CORRENTROPY-BASED ADAPTIVE FILTERING OF NONCIRCULAR COMPLEX DATA

Bruno Scalzo Dees[1], *Yili Xia*[2], *Scott C. Douglas*[3], *Danilo P. Mandic*[1]

[1] Department of EEE, Imperial College London, London, SW7 2BT, UK
[2] School of Information Science and Engineering, Southeast University, Nanjing 210096, China
[3] Department of EE, Southern Methodist University, Dallas, Texas 75275, US
Emails: {bs1912, d.mandic}@imperial.ac.uk, yili_xia@seu.edu.cn, douglas@lyle.smu.edu

ABSTRACT

Real world complex-valued signals typically exhibit rotation-dependent distributions (noncircularity), and significant performance gains in learning algorithms can be obtained by accounting for information beyond the standard second-order noncircularity (impropriety). To this end, we introduce a new closed form definition of complex correntropy which is general enough to cater for both circular and noncircular distributions in complex data, and serves as a basis for a novel cost function for widely linear adaptive filtering, termed the maximum improper complex correntropy criterion (MICCC). A stochastic gradient adaptive filtering algorithm is developed based on the MICCC, and its standard and complementary convergence and stability analyses are conducted with respect to both the circularity of the estimation error and the kernel size in the underlying Parzen estimator. Performance advantages over the strictly linear correntropy algorithm (MCCC) and the mean square error based complex least mean square (CLMS) and augmented CLMS (ACLMS) are demonstrated through analysis and simulations.

***Index Terms*—** Complex-valued signal processing, complementary mean square analysis, complex correntropy, improperness, maximum improper complex correntropy criterion (MICCC).

1. INTRODUCTION

Standard covariance and entropy-based statistical measures either cannot model higher-order statistics within a time series or/and employ a rigid assumption that the signal of interest is at least second-order stationary. However, in practice, most measured quantities exhibit a degree of non-Gaussianity and nonstationarity, which for some problems can be known in advance. To this end, an extension of the fundamental definition of correlation for random processes was proposed in [1], termed the *correntropy*, to address the problem that most of the conventional information theoretic learning (ITL) measures [2] do not use all the information in the case of temporally correlated (non-white) input signals. Unlike standard correlation, this measure contains higher-order moments of the probability distribution function (pdf), but is much simpler to estimate directly from the samples than conventional moment expansions. The concept was initially introduced for univariate random processes, and was extended to a more general case of two arbitrary random variables in [3].

Recent application studies have validated correntropy as an efficient tool for analyzing higher-order statistical moments in non-Gaussian signals [3, 4, 5]. Especially successful has been its application as a cost function in linear adaptive filters, within the framework called *maximum correntropy* criterion (MCC) [6]. Tools developed based on this concept include the MCC-based variable step-size least mean square (LMS) algorithm [7] and a closed-form fixed-point recursion filter [8, 9].

More recently, the correntropy framework has been extended to complex-valued time series, through *complex correntropy*, and the corresponding *maximum complex correntropy criterion* (MCCC), the utility of which as a cost function for the complex-valued least-mean squares (CLMS) and complex-valued fixed-point recursion filters was demonstrated in [10].

Although complex correntropy and the MCCC cost function have shown significant potential in complex-valued signal processing, there remain several issues that need to be addressed prior to its more widespread application, these include: (i) complex correntropy was derived with the assumption of proper (second-order circular) random variables, however, this is very restrictive as real world signals are typically second-order noncircular (improper); (ii) the existing MCCC cost function is only effective for *strictly linear* (SL) models which are inadequate for the widely linear systems, typical in the real world, and the associated improperness (second-order noncircularity) of data.

To this end, we introduce novel correntropy measures within a widely linear framework [11], and illustrate their effectiveness for robust adaptive filtering of noncircular data. This is achieved based on the augmented complex statistics which employs both the covariance, $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$, and the pseudo-covariance, $\mathbf{P} = E\{\mathbf{x}\mathbf{x}^T\}$, matrices, in order to cater for both circular (rotation-invariant probability distributed) and second-order noncircular (improper) signals with rotation dependent pdfs. The introduced maximum improper correntropy criterion (MICCC) is used as a cost function for a corresponding general widely linear adaptive filter. Its convergence is analysed in terms of the mean and mean-square convergence, and uniquely by employing the complementary convergence analysis to assess the degree of circularity of the output error along the iterations. Illustrative simulations demonstrate the MICCC outperforms the MCCC-CLMS, CLMS and augmented CLMS (ACLMS).

2. MAXIMUM COMPLEX CORRENTROPY CRITERION

The probabilistic interpretation of complex correntropy is based on estimating the probability of the event $\mathbf{x} = \mathbf{y}$, for random complex variables $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$. This is equivalent to considering the joint probability of the events $\Re\{\mathbf{x}\} = \Re\{\mathbf{y}\}$ and $\Im\{\mathbf{x}\} = \Im\{\mathbf{y}\}$ [10]. Using a complex-valued Gaussian pdf, denoted by $\kappa_\sigma(\cdot)$, the calculation of correntropy between variables x and y is then equivalent to estimating the probability of the estimation error, $\mathbf{e} = \mathbf{x} - \mathbf{y}$, that is

$$V_\sigma(\mathbf{x}, \mathbf{y}) = E\{\kappa_\sigma(\mathbf{e})\} = \frac{1}{\pi\sigma^2} E\left\{\exp\left[-\frac{\mathbf{e}^H\mathbf{e}}{\sigma^2}\right]\right\}, \qquad (1)$$

whereby σ^2 is the variance of $\kappa_\sigma(\mathbf{e})$ and $V_\sigma(\mathbf{x}, \mathbf{y})$ denotes an appropriate Parzen estimator [12], which has the form

$$V_\sigma(\mathbf{x}, \mathbf{y}) = \frac{1}{\pi\sigma^2}\frac{1}{N}\sum_{n=1}^{N} \exp\left[-\frac{|(x_n - y_n)|^2}{\sigma^2}\right], \qquad (2)$$

and represents a measure of maximum similarity between the random variables $\mathbf{x} = [x_1, ..., x_N]^T$ and $\mathbf{y} = [y_1, ..., y_N]^T$.

