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

Imperial College London

東南大學 SOUTHEAST UNIVERSITY

SMU.

Introduction

Opportunity: Recent advances in signal processing have highlighted demands for enhanced dealing with complex-valued data.

Challenges: The issues in processing such data include:

- **Noncircularity:** Real-world complex signals are often noncircular (rotation-dependent pdf). This can be accounted for by **widely linear modeling**, which caters for the **full available second-order statistics**.
- **Streaming Data:** Processing large data sets with block-based algorithms can be computationally prohibitive. To this end, **online algorithms** have proven effective and cater for non-stationary statistics.
- **Outliers:** Mean square error (MSE) based algorithms are sensitive to outliers. Robust cost functions such as **Maximum Complex Correntropy Criterion (MCCC)** have proven effective in the presence of outliers, and cater for impulsive non-Gaussian environments.

The standard MCCC cost function assumes second-order circular (proper) estimation error and **we here introduce a new, more comprehensive, definition of complex correntropy** to address the above challenges.

$\Longrightarrow$ The proposed **Maximum Improper Complex Correntropy Criterion (MICCC)** offers robust estimation for streaming noncircular data with non-stationary statistics in impulsive non-Gaussian environments.

Background – Correntropy as a cost function

Filtering: Estimate the output, $y \in \mathbb{R}$, from the input vector, $\boldsymbol{x} \in \mathbb{R}^N$, through a linear model given by

$$y = \boldsymbol{w}^T \boldsymbol{x} \tag{1}$$

The **MSE-based Wiener** solution has the form

$$\boldsymbol{w}_{\text{Wiener}} = E\left\{\boldsymbol{x}\boldsymbol{x}^T\right\}^{-1} E\left\{\boldsymbol{x}y\right\} \tag{2}$$

Problem: If the input $\boldsymbol{x}$ contains outliers $\rightarrow$ the Wiener solution is unreliable.

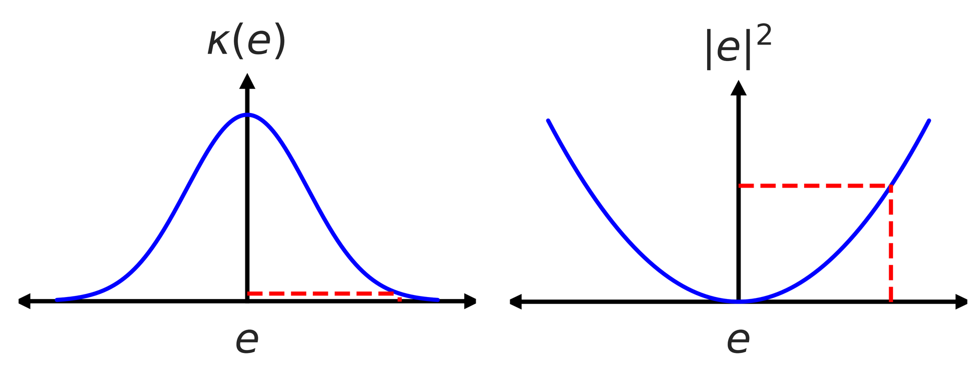
Maximum Correntropy Criterion (MCC): The MCC counteracts the sensitivity to outliers by interpreting the filtering problem as the estimation of the probability of the event $y = \boldsymbol{w}^T\boldsymbol{x}$. Using the Gaussian pdf, $\kappa(\cdot)$, the optimum filter weights are derived by maximizing the probability of estimation error, $e = y - \boldsymbol{w}^T\boldsymbol{x}$,

$$\max_{\boldsymbol{w}} \quad E\left\{\kappa(e)\right\} = \frac{1}{\sqrt{2\pi\sigma^2}} E\left\{\exp\left[-\frac{|e|^2}{2\sigma^2}\right]\right\} \tag{3}$$

The **MCC-Wiener-based estimation solution** then becomes

$$\boldsymbol{w}_{\text{MCC}} = E\left\{\kappa(e)\boldsymbol{x}\boldsymbol{x}^T\right\}^{-1} E\left\{\kappa(e)\boldsymbol{x}y\right\} \tag{4}$$

which is straightforwardly calculated using an appropriate Parzen window.



The maximization of the MCC **attenuates** the influence of outliers, whereas the minimization of the MSE **amplifies** the sensitivity to outliers.

Figure 1: Robustness to outliers of MCC (left panel) vs. MSE (right panel).

