf{w}^*$ combined with the ℓ_0–approximation $\nabla \|\mathbf{w}\|_0$, which are both detailed in the follow Sections IV-A and IV-B.

A. MAXIMUM COMPLEX CORRENTROPY CRITERION

In order to deal with outliers in complex-valued measurements, let us use the MCCC to maximize the similarity between the measurements $\mathbf{b}$ and the estimated output $\mathbf{y} = \mathbf{Aw}$. Then:

$$V_\sigma^c(\mathbf{b}, \mathbf{y}) = V_\sigma^c(\mathbf{b}, \mathbf{Aw}) = \frac{1}{2\pi\sigma^2} \frac{1}{M} \sum_{i=1}^M \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right), \quad (7)$$

where $e_i = (b_i - y_i) = (b_i - \phi_i \mathbf{w})$ is a scalar at the position i from the error vector $\mathbf{e}$; $\phi_i = \mathbf{A}(i, .)$ is a vector with size $1 \times L$ extracted from the i–*th* row of the sensing matrix $\mathbf{A}$; σ is the kernel size, a free parameter from the complex correntropy.

It is important to say that, as stated in Property 2, while depending on complex-valued vectors $(\mathbf{b}, \mathbf{y})$, (7) is always a real-valued function. Then, the Cauchy-Riemann conditions are violated [46], making (6) not analytical in the complex domain. Hence, standard differentiation can not be applied to obtain a gradient function. To address this problem the Wirtinger Calculus, which is based on the duality between spaces $\mathbb{C}$ and $\mathbb{R}^2$, is used [35]. Let $f : \mathbb{C} \to \mathbb{C}$ be a complex function defined in $\mathbb{C}$. Such function can also be defined in $\mathbb{R}^2$ (i.e., $f(x+jy) = f(x, y)$). The Conjugate Wirtinger derivative of f at a point c is defined as follows [47]

$$\frac{\partial f}{\partial z^*}(c) = \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) + j\frac{\partial f}{\partial y}(c)\right). \quad (8)$$

The work [47] presents more details on the Wirtinger Calculus properties. Briefly, in order to compute the Conjugate 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. For example, considering f as $f(z) = zz^*$ implies that

$$\frac{\partial f}{\partial z} = z^* \quad \text{and} \quad \frac{\partial f}{\partial z^*} = z. \quad (9)$$

Then, using this concept to achieve the derivative from (7):

$$\frac{\partial V_\sigma^c(\mathbf{b}, \mathbf{y})}{\partial \mathbf{w}^*} = \frac{1}{2\pi\sigma^2} \frac{1}{M} \sum_{i=1}^M \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) \frac{(-1)}{2\sigma^2} \frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*}, \quad (10)$$

where

$$e_i e_i^* = (b_i - \phi_i \mathbf{w})(b_i - \phi_i \mathbf{w})^*$$
$$= (b_i - \phi_i \mathbf{w})(b_i^* - \phi_i^* \mathbf{w}^*)$$
$$= b_i b_i^* - b_i \phi_i^* \mathbf{w}^* - b_i^* \phi_i \mathbf{w}$$
$$+ \phi_i \mathbf{w} \phi_i^* \mathbf{w}^*. \quad (11)$$

Then, using Wirtinger Calculus, one could say that

$$\frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} = 0 - (b_i \phi_i^*)^T - 0 + (\phi_i \mathbf{w} \phi_i^*)^T, \quad (12)$$

which could be rewritten as

$$\frac{\partial(e_i e_i^*)}{\partial \mathbf{w}^*} = -b_i \phi_i^H + (\phi_i \mathbf{w})\phi_i^H = -e_i \phi_i^H, \quad (13)$$

where $[.]^H = ([.]^T)^*$ is the Hermitian operator. Thus, using (13) in (10), one could obtain

$$\frac{\partial V_\sigma^c}{\partial \mathbf{w}^*} = \frac{1}{4\pi\sigma^4} \frac{1}{M} \sum_{i=1}^{M} \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) e_i \phi_i^H. \quad (14)$$

B. ℓ_0–APPROXIMATION

This paper uses the same ℓ_0–approximation gradient from the paper [44], which defines the ℓ_0–MCC algorithm, but updates the strategy for complex-valued data. Let w_k be the k–th element of the sparse vector $\mathbf{w}$, which can be rewritten as $w_k = w_k^{re} + jw_k^{im}$, where j is the imaginary unit. One can say that

$$\nabla\|w_k\|_0 \approx Z_\beta(w_k) = z_\beta(w_k^{re}) + jz_\beta(w_k^{im}), \quad (15)$$

where

$$z_\beta(w) = \begin{cases} \beta^2 w + \beta, & -\frac{1}{\beta} \leq w < 0; \\ \beta^2 w - \beta, & 0 < w \leq \frac{1}{\beta}. \end{cases} \quad (16)$$

The $z_\beta(w)$ is called zero attraction term and β is a free parameter that controls the attraction region for small coefficients within the interval $[-\frac{1}{\beta}, \frac{1}{\beta}]$. The product $\beta^2 w \pm \beta$ controls how close from zero the non-zero elements can get, in other words, the algorithm precision.