The "proper" MCCC, proposed in [10], assumes the form in (2) and produces the probabilistic difference, $e = d - y$, between the desired signal $d \in \mathbb{C}$ and the filter output $y \in \mathbb{C}$. In other words, for a strictly linear model $y = \mathbf{h}^H\mathbf{x}$, where $\mathbf{x} \in \mathbb{C}^N$ is the complex-valued input and $\mathbf{h} \in \mathbb{C}^N$ a weight vector, the MCCC represents the maximum complex correntropy between the random variables d and y, that is

$$J_{\text{MCCC}} = V_\sigma(d, y) = E\{\kappa_\sigma(e)\}. \tag{3}$$

3. MAXIMUM IMPROPER COMPLEX CORRENTROPY CRITERION

The pdf of a general zero-mean noncircular complex Gaussian distributed random variable, $\mathbf{x} \in \mathbb{C}^N$, is defined as [13, 11, 14]

$$\kappa_{\sigma,\varrho}(\mathbf{x}) = \frac{1}{\pi\sigma^2\sqrt{1-|\varrho|^2}} \exp\left[-\frac{|\mathbf{x}|^2 - \Re\{\varrho\mathbf{x}^{*2}\}}{\sigma^2(1-|\varrho|^2)}\right], \tag{4}$$

where $\varrho = E\{\mathbf{x}^2\}/E\{|\mathbf{x}|^2\}$ is the circularity quotient of $\mathbf{x}$ [15].

To introduce a measure of improper complex correntropy, as an extension of the probabilistic interpretation in [10], consider an improper complex random variable, $\mathbf{e} = \mathbf{x} - \mathbf{y} \in \mathbb{C}^N$, with $\mathbf{x} = [x_1, ..., x_N]^T$ and $\mathbf{y} = [y_1, ..., y_N]^T$. Then, the complex correntropy is estimated through an appropriate Parzen estimator, given by

$$V_{\sigma,\varrho}(\mathbf{x}, \mathbf{y}) = \frac{1}{\pi\sigma^2\sqrt{1-|\varrho|^2}} \frac{1}{N} \sum_{n=1}^{N} \exp\left[-\frac{|e_n|^2 - \Re\{\varrho e_n^{*2}\}}{\sigma^2(1-|\varrho|^2)}\right] \tag{5}$$

where $e_n = x_n - y_n$ and $\varrho = E\{\mathbf{e}^T\mathbf{e}\}/E\{\mathbf{e}^H\mathbf{e}\}$. Well established methods exist for determining the optimal value for the kernel size, σ^2 in (5) [16, 15], which without loss of generality is assumed to be a constant in this work.

A further insight into the improper complex correntropy is conveniently provided through its Taylor series expansion

$$V_{\sigma,\varrho}(\mathbf{x}, \mathbf{y}) = \frac{1}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} E\left\{\frac{\left[\mathbf{e}^H\mathbf{e} - \Re\{\varrho\mathbf{e}^H\mathbf{e}^*\}\right]^n}{\sigma^{2(n+1)}(1-|\varrho|^2)^{n+\frac{1}{2}}}\right\}. \tag{6}$$

Remark 1: Observe that with an increase in the kernel size, σ^2, the higher-order terms in (6) decay faster than the second-order terms, and that, contrary to the case of a proper $\mathbf{x}$, as desired the circularity quotient ϱ is involved too. The larger the circularity coefficient $|\varrho|$, the greater the contribution of the higher-order terms.

Remark 2: The only case where the proposed improper complex correntropy behaves like the covariance is when the kernel size, σ^2, tends to infinity and the circularity quotient, ϱ, vanishes. In this way, the involvement of the circularity quotient within the higher-order terms overcomes the undermodeling problem of the proper correntropy model in [10] when applied to noncircular data.

To support the development of correntropy-based adaptive signal processing algorithms for noncircular data, we next introduce the *maximum improper complex correntropy criterion* (MICCC) which is based on (5), accounts for the complex impropriety, and can be used for both circular and noncircular inputs.

4. ROBUST WIDELY LINEAR FILTERING

Consider a widely linear (WL) model in the form

$$y = \mathbf{h}^H\mathbf{x} + \mathbf{g}^H\mathbf{x}^* = \underline{\mathbf{w}}^H\underline{\mathbf{x}}, \tag{7}$$

where $\underline{\mathbf{x}} = \left[\mathbf{x}^T, \mathbf{x}^H\right]^T$ and $\underline{\mathbf{w}} = \left[\mathbf{h}^T, \mathbf{g}^T\right]^T$ are respectively the augmented input and coefficient vectors, with $\mathbf{x}, \mathbf{h}, \mathbf{g} \in \mathbb{C}^N$ [17]. Define the estimation error, $e = d - y$, as the difference between the desired signal $d \in \mathbb{C}$ and the filter output $y \in \mathbb{C}$. The new cost function is then defined as the maximum improper complex correntropy between the random variables d and y, and is given by

$$J_{\text{MICCC}} = V_{\sigma,\varrho}(d, y) = E\{\kappa_{\sigma,\varrho}(e)\}. \tag{8}$$

4.1. MICCC-based stochastic gradient adaptive filter

Following on the work in [6], we now derive a gradient-based adaptive learning algorithm using the MICCC as a cost function. For an input signal $\mathbf{x}_k = [x_{k-N+1}, ..., x_k]^T$ at time instant k, the improper correntropy between the desired signal $\mathbf{d}_k = [d_{k-N+1}, ..., d_k]^T$ and the filter output $\mathbf{y}_k = [y_{k-N+1}, ..., y_k]^T$ is computed using a sliding window of N samples, to give

$$J_k = \frac{1}{\pi\sigma^2\sqrt{1-|\varrho|^2}} \frac{1}{N} \sum_{i=k-N+1}^{k} \exp\left[-\frac{e_i e_i^* - \Re\{\varrho e_i^{2*}\}}{\sigma^2(1-|\varrho|^2)}\right] \tag{9}$$

where $e_i = d_i - \underline{\mathbf{w}}_k^H\underline{\mathbf{x}}_i$. The cost function J_k is maximised with respect to $\underline{\mathbf{w}}_k$ using gradient ascent [11], that is, based on $\underline{\mathbf{w}}_{k+1} = \underline{\mathbf{w}}_k + \mu\frac{\partial J_k}{\partial \underline{\mathbf{w}}_k^*}$. The computation of the derivative $\frac{\partial J_k}{\partial \underline{\mathbf{w}}_k^*}$ can be simplified through the $\mathbb{CR}$ (or Wirtinger) derivative chain rule [18, 11]

$$\frac{\partial J_k}{\partial \underline{\mathbf{w}}^*} = \frac{\partial J_k}{\partial e}\frac{\partial e}{\partial \underline{\mathbf{w}}^*} + \frac{\partial J_k}{\partial e^*}\frac{\partial e^*}{\partial \underline{\mathbf{w}}^*}. \tag{10}$$