MICCC-based stochastic gradient adaptive filter

MICCC: A complex correntropy-based cost function suitable for **noncircularly distributed errors** is introduced as an extension to the work in [1]. For an improper complex random variable, $\boldsymbol{e} = [e_1, \dots, e_N]^T \in \mathbb{C}^N$, the complex correntropy is estimated through an appropriate Parzen estimator, given by

$$E\{\kappa_{\sigma,\varrho}(\boldsymbol{e})\} = \frac{1}{\pi\sigma^2\sqrt{1-|\varrho|^2}} \frac{1}{N}\sum_{n=1}^{N} \exp\left[-\frac{|e_n|^2 - \Re\{\varrho e_n^{*2}\}}{\sigma^2(1-|\varrho|^2)}\right] \tag{5}$$

where $\varrho = E\{\boldsymbol{e}^T\boldsymbol{e}\}/E\{\boldsymbol{e}^H\boldsymbol{e}\}$ is the **circularity quotient** of $\boldsymbol{e}$.

Robust widely linear model: We consider a **widely linear model** of the form

$$\hat{y} = \boldsymbol{h}^H\boldsymbol{x} + \boldsymbol{g}^H\boldsymbol{x}^* = \underline{\boldsymbol{w}}^H\underline{\boldsymbol{x}}, \tag{6}$$

where $\underline{\boldsymbol{x}} = \left[\boldsymbol{x}^T, \boldsymbol{x}^H\right]^T$ and $\underline{\boldsymbol{w}} = \left[\boldsymbol{h}^T, \boldsymbol{g}^T\right]^T$ are respectively the **augmented** input and coefficient vectors, with $\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{g} \in \mathbb{C}^N$.

Define the estimation error, $e = y - \hat{y}$, as the difference between the desired signal $y \in \mathbb{C}$ and the filter output $\hat{y} \in \mathbb{C}$. The new cost function is then defined as the MICCC between the random variables y and $\hat{y}$, and is given by

$$J_{\text{MICCC}} = E\{\kappa_{\sigma,\varrho}(e)\}. \tag{7}$$

MICCC-based stochastic gradient algorithm: For an input signal $\boldsymbol{x}_k = [x_{k-N+1}, \dots, x_k]^T \in \mathbb{C}^N$ at time instant k, the improper correntropy between the desired signal $\boldsymbol{y}_k = [y_{k-N+1}, \dots, y_k]^T \in \mathbb{C}^N$ and the filter output $\hat{\boldsymbol{y}}_k = [\hat{y}_{k-N+1}, \dots, \hat{y}_k]^T \in \mathbb{C}^N$ is computed using (5), where $e_i = y_i - \underline{\boldsymbol{w}}_k^H\underline{\boldsymbol{x}}_i$.

The value of J_{MICCC} at time intsant k, J_k, is maximised with respect to $\underline{\boldsymbol{w}}_k$ using **gradient ascent** [2], that is, based on $\underline{\boldsymbol{w}}_{k+1} = \underline{\boldsymbol{w}}_k + \mu\frac{\partial J_k}{\partial \underline{\boldsymbol{w}}_k^*}$. The computation of the derivative $\frac{\partial J_k}{\partial \underline{\boldsymbol{w}}_k^*}$ can be simplified through the **$\mathbb{CR}$ (or Wirtinger) derivative chain rule**, as

$$\frac{\partial J_k}{\partial \underline{\boldsymbol{w}}^*} = \frac{\partial J_k}{\partial e}\frac{\partial e}{\partial \underline{\boldsymbol{w}}^*} + \frac{\partial J_k}{\partial e^*}\frac{\partial e^*}{\partial \underline{\boldsymbol{w}}^*}. \tag{8}$$

With $\frac{\partial e}{\partial \underline{\boldsymbol{w}}^*} = -\underline{\boldsymbol{x}}$ and $\frac{\partial e^*}{\partial \underline{\boldsymbol{w}}^*} = \boldsymbol{0}$, equation (8) reduces to

$$\frac{\partial J_k}{\partial \underline{\boldsymbol{w}}^*} = -\frac{\partial J_k}{\partial e}\underline{\boldsymbol{x}} = -\frac{\partial \kappa_{\sigma,\varrho}(e)}{\partial e}\underline{\boldsymbol{x}}. \tag{9}$$

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

$$\frac{\partial J_k}{\partial \underline{\boldsymbol{w}}^*} = E\left\{\frac{\kappa_{\sigma,\varrho}(e)}{\sigma^2\left(1-|\varrho|^2\right)}(e^* - \varrho^* e)\underline{\boldsymbol{x}}\right\}. \tag{10}$$

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

$$\underline{\boldsymbol{w}}_{k+1} = \underline{\boldsymbol{w}}_k + \mu\frac{\kappa_{\sigma,\varrho}(e_k)(e_k^* - \varrho^* e_k)\underline{\boldsymbol{x}}_k}{\sigma^2(1-|\varrho|^2)}. \tag{11}$$