C. ℓ_0–MCCC ALGORITHM

This paper introduces the ℓ_0–MCCC algorithm, which is capable of solving compressive sensing problems for complex-valued signals using the MCCC as the restriction. To achieve that, the goal is to solve (6). Since the derivative of V_σ^c with respect to $\mathbf{w}$ is shown in (14), one could use stochastic gradient solution to reconstruct complex-valued signals with the following update rule:

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \eta \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) e_i \phi_i^H, \quad (17)$$

and then apply the $\nabla\|\mathbf{w}\|_0$ obtained in (15)

$$\mathbf{w}_{i+1} = \mathbf{w}_{i+1} + \eta\lambda Z_\beta(\mathbf{w}_{i+1}), \quad (18)$$

The algorithm 1 summarizes the procedure.

As in the real-valued case [44], the number of data may be insufficient to ensure the convergence. Thus, line 3 from the algorithm makes $\phi_i = \phi_{i+M}$ and $b_{i+M} = b_i$.

Complex correntropy generalizes the real-valued correntropy [35] used to define the ℓ_0–MCC algorithm. Also, as property 3 highlights, by making the kernel size $\sigma \to \infty$, complex correntropy tends to correlation. Then, one can say that the ℓ_0-MCCC algorithm generalizes both ℓ_0–MCC and ℓ_0–LMS to work with complex-valued data.

V. NOISE SIMULATION

In order to evaluate the performance of the proposed algorithm in a noisy environment, this paper uses the Lévy alpha-stable distribution, also called stable distribution, to simulate either impulsive or Gaussian noise.

Algorithm 1 ℓ_0–MCCC

Initialization:
Choose step size η, Gaussian kernel width σ, regularization parameter λ, and zero attraction parameter β. Initial iteration counter $i = 0$ and sparse vector $\mathbf{w}_0 = \mathbf{0}$. Set error tolerance ϵ and maximum iteration number C.
Computation:
1: **procedure** ℓ_0–MCC:
2:     **while** $(i < C)$ **do**
3:       % extract the input vector and corresponding output vector from sensing matrix A and the noisy measurements b:
4:       $r = mod(i, M) + 1; \phi_i = \mathbf{A}(r, .); d_i = b_r$
5:       % computing the output
6:       $y_i = \phi_i w_i$
7:       % computing the error
8:       $e_i = d_i - y_i$
9:       % updating $\mathbf{w}$ using MCCC gradient step
10:      $\mathbf{w}_{i+1} = \mathbf{w}_i + \eta \exp\left(-\frac{e_i e_i^*}{2\sigma^2}\right) e_i \phi_i^H$
11:      % updating $\mathbf{w}$ using the zero attraction strategy
12:      $k = 1; \bar{\mathbf{w}} = \mathbf{w}_{i+1};$
13:      **for** $(k \leq L)$ **do**
14:        $\bar{w}_k = Z_\beta(\bar{w}_k)$
15:        $k = k + 1$
16:      **end for**
17:      $\mathbf{w}_{i+1} = \mathbf{w}_{i+1} + \eta\lambda\bar{\mathbf{w}}$
18:      **if** $\|\mathbf{w}_{i+1} - \mathbf{w}_i\|_2^2 < \epsilon$ **then**
19:        **break**
20:      **end if**
21:      % update iteration number
22:      $i = i + 1$
23:     **end while**
24: **end procedure**

The stable distribution is characterized by four parameters: the index of stability ($0 < \alpha \leq 2$), the skewness parameter ($-1 \leq \beta \leq 1$), the scale parameter ($\sigma \geq 0$), and a shift parameter ($\mu \in \mathbb{R}$) [48]. A common way to introduce stable random variables is to define their characteristic function, which is given by

$$E[\exp(j\theta X)]$$
$$= \exp\left\{-\sigma^\alpha|\theta|^\alpha\left(1 - j\beta \operatorname{sign}(\theta)\tan\frac{\pi\alpha}{2}\right) + j\mu\theta\right\}, \quad (19)$$

if $\alpha \neq 1$, and

$$E[\exp(j\theta X)]$$
$$= \exp\left\{-\sigma|\theta|\left(1 - \frac{2j\beta}{\pi}\operatorname{sign}(\theta)\ln|\theta|\right) + j\mu\theta\right\}, \quad (20)$$

if $\alpha = 1$. The sign function is defined as

$$\operatorname{sign}(\theta) = \begin{cases} -1 & \theta < 1, \\ 0 & \theta = 0, \\ 1 & \theta > 1. \end{cases} \quad (21)$$

TABLE 1. Average SER in dB of 10^3 computer simulations for each one of the three algorithms and three different noise scenarios.

	Noiseless	Gaussian Noise	Impulsive Noise
NESTA	59.21	17.07	7.25
ℓ_0–**LMS**	34.39	23.47	9.51
ℓ_0–**MCCC**	34.28	25.16	24.07

This paper works with symmetric alpha-stable distribution, then β and μ are set to 0 in all simulations. The index of stability α is responsible for controlling the impulsive of the density function i.e. the smaller the value of α, the heavier the distribution tail will be. Particularly for $\beta, \mu = 0$, and $\alpha = 2$, the stable distribution is equivalent to a Gaussian distribution [48].