With $\frac{\partial e}{\partial \underline{\mathbf{w}}^*} = -\underline{\mathbf{x}}$ and $\frac{\partial e^*}{\partial \underline{\mathbf{w}}^*} = \mathbf{0}$, equation (10) reduces to

$$\frac{\partial J_k}{\partial \underline{\mathbf{w}}^*} = -\frac{\partial J_k}{\partial e}\underline{\mathbf{x}} = -\frac{\partial \kappa_{\sigma,\varrho}(e)}{\partial e}\underline{\mathbf{x}}. \tag{11}$$

To simplify the derivation of $\frac{\partial J_k}{\partial e}$, we assume an unbiased estimation with $E\{e\} = 0$, such that $\frac{\partial \varrho}{\partial e} = \frac{2E\{e\}}{\sigma^2} = 0$, to give

$$\frac{\partial J_k}{\partial \underline{\mathbf{w}}^*} = E\left\{\frac{\kappa_{\sigma,\varrho}(e)}{\sigma^2(1-|\varrho|^2)}(e^* - \varrho^* e)\underline{\mathbf{x}}\right\}. \tag{12}$$

Therefore, the weight update for the filter in (7) becomes

$$\underline{\mathbf{w}}_{k+1} = \underline{\mathbf{w}}_k + \mu\frac{\sum_{i=k-N+1}^{k} \kappa_{\sigma,\varrho}(e_i)(e_i^* - \varrho^* e_i)\underline{\mathbf{x}}_i}{\sigma^2(1-|\varrho|^2)N}. \tag{13}$$

The instantaneous approximation ($N = 1$) finally yields the weight update of the proposed widely linear correntropy adaptive filter in the form

$$\underline{\mathbf{w}}_{k+1} = \underline{\mathbf{w}}_k + \mu\frac{\kappa_{\sigma,\varrho}(e_k)(e_k^* - \varrho^* e_k)\underline{\mathbf{x}}_k}{\sigma^2(1-|\varrho|^2)}. \tag{14}$$

5. CONVERGENCE ANALYSIS

The mean and mean-square convergence analyses use the following standard independence assumptions:

A1. The desired response is produced by a WL model given by

$$d_k = \mathbf{h}_{opt}^H\mathbf{x} + \mathbf{g}_{opt}^H\mathbf{x}^* + \eta_k = \underline{\mathbf{w}}_{opt}^H\underline{\mathbf{x}}_k + \eta_k, \tag{15}$$

where η_k is complex circular Gaussian noise, that is $E\{\eta_k^2\} = 0$, which is uncorrelated with $\underline{\mathbf{x}}_k$, and $\underline{\mathbf{w}}_{opt}$ is the optimal weight vector.

A2. The input $\underline{\mathbf{x}}_k$ is correlated second-order noncircular such that the off-diagonal elements of $\underline{\mathbf{R}}_k = E\{\underline{\mathbf{x}}_k\underline{\mathbf{x}}_k^H\}$ and $\underline{\mathbf{P}}_k = E\{\underline{\mathbf{x}}_k\underline{\mathbf{x}}_k^T\}$ do exist.

A3. The error nonlinearity $\kappa_{\sigma,\varrho}(e_k)$ is asymptotically uncorrelated with $E\{\underline{\mathbf{x}}_k\underline{\mathbf{x}}_k^H\}$ and $E\{\underline{\mathbf{x}}_k\underline{\mathbf{x}}_k^T\}$ at the steady state.

A4. The filter is long enough such that the a priori error is zero-mean Gaussian.

The assumption A1 is common, while for a long enough filter the assumption A3 also becomes realistic. Assumption A4 is reasonable owing to the central limit theorem, and also remains valid in the whole adaptation stage [19, 20, 21, 22]. For detailed derivations, we refer to [23].

5.1. Convergence in the mean

Consider the weight error vector, given by $\underline{\mathbf{v}}_k = \underline{\mathbf{w}}_k - \underline{\mathbf{w}}_{opt}$, so that the estimation error can be expressed in terms of $\mathbf{v}_k$, as

$$e_k = d_k - \left(\underline{\mathbf{v}}_k + \underline{\mathbf{w}}_{opt}\right)^H \underline{\mathbf{x}}_k = \eta_k - \underline{\mathbf{v}}_k^H \underline{\mathbf{x}}_k. \quad (16)$$

For convenience, we introduce the variable

$$\overline{\mu} = \frac{\mu\kappa_{\sigma,\varrho}(e_k)}{\sigma^2(1-|\varrho|^2)}. \quad (17)$$

Upon inserting (16-17) into (14) we arrive at

$$\underline{\mathbf{v}}_{k+1} = \underline{\mathbf{v}}_k + \overline{\mu}\left(\eta_k^* - \underline{\mathbf{x}}_k^H \underline{\mathbf{v}}_k - \varrho^*\eta_k + \varrho^*\underline{\mathbf{x}}_k^T\underline{\mathbf{v}}_k^*\right)\underline{\mathbf{x}}_k. \quad (18)$$

Under the convergence assumptions A1-A4, and upon taking the statistical expectations on both sides, we obtain

$$E\{\underline{\mathbf{v}}_{k+1}\} = E\{\underline{\mathbf{v}}_k\} + E\{\overline{\mu}\}\left(\varrho^*\underline{\mathbf{P}}_k E\{\underline{\mathbf{v}}_k^*\} - \underline{\mathbf{R}}_k E\{\underline{\mathbf{v}}_k\}\right) \quad (19)$$

where $\underline{\mathbf{R}}_k = E\{\underline{\mathbf{x}}_k\underline{\mathbf{x}}_k^H\}$ and $\underline{\mathbf{P}}_k = E\{\underline{\mathbf{x}}_k\underline{\mathbf{x}}_k^T\}$ are the covariance and pseudo-covariance matrices of the input data.