Simulations and Applications

Fig. 2 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. The weight signal-to-noise ratio (WSNR), defined as

$$\text{WSNR}_{dB} = 10\log_{10}\left(\frac{\underline{\boldsymbol{w}}_{opt}^H\underline{\boldsymbol{w}}_{opt}}{\left(\underline{\boldsymbol{w}}_{opt} - \underline{\boldsymbol{w}}_k\right)^H\left(\underline{\boldsymbol{w}}_{opt} - \underline{\boldsymbol{w}}_k\right)}\right) \tag{12}$$

was used to quantify both convergence and misadjustment.

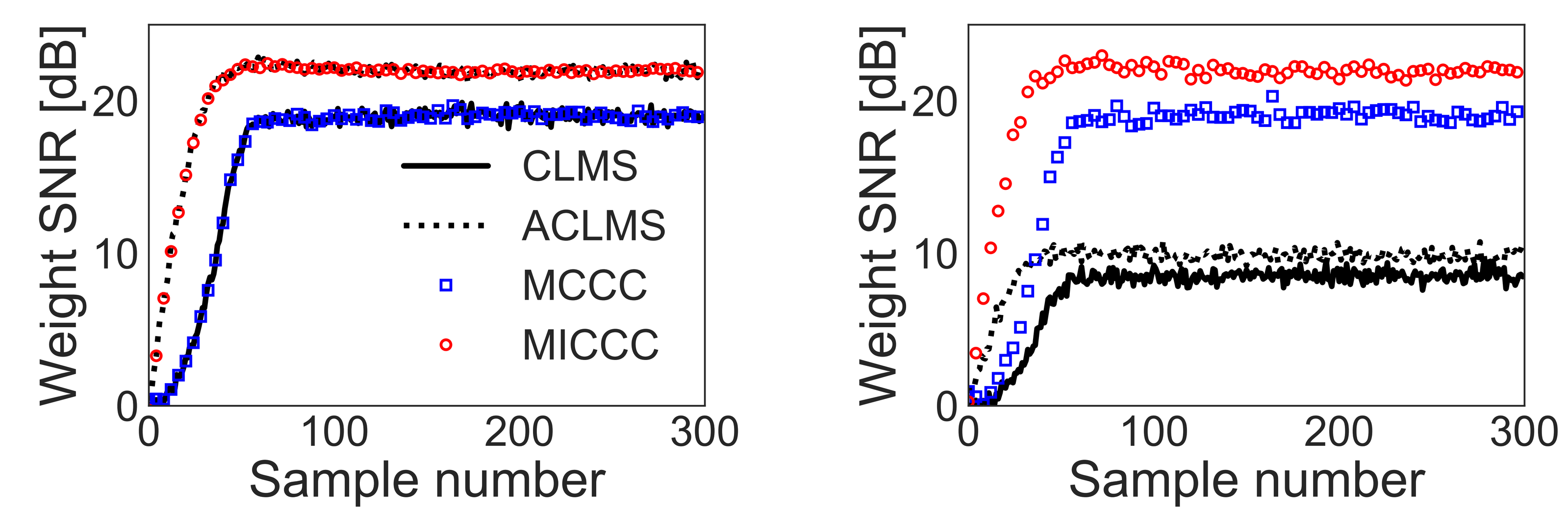


Figure 2: WSNR of MICCC, MCCC, CLMS and ACLMS under proper Gaussian noise (left panel) and impulsive improper noise (right panel).

Synthetic data: 1000 realisations of proper Gaussian noise, $\underline{\boldsymbol{x}}$, were generated, with the real and and imaginary parts of the noise, η_k, characterized by the respective pdfs $0.9\mathcal{N}(0,1)$ and $\mathcal{N}(0,10)$. The optimum weights were given by $\boldsymbol{h}_{\text{opt}} = [1-2\jmath, -3+4\jmath]^T$ and $\boldsymbol{g}_{\text{opt}} = [2+0.5\jmath, -2+2\jmath]^T$.

Conclusions

We have extended the definition of **complex correntropy** to account for complex-valued data with **noncircular distributions**. This 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.

Selected References

1. J.P.F Guimaraes and A.I.R. Fontes, J.B.A. Rego, M.A. Martins, and J.C. Principe, "Complex Correntropy: Probabilistic Interpretation and Application to Complex-Valued Data," in *IEEE Signal Processing Letters*, vol. 24, no. 1, pp. 42–45, 2016.
2. D.P. Mandic and V.S.L. Goh, "Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models." New York: Wiley, 2009.
3. 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.