The Generalized Signal-to-Noise Ratio in dB (GSNR) [49], defined as

$$\text{GSNR} = 10 \log_{10} \frac{P_S}{\sigma^\alpha}, \qquad (22)$$

is used to determine the scale parameter σ, where P_S is the power of the noiseless signal, and α is the index of stability. For given values of GSNR and α, we can determine the scale parameter σ using (22).

VI. RESULTS AND DISCUSSION

In this section, we evaluate the performance of the proposed method ℓ_0–MCCC and compare it with NESTA and ℓ_0–LMS. The Signal-to-Error Ratio in dB (SER) is used throughout this section to evaluate the reconstructed sparse vector when compared with the original sparse vector. It is calculated as

$$\text{SER} = 10 \log_{10} \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{e}\|_2^2}, \qquad (23)$$

where $\|.\|$ is the ℓ_2–norm operator, $\mathbf{x}$ is the noise-free original sparse vector, and $\mathbf{e} = \mathbf{x} - \mathbf{w}$ is the error between the original sparse vector and the reconstructed vector $\mathbf{w}$.

First, three scenarios are considered in the comparisons: noiseless, Gaussian noise, and impulsive noise. The stable distribution, as described by Section V, was used to simulate both the Gaussian and the impulsive noise distributions. Table 1 shows the average from 10^3 SER values for each scenario and the corresponding algorithm. All the simulations were set up by using the same configuration: $M = 512$ measurements, a randomized sparse vector $\mathbf{x}$ with size $L = 1024$ and $S = 32$ non-zero elements. We tested NESTA with a learning rate $\mu = 0.001$, and ℓ_0–LMS and ℓ_0-MCCC algorithms with the following parameters: learning rate $\mu = 0.0005$, $\lambda = 0.07$, $\beta = 0.85$. In addition, we used the ℓ_0–MCCC algorithm with a kernel size set to $\sigma = 1.83$, tolerance $\epsilon = 10^{-6}$, and a maximum number of iterations $C = 2.10^5$.

In the noiseless scenario, all algorithms are able to reconstruct the sparse vector. In fact, accuracy is one of the strengths of NESTA [11], which is the method that achieves the highest SER level in the noiseless environment, as can be seen in Table 1. Because ℓ_0–LMS and ℓ_0–MCCC use the same strategy to approximate the ℓ_0 gradient, they achieve

similar SER levels that are numerically limited by the relation $\beta^2 w \pm \beta$ that composes (16).

The Gaussian noise is generated using a stable distribution with parameters GSNR = 20 dB and $\alpha = 2$. As Table 1 shows, NESTA performance drops significantly when compared with the other two algorithms. The attractor used by the ℓ_0 gradient strategy from the ℓ_0–LMS and ℓ_0–MCCC algorithms tends to reduce the effect from the noise. In this scenario, the proposed ℓ_0–MCCC algorithm has equivalent performance of the literature ℓ_0–LMS.

To evaluate the performance of the algorithms to impulsive noise, the parameter α from the stable distribution was changed from 2 to 1.5, while keeping the GSNR level in 20 dB. Using this parameter configuration, we added outliers to the measurements and reconstructed the original signal using each one of the three algorithms. A typical result with impulsive noise is shown in Fig. 1. The top row shows an example with NESTA, the middle row with the ℓ_0–LMS, and the bottom row with the ℓ_0–MCCC. The left-hand column of Fig. 1 corresponds to the real part of the signals and the right-hand column to the imaginary part. As expected, the presence of outliers deteriorated the performance of the two algorithms based on second-order constraints, NESTA and the ℓ_0–LMS algorithms. In contrast, the method based on complex correntropy was able to better approximate the original complex-valued vector and achieved a higher SER. The average SER of 10^3 simulations for each algorithm is summarized in Table 1.

In order to evaluate the robustness of the ℓ_0–MCCC algorithm to impulsive noise, we ran 10^3 Monte Carlo trials with GSNR in the interval [10, 30] dB and fixed $\alpha = 1.5$. The results plotted in Fig. 2 correspond to a signal length $L = 1024$, two different sparsity levels $S = \{16, 32\}$, and a number of measurements $M = 512$. As we can see from Fig. 2, the ℓ_0–MCCC algorithm is the one that achieves the highest SER levels. As expected, independently of the algorithm tested, by increasing the number of non-zero elements S in the sparse vector, the reconstruction error increases.

An important parameter in the compressive sensing problem is the number of measurements taken. Simulations were made varying the number of measurements M from 64 to 512 in an alpha-stable environment using GSNR = 25 dB and $\alpha = 1.5$. Fig. 3 shows the average SER of 10^3 simulations for each algorithm and sparsity levels $S = \{16, 32\}$. As we can see from Fig. 3, NESTA shows relatively good performance with a small number of measurements, but as the number of measurements taken increases, its performance in the presence of impulsive noise degrades fast. On the other hand, both the ℓ_0–LMS and the ℓ_0–MCCC algorithms show, in general, higher SER levels as the number of measurements increases. However, since the ℓ_0–MCCC algorithm uses the correntropy as a cost function, it can extract more information from data than ℓ_0–LMS, achieving the higher SER levels.