At the steady state the terms that include high powers of e_k can be neglected [24], unless the measurements of the desired signals d_k are extremely noisy. Further, it is insightful to inspect the first two terms of the Taylor series expansion of $E\{\overline{\mu}\}$, based on (6), that is

$$E\{\overline{\mu}\} \approx \frac{\mu}{\pi\sigma^4(1-|\varrho|^2)^{\frac{3}{2}}}\left[1 - \frac{\sigma_\eta^2 + \mathrm{tr}\left(\underline{\mathbf{R}}_k\underline{\mathbf{K}}_k - \Re\{\varrho\underline{\mathbf{P}}_k^*\underline{\mathbf{G}}_k\}\right)}{\sigma^2(1-|\varrho|^2)^{\frac{1}{2}}}\right] \quad (20)$$

where $\underline{\mathbf{K}}_k = E\{\underline{\mathbf{v}}_k\underline{\mathbf{v}}_k^H\}$ and $\underline{\mathbf{G}}_k = E\{\underline{\mathbf{v}}_k\underline{\mathbf{v}}_k^T\}$ are respectively the covariance and pseudo-covariance matrices of the weight error vector. Observe that the mean characteristics of the MICCC stochastic gradient algorithm depend on the kernel size, σ^2, the input data, $\underline{\mathbf{x}}_k$, the weight error vector, $\underline{\mathbf{v}}_k$, the circularity of the estimation error, ϱ, and the minimum MSE, σ_η^2.

Remark 3: With the assumptions of circular estimation error, that is $\varrho = 0$, and sufficiently large kernel size, σ^2, the mean behaviour of the stochastic gradient correntropy filter approaches that of the ACLMS, given by $E\{\underline{\mathbf{v}}_{k+1}\} = (\mathbf{I} - \mu\underline{\mathbf{R}}_k)\,E\{\underline{\mathbf{v}}_k\}$.

5.2. Convergence in the mean-square

Consider the evolution of the weight error covariance matrix, $\underline{\mathbf{K}}_k$, which can be used to determine the MSE through the relation $E\{|e_k|^2\} = \sigma_\eta^2 + \mathrm{tr}(\underline{\mathbf{R}}_k\underline{\mathbf{K}}_k)$, where $\mathrm{tr}(\underline{\mathbf{R}}_k\underline{\mathbf{K}}_k)$ is the excess MSE at time instant k [25]. The computation of $E\{\underline{\mathbf{v}}_{k+1}\underline{\mathbf{v}}_{k+1}^H\}$ is based on (18), and upon taking the expectations of the fourth-order moments using Isserlis' theorem [26] for Gaussian vectors, we obtain

$$\begin{aligned}\underline{\mathbf{K}}_{k+1} = {} & \underline{\mathbf{K}}_k + E\{\overline{\mu}\}\left[-\underline{\mathbf{R}}_k\underline{\mathbf{K}}_k + \varrho^*\underline{\mathbf{P}}_k\underline{\mathbf{G}}_k^* - \underline{\mathbf{K}}_k\underline{\mathbf{R}}_k + \varrho\underline{\mathbf{G}}_k\underline{\mathbf{P}}_k^*\right] \\ & + E\{\overline{\mu}^2\}\left[\left(1+|\varrho|^2\right)\left(\sigma_\eta^2\underline{\mathbf{R}}_k + \underline{\mathbf{R}}_k\underline{\mathbf{K}}_k\underline{\mathbf{R}}_k + \underline{\mathbf{P}}_k\underline{\mathbf{K}}_k^T\underline{\mathbf{P}}_k^*\right)\right. \\ & - 2\varrho\underline{\mathbf{R}}_k\underline{\mathbf{G}}_k\underline{\mathbf{P}}_k^* - 2\varrho^*\underline{\mathbf{P}}_k\underline{\mathbf{G}}_k^*\underline{\mathbf{R}}_k \\ & \left. + \underline{\mathbf{R}}_k\mathrm{tr}\left(\left(1+|\varrho|^2\right)\underline{\mathbf{R}}_k\underline{\mathbf{K}}_k - 2\Re\left\{\varrho^*\underline{\mathbf{P}}_k\underline{\mathbf{G}}_k^*\right\}\right)\right].\end{aligned} \quad (21)$$

Next, consider the unitary matrix $\mathbf{Q}$ derived from the approximate uncorrelating transform (AUT) [27] which diagonalizes the pseudo-covariance matrix, $\mathbf{P} = E\{\mathbf{x}\mathbf{x}^T\}$, as $\mathbf{P} = \mathbf{Q}\boldsymbol{\Lambda}_{\mathbf{P}}\mathbf{Q}^T$, with $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$ and $\boldsymbol{\Lambda}_{\mathbf{P}} = \mathrm{diag}\{p_{\max}, ..., p_{\min}\}$ being a diagonal matrix of real-valued entries (circularity coefficients). The AUT also simultaneously approximately diagonalizes the covariance matrix, $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$, as $\mathbf{R} \approx \mathbf{Q}\boldsymbol{\Lambda}_{\mathbf{R}}\mathbf{Q}^H$ with $\boldsymbol{\Lambda}_{\mathbf{R}} = \mathrm{diag}\{\lambda_{\max}, ..., \lambda_{\min}\}$ being the diagonal matrix with the eigenvalues of $\mathbf{R}$. Therefore, in this way $\underline{\mathbf{R}}_k$ and $\underline{\mathbf{P}}_k$ can be jointly diagonalized as $\underline{\mathbf{P}}_k = \underline{\mathbf{Q}}\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\underline{\mathbf{Q}}^T$ and $\underline{\mathbf{R}}_k \approx \underline{\mathbf{Q}}\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\underline{\mathbf{Q}}^H$, where $\underline{\mathbf{Q}}$ has the form [25]

$$\underline{\mathbf{Q}} = \frac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{Q} & -\mathbf{Q} \\ \mathbf{Q}^* & \mathbf{Q}^*\end{bmatrix}. \quad (22)$$