Fig. 4 shows the performance of the ℓ_0–MCCC algorithm as a function of both α and GSNR parameters
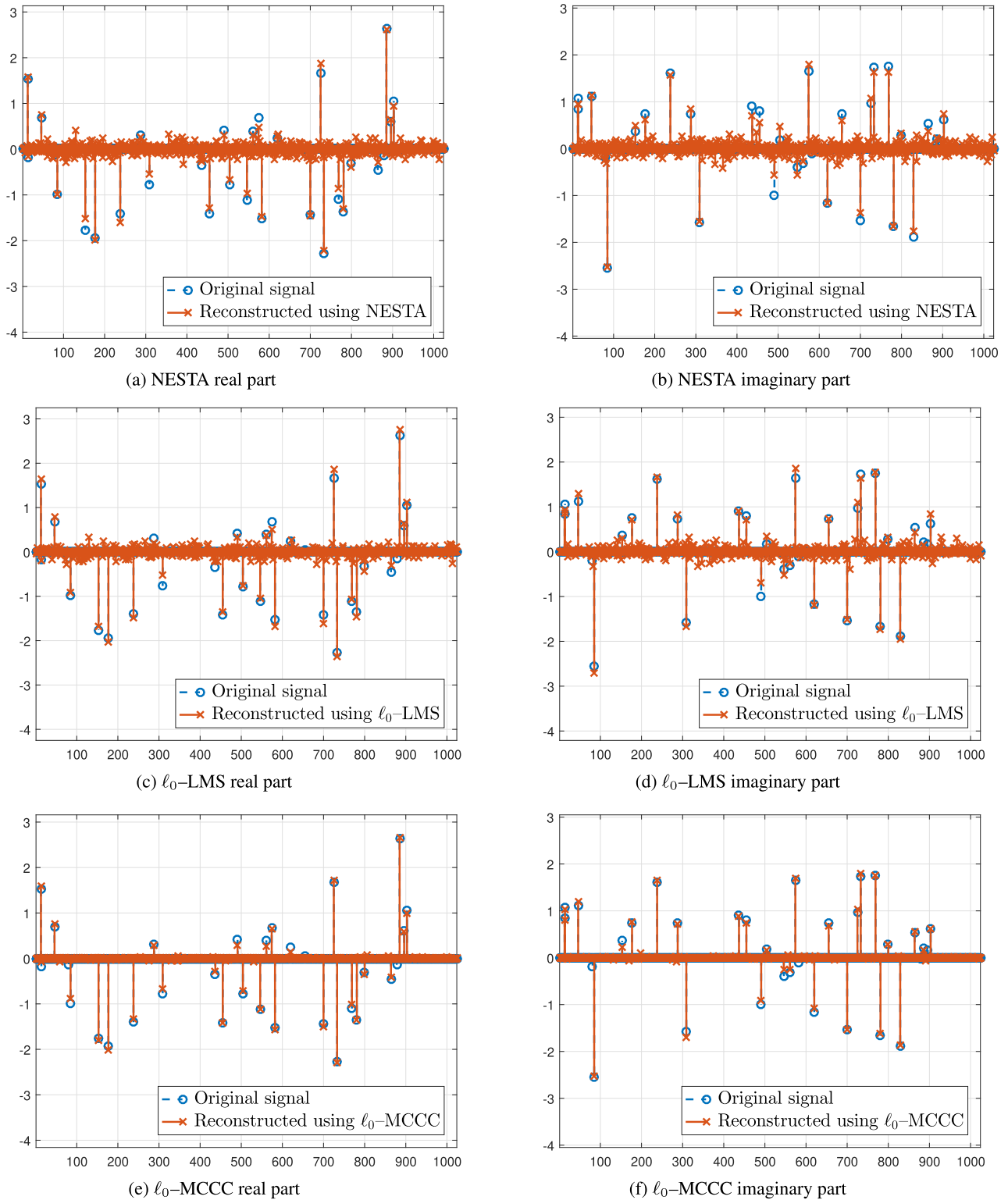
FIGURE 1. Typical reconstruction results for a sparse signal with sparsity $S = 32$, length $L = 1024$ and $M = 512$ measurements contaminated with outliers ($\alpha = 1.5$, GNSR = 20 dB). (a-b) NESTA achieved an SER of 7.74 dB; (c-d) ℓ_0–LMS achieved an SER of 9.73 dB; and (e-f) ℓ_0–MCCC achieved an SER of 27.70 dB. The ℓ_0–MCCC is able to reject the outliers achieving a better performance than ℓ_0–LMS and NESTA, which are based on second-order constraints.

from the alpha-stable distribution. The figure shows that the ℓ_0–MCCC performance decreases, as the noise power increases.

The kernel size plays an important role in the ℓ_0–MCCC algorithm. It is a free parameter of the correntropy function that controls the convergence rate, robustness, and
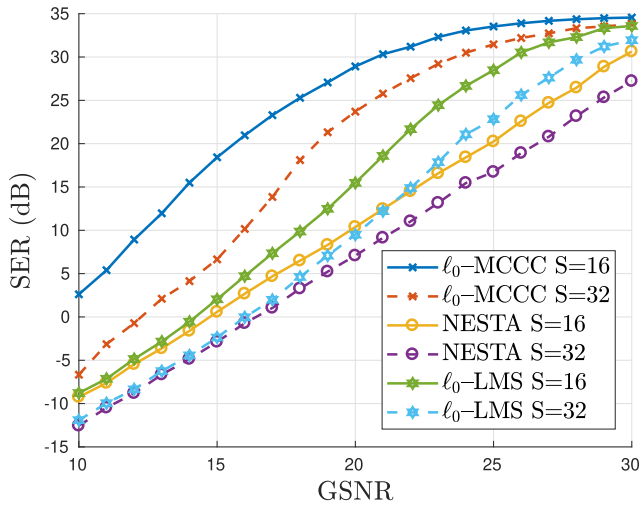
FIGURE 2. Average SER of 10^3 simulations showing the algorithms' performance in reconstructing complex-valued sparse vectors over multiple values of GSNR and fixed $\alpha = 1.5$. S is the number of non-zero elements of the sparse vector. Note that the higher the GSNR, the lower the noise power.
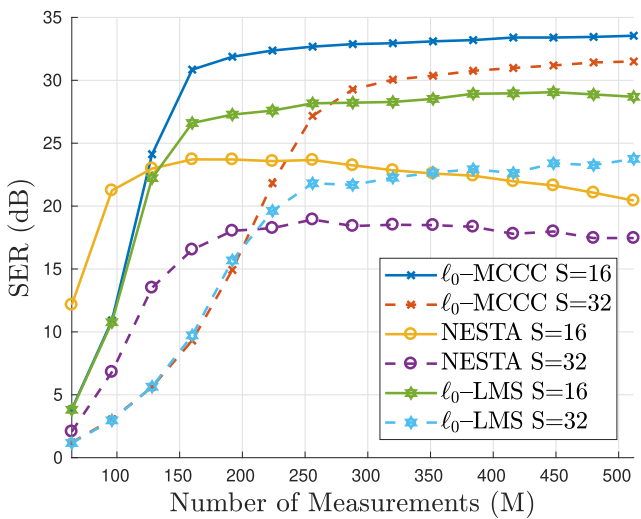

FIGURE 3. Average SER of 10^3 simulations showing the algorithms' performance over multiple numbers of measurements M and fixed parameters GSNR = 25 dB and $\alpha = 1.5$.
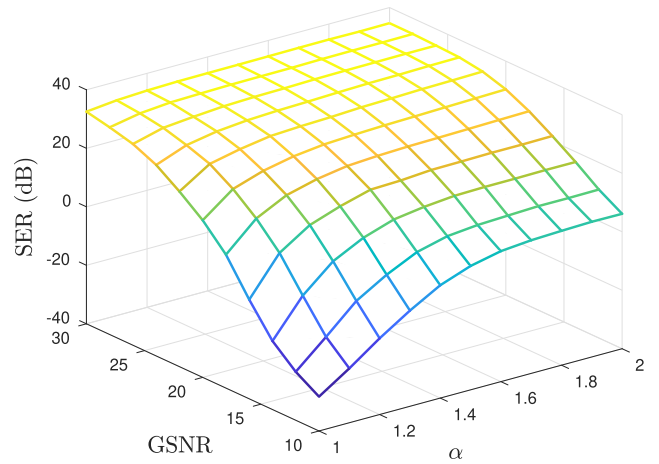

FIGURE 4. Average SER of 10^2 simulations for the ℓ_0-MCCC algorithm as a function of the noise power GSNR and index of stability α.

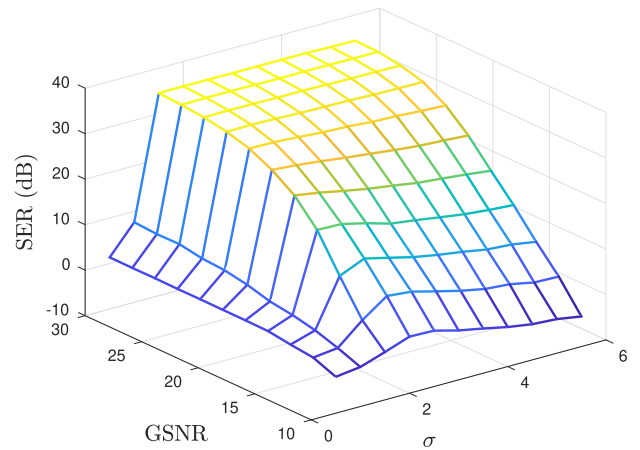**FIGURE 5.** Average of 10^2 simulations for the ℓ_0-MCCC with $\alpha = 1.5$ as a function of the noise power GSNR and kernel size σ.
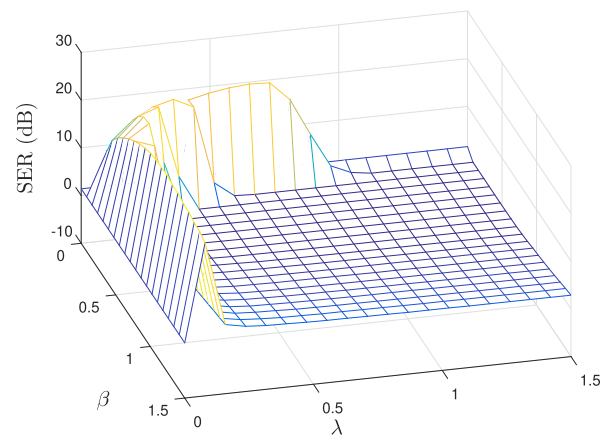