Notice that the diagonal matrices $\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}$ and $\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}$ are identical except for the opposite signs of the last N diagonal elements, that is

$$\underline{\boldsymbol{\Lambda}}_{\mathbf{R}} = \begin{bmatrix}\boldsymbol{\Lambda}_{\mathbf{R}} + \boldsymbol{\Lambda}_{\mathbf{P}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\mathbf{R}} - \boldsymbol{\Lambda}_{\mathbf{P}}\end{bmatrix},\ \underline{\boldsymbol{\Lambda}}_{\mathbf{P}} = \begin{bmatrix}\boldsymbol{\Lambda}_{\mathbf{R}} + \boldsymbol{\Lambda}_{\mathbf{P}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\mathbf{P}} - \boldsymbol{\Lambda}_{\mathbf{R}}\end{bmatrix}. \quad (23)$$

Now, pre- and post- multiply both sides of (21) with the unitary matrices $\underline{\mathbf{Q}}^H$ and $\underline{\mathbf{Q}}$ to arrive at

$$\begin{aligned}\underline{\tilde{\mathbf{K}}}_{k+1} = {} & \underline{\tilde{\mathbf{K}}}_k + 2E\{\overline{\mu}\}\left[\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\Re\left\{\varrho^*\underline{\tilde{\mathbf{G}}}_k^*\right\} - \underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\underline{\tilde{\mathbf{K}}}_k\right] \\ & + E\{\overline{\mu}^2\}\left[\left(1+|\varrho|^2\right)\left(\sigma_\eta^2\underline{\boldsymbol{\Lambda}}_{\mathbf{R}} + \underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2\underline{\tilde{\mathbf{K}}}_k + \underline{\boldsymbol{\Lambda}}_{\mathbf{P}}^2\underline{\tilde{\mathbf{K}}}_k^T\right)\right. \\ & + \underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\mathrm{tr}\left(\left(1+|\varrho|^2\right)\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\underline{\tilde{\mathbf{K}}}_k - 2\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\Re\left\{\varrho^*\underline{\tilde{\mathbf{G}}}_k^*\right\}\right) \\ & \left. - 4\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\Re\left\{\varrho^*\underline{\tilde{\mathbf{G}}}_k^*\right\}\right].\end{aligned} \quad (24)$$

where $\underline{\tilde{\mathbf{K}}}_k = \underline{\mathbf{Q}}^H\underline{\mathbf{K}}_k\underline{\mathbf{Q}}$ and $\underline{\tilde{\mathbf{G}}}_k = \underline{\mathbf{Q}}^H\underline{\mathbf{G}}_k\underline{\mathbf{Q}}$ are the rotated weight error covariance and pseudo-covariance matrices.

The diagonal elements of $\underline{\tilde{\mathbf{K}}}_k$, $\underline{\tilde{\mathbf{G}}}_k$ can be combined into the respective vectors $\boldsymbol{\kappa}_k$ and $\boldsymbol{\gamma}_k$ and admit the recursion

$$\boldsymbol{\kappa}_{k+1} = \mathbf{A}\boldsymbol{\kappa}_k + \mathbf{B}\Re\{\varrho^*\boldsymbol{\gamma}_k^*\} + E\{\overline{\mu}^2\}\left(1+|\varrho|^2\right)\sigma_\eta^2\mathbf{r}, \quad (25)$$

$$\mathbf{A} = \left[\mathbf{I} - 2E\{\overline{\mu}\}\underline{\boldsymbol{\Lambda}}_{\mathbf{R}} + \left(1+|\varrho|^2\right)E\{\overline{\mu}^2\}\left(2\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2 + \mathbf{r}\mathbf{r}^T\right)\right], \quad (26)$$

$$\mathbf{B} = 2\left[E\{\overline{\mu}\}\underline{\boldsymbol{\Lambda}}_{\mathbf{P}} - 2E\{\overline{\mu}^2\}\left(\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\underline{\boldsymbol{\Lambda}}_{\mathbf{P}} + \mathbf{r}\mathbf{p}^T\right)\right], \quad (27)$$

where $\mathbf{r}$ and $\mathbf{p}$ contain the diagonal elements of $\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}$ and $\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}$, respectively. We have exploited the equivalence $\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2 = \underline{\boldsymbol{\Lambda}}_{\mathbf{P}}^2$ in (26).

Remark 4: With the relaxing assumptions of circular estimation error, $\varrho = 0$, and sufficiently large kernel size, σ^2, the mean-square behaviour of MICCC degenerates into that of the ACLMS algorithm in [25].

Remark 5: With the inclusion of $\boldsymbol{\gamma}_k$ in (25), the standard convergence analysis is not sufficient for a complete convergence analysis. Therefore it is necessary to perform the complementary convergence analysis [28] as well.

5.3. Complementary convergence in the mean

We next consider the evolution of the weight error pseudo-covariance matrix, $\underline{\mathbf{G}}_k$, which is necessary to determine the complementary MSE (CMSE) through the relation $E\{e_k^2\} = \mathrm{tr}(\underline{\mathbf{P}}_k\underline{\mathbf{G}}_k^*)$ [28], which is the excess CMSE at time instant k [25]. Upon computing $E\{\underline{\mathbf{v}}_{k+1}\underline{\mathbf{v}}_{k+1}^T\}$ based on (18), and taking the expectations of the fourth-order moments using Isserlis' theorem [26] for Gaussian vectors, we obtain

$$\begin{aligned}\underline{\mathbf{G}}_{k+1} = {} & \underline{\mathbf{G}}_k + E\{\overline{\mu}\}\left[-\underline{\mathbf{G}}_k\underline{\mathbf{R}}_k^T + \varrho^*\underline{\mathbf{K}}_k\underline{\mathbf{P}}_k - \underline{\mathbf{R}}_k\underline{\mathbf{G}}_k + \varrho^*\underline{\mathbf{P}}_k\underline{\mathbf{K}}_k^T\right] \\ & + E\{\overline{\mu}^2\}\left[-2\varrho^*\sigma_\eta^2\underline{\mathbf{P}}_k + 2\underline{\mathbf{R}}_k\underline{\mathbf{G}}_k\underline{\mathbf{R}}_k^T + 2\varrho^{*2}\underline{\mathbf{P}}_k\underline{\mathbf{G}}_k^*\underline{\mathbf{P}}_k\right. \\ & - 2\varrho^*\left(\underline{\mathbf{R}}_k\underline{\mathbf{K}}_k\underline{\mathbf{P}}_k + \underline{\mathbf{P}}_k\underline{\mathbf{K}}_k^T\underline{\mathbf{R}}_k^T\right) \\ & \left. + \underline{\mathbf{P}}_k\mathrm{tr}\left(\underline{\mathbf{P}}_k^*\underline{\mathbf{G}}_k - 2\varrho^*\underline{\mathbf{R}}_k\underline{\mathbf{K}}_k + \varrho^{*2}\underline{\mathbf{P}}_k\underline{\mathbf{G}}_k^*\right)\right].\end{aligned} \quad (28)$$