FIGURE 6. ℓ_0-MCCC performance varying the parameters λ, which weighs the ℓ_0 approximation importance, and β, which controls how strong is the zero attraction factor. The plot is the average from 10^2 Monte Carlo trials.

steady-state performance. Fig. 5 shows the SER values of the ℓ_0-MCCC algorithm to different values of GSNR and kernel size σ. As can be noticed, the performance is strictly related to the kernel size selection even at the same GSNR level. As discussed in Section IV, for large kernel sizes, ℓ_0-MCCC generalizes the ℓ_0-LMS algorithm. So, it is worth to mention that, choosing a larger kernel size i.e. $\sigma = 6$ makes the SER level of the ℓ_0-MCCC be around 15dB at GSNR 20dB for example, which is similar to the ℓ_0-LMS results from Fig. 2. By properly setting the kernel size, the ℓ_0-MCCC method can achieve higher SER levels than ℓ_0-LMS and NESTA in an impulsive noise environment.

The ℓ_0-MCCC performance as a function of the parameters λ and β is indicated int the plot of Fig. 6. The plot

corresponds to the average of 10^2 simulations. The goal is to reconstruct an $L = 1024$ sparse vector with $S = 32$ non-zero elements using $M = 512$ measurements, which are

contaminated with an alpha-stable noise of GSNR = 20 dB and $\alpha = 1.5$. The parameter β in (16) controls the attraction region of the ℓ_0 approximation, and the parameter λ controls the ℓ_0 approximation importance as stated in (17). As can be seen in the Fig. 6, there is a small region of λ and β values which makes the ℓ_0–MCCC achieve a high SER level, which indicates how sensitive the algorithm is to these parameters.

In summary, the proposed method ℓ_0–MCCC has the same free parameters as the ℓ_0–MCC: gradient step size μ; the regularization parameter λ; the zero attraction parameter β; and the kernel size σ. As shown in this section, by proper tuning the free parameters, the ℓ_0–MCCC is able to successfully reconstruct complex-valued sparse vectors and impulsive noise, achieving better performance than NESTA and ℓ_0–LMS algorithms. Although the ℓ_0–MCCC computational cost is equivalent to the real-valued version ℓ_0–MCC and ℓ_0–LMS, they are bigger than NESTA. Then, a trade-off between computational time and performance under impulsive noise is added to the problem. This must be taken into account specially when a larger sparse vector is required.

VII. CONCLUSION

This paper investigates the use of the complex correntropy function to generalize both the ℓ_0–LMS and the ℓ_0–MCC algorithms to deal with non-Gaussian noise and complex-valued data in compressive sensing problems. This new method is derived by using Wirtinger Calculus and denoted by ℓ_0–MCCC. In particular, the performance of the ℓ_0–MCCC depends from the selection of the free parameters: the kernel width σ, the regularization parameter λ, and the zero attractor region β, which all of them should be selected according to the application. Simulations explore this dependence and show a comparison of our new proposed algorithm with the previous methods NESTA and ℓ_0–LMS. As in the paper by Yicong He et al. (2019) on real-valued data and impulsive noise, our algorithm based on the complex correntropy constraint outperformed the algorithms based on second-order constraints. Future work aims in investigating alternatives to increase the algorithm performance for reducing both the computational cost and the number of measurements required for reconstruct the sparse vector.

REFERENCES

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[2] H. Sun, F. Yang, F. Meng, Z. Zhang, C. Gao, and M. Liu, "A topographic kirchhoff dynamic focused beam migration method based on compressed sensing," *IEEE Access*, vol. 6, pp. 56666–56674, 2018.

[3] L. Wang and M. Fatemi, "Compressive sensing holographic microwave random array imaging of dielectric inclusion," *IEEE Access*, vol. 6, pp. 56477–56487, 2018.

[4] J. Pagan, R. Fallahzadeh, M. Pedram, J. L. Risco-Martin, J. M. Moya, J. L. Ayala, and H. Ghasemzadeh, "Toward ultra-low-power remote health monitoring: An optimal and adaptive compressed sensing framework for activity recognition," *IEEE Trans. Mobile Comput.*, vol. 18, no. 3, pp. 658–673, Jun. 2018.

[5] Y. Bao, Z. Shi, X. Wang, and H. Li, "Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring," *Struct. Health Monit.*, vol. 17, no. 4, pp. 823–836, 2018. doi: 10.1177/1475921717721457.

[6] M. Jayawardhana, X. Zhu, R. Liyanapathirana, and U. Gunawardana, "Compressive sensing for efficient health monitoring and effective damage detection of structures," *Mech. Syst. Signal Process.*, vol. 84, pp. 414–430, Feb. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327016302503

[7] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[8] J. Zhu and D. Baron, "Performance regions in compressed sensing from noisy measurements," in *Proc. 47th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2013, pp. 1–6.

[9] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comp. Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, May 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1631073X08000964

[10] F. Lin, "Robust compressive reconstruction algorithm for eliminating sampling and measurement impulsive noise," in *Proc. 9th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, vol. 2, Aug. 2017, pp. 89–92.

[11] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1–39, 2011. doi: 10.1137/090756855.

[12] C. Zhang, Y. Guo, F. Wang, and B. Chen, "Generalized maximum correntropy-based echo state network for robust nonlinear system identification," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2018, pp. 1–6.

[13] A. I. R. Fontes, A. de M. Martins, L. F. Q. Silveira, and J. C. Principe, "Performance evaluation of the correntropy coefficient in automatic modulation classification," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 1–8, Jan. 2015.

[14] *Information Theoretic Learning*. New York, NY, USA: Springer, 2010. [Online]. Available: https://books.google.com.br/books?id=DLeSdOqojs8C

[15] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[16] A. I. R. Fontes, J. B. A. Rego, A. M. de Martins, L. F. Q. Silveira, and J. C. Principe, "Cyclostationary correntropy: Definition and applications," *Expert Syst. Appl.*, vol. 69, pp. 110–117, Mar. 2017.

[17] S. Y. Luan, T. Qiu, Y. Zhu, and L. Yu, "Cyclic correntropy and its spectrum in frequency estimation in the presence of impulsive noise," *Signal Process.*, vol. 120, pp. 503–508, Mar. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168415003254

[18] T. Liu, T. Qiu, and S. Luan, "Cyclic correntropy: Foundations and theories," *IEEE Access*, vol. 6, pp. 34659–34669, 2018.

[19] A. I. R. Fontes, P. T. V. Souza, A. D. Neto, A. D. M. Martins, and L. F. Q. Silveira, "Classification system of pathological voices using correntropy," *Math. Problems Eng.*, vol. 42, no. 1, pp. 1–8, 2014.

[20] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 880–884, Jul. 2014.

[21] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.

[22] Z. Wu, S. Peng, B. Chen, and H. Zhao, "Robust Hammerstein adaptive filtering under maximum correntropy criterion," *Entropy*, vol. 17, no. 10, pp. 7149–7166, 2015.

[23] R. He, B.-G. Hu, W.-S. Zheng, and X.-W. Kong, "Robust principal component analysis based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1485–1494, Jun. 2011.

[24] L. Chen, H. Qu, J. Zhao, B. Chen, and J. C. Principe, "Efficient and robust deep learning with correntropy-induced loss function," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1019–1031, May 2016.

[25] W. Shi, Y. Gong, X. Tao, and N. Zheng, "Training DCNN by combining max-margin, max-correlation objectives, and correntropy loss for multilabel image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2896–2908, Jul. 2018.

[26] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.

[27] Y. He, F. Wang, S. Wang, J. Cao, and B. Chen, "Maximum correntropy adaptation approach for robust compressive sensing reconstruction," *Inf. Sci.*, vol. 480, pp. 381–402, Apr. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025518309940

[28] J. Jin, Y. Gu, and S. Mei, "A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework," IEEE J. Sel. Topics Signal Process., vol. 4, no. 2, pp. 409–420, Apr. 2010.

[29] X. X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," IEEE Trans. Geosci. Remote Sens., vol. 50, no. 1, pp. 247–258, Jan. 2012.

[30] C. K. Sung, F. de Hoog, Z. Chen, P. Cheng, and D. C. Popescu, "Interference mitigation based on Bayesian compressive sensing for wireless localization systems in unlicenced band," IEEE Trans. Veh. Technol., vol. 66, no. 8, pp. 7038–7049, Aug. 2017.

[31] S. Bettens, C. Schretter, N. Deligiannis, and P. Schelkens, "Bounds and conditions for compressive digital holography using wavelet sparsifying bases," IEEE Trans. Comput. Imag., vol. 3, no. 4, pp. 592–604, Dec. 2017.

[32] F. H. C. Tivive and A. Bouzerdoum, "A compressed sensing method for complex-valued signals with application to through-the-wall radar imaging," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2013, pp. 2144–2148.

[33] X. X. Zhu and R. Bamler, "A fundamental bound for super-resolution—With application to 3D SAR imaging," in Proc. Joint Urban Remote Sens. Event, Apr. 2011, pp. 181–184.

[34] J. P. F. Guimarães, A. I. R. Fontes, J. B. A. Rego, A. M. de Martins, and J. C. Principe, "Complex correntropy: Probabilistic interpretation and application to complex-valued data," IEEE Signal Process. Lett., vol. 24, no. 1, pp. 42–45, Jan. 2017.