Now, pre- and post- multiply both sides of the above equations with the unitary matrices $\underline{\mathbf{Q}}^H$ and $\underline{\mathbf{Q}}$, derived from the AUT, to yield

$$\begin{aligned}\underline{\tilde{\mathbf{G}}}_{k+1} = {} & \underline{\tilde{\mathbf{G}}}_k + E\{\overline{\mu}\}\left[\varrho^*\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\left(\underline{\tilde{\mathbf{K}}}_k + \underline{\tilde{\mathbf{K}}}_k^T\right) - 2\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\underline{\tilde{\mathbf{G}}}_k\right] \\ & + E\{\overline{\mu}^2\}\left[-2\varrho^*\sigma_\eta^2\underline{\boldsymbol{\Lambda}}_{\mathbf{P}} + 2\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2\underline{\tilde{\mathbf{G}}}_k + 2\varrho^{*2}\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}^2\underline{\tilde{\mathbf{G}}}_k^*\right. \\ & - 2\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\varrho^*\left(\underline{\tilde{\mathbf{K}}}_k + \underline{\tilde{\mathbf{K}}}_k^T\right) \\ & \left. + \underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\mathrm{tr}\left(\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\underline{\tilde{\mathbf{G}}}_k - 2\varrho^*\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\underline{\tilde{\mathbf{K}}}_k + \varrho^{*2}\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}\underline{\tilde{\mathbf{G}}}_k^*\right)\right].\end{aligned} \quad (29)$$

The diagonal elements of this expression admit the recursion

$$\boldsymbol{\gamma}_{k+1} = \left[\mathbf{I} - 2E\{\overline{\mu}\}\underline{\boldsymbol{\Lambda}}_{\mathbf{R}} + E\{\overline{\mu}^2\}\left(2\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2 + \mathbf{r}\mathbf{r}^T\right)\right]\boldsymbol{\gamma}_k \qquad (30)$$
$$+ E\{\overline{\mu}^2\}\varrho^{*2}\left[2\underline{\boldsymbol{\Lambda}}_{\mathbf{P}}^2 + \mathbf{r}\mathbf{r}^T\right]\boldsymbol{\gamma}_k^* + \varrho^*\mathbf{B}^T\boldsymbol{\kappa}_k - 2E\{\overline{\mu}^2\}\varrho^*\sigma_\eta^2\mathbf{r}.$$

The real-valued nature (decoupled real and imaginary parts) of the evolution of $\boldsymbol{\kappa}_k$ guarantees that the real and imaginary parts of $\boldsymbol{\gamma}_k$ evolve independently, so that we arrive at

$$\Re\left\{\varrho^*\boldsymbol{\gamma}_{k+1}^*\right\} = \mathbf{A}\Re\left\{\varrho^*\boldsymbol{\gamma}_k^*\right\} + |\varrho|^2\mathbf{B}^T\boldsymbol{\kappa}_k - 2E\{\overline{\mu}^2\}|\varrho|^2\sigma_\eta^2\mathbf{r}. \qquad (31)$$

Upon combining (25) and (31), we obtain the recursion for the augmented variable $\underline{\mathbf{S}}_k = \left[\boldsymbol{\kappa}_k^T, \ \Re\left\{\varrho^*\boldsymbol{\gamma}_k^H\right\}\right]^T$, which assumes the form

$$\underline{\mathbf{S}}_{k+1} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ |\varrho|^2\mathbf{B}^T & \mathbf{A} \end{bmatrix}}_{\underline{\mathbf{A}}}\underline{\mathbf{S}}_k + \begin{bmatrix} \left(1+|\varrho|^2\right) \\ -2|\varrho|^2 \end{bmatrix} E\{\overline{\mu}^2\}\sigma_\eta^2\mathbf{r}. \qquad (32)$$

Remark 6: The recursion in (32) depends on both the standard and complementary convergence analyses, and reduces to the standard convergence of the ACLMS [25] for circular estimation error, that is, for $\varrho = 0$.

5.4. Mean-square stability

For the recursion (32) to converge, the eigenvalues of $\underline{\mathbf{A}}$ have to be less than unity. Instead of attempting to determine the eigenvalues of $\underline{\mathbf{A}}$ directly, we use majorization inequalities of Hermitian block matrices [29], the Gantmacher theorem [30] and the Weyl inequality, to state that for a Hermitian positive semidefinite 2×2 block matrix $\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{M} & \mathbf{K} \\ \mathbf{K}^* & \mathbf{N} \end{bmatrix}$, then $\overline{\lambda}\left[\mathbf{M}\right] + \overline{\lambda}\left[\mathbf{N}\right] \geq \overline{\lambda}\left[\mathbf{M}+\mathbf{N}\right] \geq \overline{\lambda}\left[\underline{\mathbf{H}}\right]$, where the $\overline{\lambda}\left[\cdot\right]$ operator denotes the maximum eigenvalue [31]. Now,

$$\overline{\lambda}\left[\underline{\mathbf{A}}\right] \leq 2\overline{\lambda}\left[\mathbf{I}\right] - 4E\{\overline{\mu}\}\overline{\lambda}\left[\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}\right]$$
$$+ 2\left(1+|\varrho|^2\right)E\{\overline{\mu}^2\}\left[\overline{\lambda}\left[\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2\right] + \overline{\lambda}\left[\mathbf{r}\mathbf{r}^T\right]\right], \qquad (33)$$

and since $\overline{\lambda}\left[\mathbf{r}\mathbf{r}^T\right] = \mathrm{tr}\left(\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2\right)$, the condition $\overline{\lambda}\left[\underline{\mathbf{A}}\right] < 1$ reduces to

$$1 - 2E\{\overline{\mu}\}r_{\min} + \left(1+|\varrho|^2\right)E\{\overline{\mu}^2\}\left(2r_{\max}^2 + \mathrm{tr}(\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2)\right) < 1. \qquad (34)$$