[35] J. P. F. Guimarães, A. I. R. Fontes, J. B. A. Rego, A. D. M. Martins, and J. C. Principe, "Complex correntropy function: Properties, and application to a channel equalization problem," Expert Syst. Appl., vol. 107, pp. 173–181, Oct. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417418302501

[36] J. Zhu, L. Han, and X. Meng, "An AMP-based low complexity generalized sparse Bayesian learning algorithm," IEEE Access, vol. 7, pp. 7965–7976, 2019.

[37] T. Blumensath, M. Yaghoobi, and M. E. Davies, "Iterative hard thresholding and l0 regularisation," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), vol. 3, Apr. 2007, pp. III-877–III-880.

[38] R. E. Carrillo and K. E. Barner, "Lorentzian iterative hard thresholding: Robust compressed sensing with prior information," IEEE Trans. Signal Process., vol. 61, no. 19, pp. 4822–4833, Oct. 2013.

[39] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," IEEE Trans. Signal Process., vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[40] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," Proc. Nat. Acad. Sci. USA, vol. 106, no. 45, pp. 18914–18919, 2009.

[41] X. Meng and J. Zhu, "Bilinear adaptive generalized vector approximate message passing," IEEE Access, vol. 7, pp. 4807–4815, 2019.

[42] X. Meng, S. Wu, and J. Zhu, "A unified Bayesian inference framework for generalized linear models," IEEE Signal Process. Lett., vol. 25, no. 3, pp. 398–402, Mar. 2018.

[43] S. Seth and J. C. Principe, "Compressed signal reconstruction using the correntropy induced metric," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Mar. 2008, pp. 3845–3848.

[44] Y. He, F. Wang, S. Wang, J. Cao, and B. Chen, "Maximum correntropy adaptive filtering approach for robust compressive sensing reconstruction," Jun. 2017, arXiv:1706.03226. [Online]. Available: https://arxiv.org/abs/1706.03226

[45] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[46] D. P. Mandic and V. S. L. Goh, Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models. Hoboken, NJ, USA: Wiley, 2009.

[47] P. Bouboulis and S. Theodoridis, "Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS," IEEE Trans. Signal Process., vol. 59, no. 3, pp. 964–978, Mar. 2011.

[48] G. Samorodnitsky and M. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance. New York, NY, USA: Taylor & Francis, 1994. [Online]. Available: https://books.google.com.br/books?id=wTTUfYwjksAC

[49] C. L. Nikias and M. Shao, Signal Processing With Alpha-Stable Distributions and Applications. New York, NY, USA: Wiley, 1995.

JOÃO P. F. GUIMARÃES received the B.Sc. and M.Sc. degrees in computer engineering and electrical engineering from the Federal University of Rio Grande do Norte (UFRN), Brazil, in 2010 and 2012, respectively, where he is currently pursuing the Ph.D. degree. He was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, The University of Texas at El Paso. He is also an Associate Professor with the Federal Institute of Rio Grande do Norte (IFRN). His research interests include digital signal processing and machine learning.

FELIPE B. DA SILVA received the B.Sc. and M.Sc. degrees in mathematics from the University of Brasilia, Brazil, in 2007 and 2010, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with The University of Texas at El Paso. His current research interests include signal and image processing, numerical optimization, compressive sensing, remote sensing, cryptography, big data, numerical linear algebra, group theory, and number theory.

ALUISIO I. R. FONTES received the bachelor's degree in computer engineering and the M.Sc. and Ph.D. degrees in electrical engineering and computing from the Federal University of Rio Grande do Norte (UFRN), Brazil, in 2006, 2012, and 2015, respectively. He is currently a Professor with the Federal Institute of Rio Grande do Norte (IFRN), where he is also a member of the Information Theoretic Learning Research Group. He has experience in the electrical engineering area, with an emphasis on digital signal processing for telecommunication systems, acting mainly in the following subjects: information theoretic, data processing and telecommunication systems and artificial intelligence.

RICARDO VON BORRIES received the B.Sc. degree in electrical engineering from the University of Brasilia, Brazil, in 1982, the M.Sc. degree in electrical engineering from the Federal University of Rio de Janeiro, Brazil, in 1989, and the Ph.D. degree in electrical and computer engineering from Rice University, Houston, TX, USA, in 2005. He is currently an Associate Professor of electrical and computer engineering with The University of Texas at El Paso. His main research interests include digital signal processing, wavelet transforms, Fourier transform, time–frequency analysis, compressive sensing, tomographic imaging, and overcomplete signal representation. He is a member of SIAM and ASEE.

ALLAN DE M. MARTINS received the M.S. and Ph.D. degrees in electrical engineering. He held a postdoctoral position with the Computational Neuro-Enginering Laboratory (CNEL), Florida University, and also with the Geneva Observatory, Switzerland. He is currently an Associate Professor with the Federal University of Rio Grande do Norte (UFRN). He has experience in the areas of digital signal and image processing, machine learning, control systems, and related areas.

• • •