The inequality still holds if $\mathrm{tr}(\underline{\boldsymbol{\Lambda}}_{\mathbf{R}}^2)$ is replaced by $2Nr_{\max}^2$, and therefore the MICCC-based stochastic gradient algorithm achieves mean-square stability for

$$0 < \mu < \frac{\pi\sigma^2(1-|\varrho|^2)r_{\min}}{\left(1+|\varrho|^2\right)E\{\kappa_{\sigma,\varrho}(e_k)\}\left(N+1\right)r_{\max}^2}. \qquad (35)$$

Upon dividing the numerator and denominator with $r_{\min}$ and recognizing that the maximum eigenvalue of the augmented covariance matrix, $r_{\max}$, is the sum of the eigenvalues of the covariance, $\mathbf{R}$, and pseudo-covariance, $\mathbf{P}$, matrices that is, $r_{\max} = \lambda_{\max} + p_{\max}$, gives [25]

$$0 < \mu < \frac{\pi\sigma^2(1-|\varrho|^2)}{\left(1+|\varrho|^2\right)E\{\kappa_{\sigma,\varrho}(e_k)\}\left(N+1\right)s\left[\underline{\mathbf{R}}_k\right]\left(\lambda_{\max}+p_{\max}\right)}, \qquad (36)$$

where the eigenvalue spread of the augmented covariance matrix is

$$s\left[\underline{\mathbf{R}}_k\right] = \frac{r_{\max}}{r_{\min}} = \frac{\lambda_{\max}+p_{\max}}{\lambda_{\min}-p_{\min}}. \qquad (37)$$

5.5. Steady-state analysis

The steady-state values of $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ are given by

$$\boldsymbol{\kappa}_\infty = \left[\mathbf{I}-\mathbf{A}\right]^{-1}\left[\mathbf{B}\Re\left\{\varrho^*\boldsymbol{\gamma}_\infty^*\right\} + E\{\overline{\mu}^2\}\left(1+|\varrho|^2\right)\sigma_\eta^2\mathbf{r}\right], \qquad (38)$$

$$\Re\left\{\varrho^*\boldsymbol{\gamma}_\infty^*\right\} = \left[\mathbf{I}-\mathbf{A}\right]^{-1}\left[|\varrho|^2\mathbf{B}^T\boldsymbol{\kappa}_\infty - 2E\{\overline{\mu}^2\}|\varrho|^2\sigma_\eta^2\mathbf{r}\right], \qquad (39)$$

Upon combining (38) and (39), the steady-state misadjustment can be expressed as

$$M_{\text{MICCC}} = \frac{\mathbf{r}^T\boldsymbol{\kappa}_\infty}{\sigma_\eta^2} = \left[\left[\mathbf{I}-\mathbf{A}\right] - |\varrho|^2\mathbf{B}\left[\mathbf{I}-\mathbf{A}\right]^{-1}\mathbf{B}^T\right]^{-1} \times$$
$$E\{\overline{\mu}^2\}\left[\left(1+|\varrho|^2\right)\mathbf{I} - 2\mathbf{B}\left[\mathbf{I}-\mathbf{A}\right]^{-1}\right] \qquad (40)$$

Remark 7: The steady-state misadjustment of the MICCC algorithm increases with the increase in noncircularity of the estimation error.

6. SIMULATIONS

The optimum weights in (7) were chosen arbitrarily as

$$\mathbf{h}_{\text{opt}} = \left[1-2\jmath, -3+4\jmath\right]^T, \quad \mathbf{g}_{\text{opt}} = \left[2+0.5\jmath, -2+2\jmath\right]^T \qquad (41)$$

and the complex input signal, $\underline{\mathbf{x}}$, was proper Gaussian noise. The real and imaginary parts of the noise, η_k, were characterized by the respective pdfs $0.9\mathcal{N}(0,1)$ and $\mathcal{N}(0,10)$, where the large variance of 10 provides large impulsive disturbances. The performance of the MICCC-based stochastic gradient (MICCC) was compared to its "proper" MCCC-based counterpart (MCCC) [10] and the established CLMS and augmented CLMS (ACLMS) algorithms [32, 11, 33]. The weight signal-to-noise ratio (WSNR), defined as

$$\text{WSNR}_{dB} = 10\log_{10}\left(\frac{\underline{\mathbf{w}}_{opt}^H\underline{\mathbf{w}}_{opt}}{\left(\underline{\mathbf{w}}_{opt}-\underline{\mathbf{w}}_k\right)^H\left(\underline{\mathbf{w}}_{opt}-\underline{\mathbf{w}}_k\right)}\right), \qquad (42)$$

was used to quantify both convergence and misadjustment [8], where $\underline{\mathbf{w}}_k = \left[\mathbf{h}_k^T, \mathbf{g}_k^T\right]^T$ is the weight vector computed at the time instant k for the widely linear algorithms (MICCC and ACLMS), and $\underline{\mathbf{w}}_k = \left[\mathbf{h}_k^T, \mathbf{0}\right]^T$ for the strictly linear algorithms (MCCC and CLMS).

Fig. 1 shows the average WSNR produced by 1000 Monte Carlo trials, with the initial value for the weights set to zero. The parameters in all algorithms considered were tuned such that their steady-state WSNR were equal in a Gaussian environment.

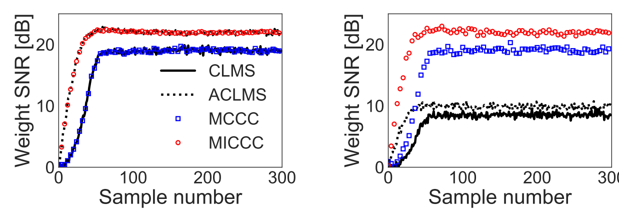

Fig. 1: Weight signal to noise ratio (WSNR) of MICCC, MCCC, CLMS and ACLMS under Gaussian proper noise (left panel) and impulsive improper noise (right panel).

Fig. 1 illustrates that the outliers in non-Gaussian environments negatively impact the performance of the MSE-based algorithms, while the correntropy-based algorithms were unaffected. Owing to its inherent account of noncircularity, ϱ, the MICCC exhibited a significantly enhanced convergence rate and WSNR over the proper MCCC and the second-order statistics-based CLMS and ACLMS.

7. CONCLUSIONS

We have extended the definition of complex correntropy to account for a general class of complex-valued data with noncircular distributions. This has been achieved through a probabilistic interpretation of the complex correntropy, and has served as a basis for a new stochastic gradient algorithm with the cost function in the form of the maximum improper correntropy criterion (MICCC). The analysis and simulations have demonstrated that, with noncircularity accounted for by MICCC, the proposed method offers faster convergence rates and greater WSNR in both Gaussian and non-Gaussian environments. Future work aims at the development of a whole class of MICCC-based algorithms, such as Kalman and MVDR filters.

8. REFERENCES

[1] I. Santamaria, P. P. Pokharel, and J. C. Principe, "Generalized Correlation Function: Definition, Properties, and Applications to Blind Equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.

[2] J. C. Principe, D. Xu, and J. Fisher, "Information theoretic learning," in *Unsupervised Adaptive Filtering*, S. Haykin, Ed. Wiley.

[3] L. Weifeng, P. P. Poharel, and J. C. Principe, "Correntropy: Properties and Applications in Non-Gaussian Signal Processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[4] A. Gunduz and J. C. Principe, "Correntropy as a Novel Measure for Nonlinearity Tests," *Signal Processing*, vol. 89, pp. 14–23, 2009.

[5] A. I. R. Fontes, M. A. M, L. F. Silveira, and J. C. Principe, "Performance Evaluation of The Correntropy Coefficient in Automatic Modulation Classification," *Expert Systems with Applications*, vol. 42, no. 1, pp. 1–8, 2015.

[6] A. Singh and J. C. Principe, "Using Correntropy as a Cost Function in Linear Adaptive Filters," *In Proceedings of the International Joint Conference on Neural Networks*, pp. 2950–2955, 2009.

[7] R. Wang, B. Chen, N. Zheng, and J. C. Principe, "A Variable Step-Size Adaptive Algorithm under Maximum Correntropy Criterion," *In Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 1–5, 2015.

[8] A. Singh and J. C. Principe, "A Closed Form Recursive Solution for Maximum Correntropy Training," *In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 2070–2073, 2010.

[9] B. Chen, J. Wang, H. Zhao, N. Zheng, and J. C. Principe, "Convergence of a Fixed-Point Algorithm under Maximum Correntropy Criterion," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1723–1727, 2015.

[10] J. P. F. Guimaraes, A. I. R. Fontes, J. B. A. Rego, M. A. M., and J. C. Principe, "Complex Correntropy: Probabilistic Interpretation and Application to Complex-Valued Data," *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 42–45, 2016.

[11] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York: Wiley, 2009.

[12] E. Parzen, "On Estimation of a Probability Density Function and the Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[13] A. van den Bos, "Complex Gradient and Hessian," *Image Signal Processing*, vol. 141, no. 6, pp. 380–382, 1994.

[14] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[15] J. Eriksson, E. Ollila, and V. Koivunen, "Essential Statistics and Tools for Complex Random Variables," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5400–5408, 2010.

[16] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall, 1986.

[17] B. Picinbono and P. Chevalier, "Widely Linear Estimation with Complex Data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[18] K. Kreutz-Delgado, "The Complex Gradient Operator and the CR-Calculus," arXiv:0906.4835v1 [math.OC]. [Online]. Access: https://arxiv.org/abs/0906.4835, 2009.

[19] T. Y. Al-Naffouri and A. H. Sayed, "Adaptive Filters with Error Nonlinearities: Mean-Square Analysis and Optimum Design," *EURASIP Journal on Applied Signal Processing*, vol. 4, no. 1, pp. 192–205, 2001.

[20] N. R. Yousef and A. H. Sayed, "A Unified Approach to the Steady-State and Tracking Analyses of Adaptive Filters," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314–324, 2001.

[21] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New Jersey: John Wiley & Sons, 2003.

[22] T. Y. Al-Naffouri and A. H. Sayed, "Transient Analysis of Adaptive Filters with Error Nonlinearities," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 653–661, 2003.

[23] B. Scalzo Dees, Y. Xia, S. C. Douglas, and D. P. Mandic, "Correntropy-Based Adaptive Filtering of Noncircular Complex Data," *IEEE Transactions on Signal Processing, Submitted*.

[24] C. Boukis, D. P. Mandic, and A. G. Constantinides, "A Class of Stochastic Gradient Algorithms with Exponentiated Error Cost Functions," *Digital Signal Processing*, vol. 19, pp. 201–212, 2009.

[25] D. P. Mandic, S. Kanna, and S. C. Douglas, "Mean Square Analysis of the CLMS and ACLMS for Non-Circular Signals: The Approximate Uncorrelating Transform Approach," *In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3531–3535, 2015.

[26] L. Isserlis, "On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in Any Number of Variables)," *Biometrika*, vol. 12, no. 1/2, pp. 134–139, 1918.

[27] C. C. Took, S. C. Douglas, and D. P. Mandic, "On Approximate Diagonalization of Correlation Matrices in Widely Linear Signal Processing," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1469–1473, 2012.

[28] X. Yili and D. P. Mandic, "Complementary Mean Square Analysis of Augmented CLMS for Second-Order Noncircular Gaussian Signals," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1413–1417, 2017.

[29] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications*. New York: Springer Series in Statistics, Springer, 2011.

[30] J. B. Foley and F. M. Boland, "A Note on the Convergence Analysis of LMS Adaptive Filters with Gaussian Data," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 7, pp. 1087–1089, 1988.

[31] A. Knutson and T. Tao, "Honeycombs and Sums of Hermitian Matrices," *Notices of the American Mathematics Society*, vol. 48, no. 2, pp. 175–186, 2001.

[32] D. P. Mandic, S. Javidi, G. Souretis, and V. S. L. Goh, "The Augmented Complex Least Mean Square Algorithm with Application to Adaptive Prediction Problems," *In Proceedings of the 1st IARP Workshop on Cognitive Information Processing, Santorini, Greece*, pp. 54–57, 2008.

[33] D. P. Mandic, S. Javidi, S. Goh, A. Kuh, and K. Aihara, "Complex-Valued Prediction of Wind Profile Using Augmented Complex Statistics," *Renewable Energy*, vol. 34, no. 1, pp. 196–201, 2009.