Fractional-order Correntropy Adaptive Filters for Distributed Processing of α-Stable Signals

Vinay Chakravarthi Gogineni, *Member, IEEE*, Sayed Pouria Talebi, Stefan Werner, *Senior Member, IEEE*, and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*—This work revisits the problem of distributed adaptive filtering in multi-agent sensor networks. In contrast to classical approaches, the formulation relaxes the Gaussian assumption on the signal and noise to the generalized setting of α-stable distributions that do not possess second- and higher-order statistical moments. Most importantly, the considered scenario allows for different characteristic exponents throughout the network. Drawing upon ideas from correntropy-type local similarity measures and fractional-order calculus, a novel class of distributed fractional-order correntropy adaptive filters, that are robust against the jittery behavior of α-stable signals, is derived and their convergence criterion is established. The effectiveness of the proposed algorithms, as compared to existing distributed adaptive filtering techniques, is demonstrated via simulation examples.**

***Index Terms*—Distributed adaptive networks, α-stable signals, fractional-order calculus, correntropy criterion, consensus fusion.**

I. INTRODUCTION

Information processing over distributed multi-agent networks has recently attracted attention in many applications. Adaptive learning is one of the most important aspects of distributed information processing. In adaptive networks, the interconnected nodes are capable of performing real-time data processing and exchanging information with their neighbors [1]. The distributed adaptive learning techniques such as consensus, incremental, and and diffusion strategies allow these nodes to carry out estimation tasks in a collaborative manner [2]–[6]. The majority of current distributed signal processing approaches assume a Gaussian model for signal and noise [1], [2]. Although the Gaussian assumption leads to mathematically tractable and computationally efficient solutions, it becomes unrealistic in a large number of modern applications, such as cases where the encountered signal exhibit sharp spikes [7]–[12]. The behavior of such signals is accurately modelled by the class of symmetric α-stable (SαS) random processes [7]–[14]. Since SαS random processes do not possess finite second-order moments, except for the Gaussian case, traditional distributed Wiener filtering techniques, that are derived by minimizing the second-order moment of an error measure, exhibit considerable performance degradation when applied to lower-order SαS signals [10], [14].

To address this issue, a distributed particle filter is proposed in [15]; however, its considerable computational burden makes impractical for use in real-time applications. More recently, using the framework of fractional-order calculus [16]–[18], cost effective adaptive filtering methods have been proposed for processing α-stable signals [19]–[21]. Although these distributed fractional-order filters [19] achieve improved performance over the conventional distributed Wiener filtering techniques, their performance is highly sensitive with respect to the characteristic exponent value. Furthermore, the propagation of residual jitters present in the local estimates cause performance degradation.

On the other hand, correntropy criterion based adaptive filtering techniques are adapt to dealing with situations where the data is corrupted by heavier-tailed noise [22]–[29]. These correntropy based adaptive filters have been successfully extended to distributed networks [30]–[32]. However, maximum correntropy criterion (MCC) [23] and generalized maximum correntropy criterion (GMCC) [25] are defined over second and higher-order error moments, which compromises their performance, specially in cases where both the signal and noise are modeled as SαS signals. In summary, a comprehensive filtering framework that can deal with jittery behaviour in both signal and noise is lacking. Importantly, to the best of our knowledge, there is no framework that can accommodate signals characterized by different values of characteristic exponent across the network.

This paper proposes a novel class of distributed adaptive filtering techniques based on maximizing a new *fractional-order correntropy* criterion. Our main contributions here are as follows:

- Based on the principles of correntropy-type local similarity measure and fractional-order calculus, a novel class of distributed fractional-order correntropy adaptive filters that effectively regulate the strong jittery behavior of SαS signals is derived;
- The derived class of filters render simultaneous learning from α-stable signals characterized by different values of characteristic exponent across the network;
- Conditions for the convergence of the proposed class of filters are established;

Mathematical Notations: We denote scalars, column vectors and matrices with lower case, bold lower case and bold uppercase letters, respectively, while $\mathbf{I}$ represents the identity matrix of appropriate size. Matrix transpose is denoted by $(\cdot)^{\mathrm{T}}$ and

This work was supported, in part, by the ERCIM Alain Bensoussan Fellowship Programme and the Research Council of Norway.

Vinay Chakravarthi Gogineni is with the SimulaMet, Simula Research Laboratory, Norway (e-mail: vcgogineni@simula.no).

Sayed Pouria Talebi and Stefan Werner are with the Norwegian University of Science and Technology, Norway (e-mail: pouria@ntnu.no and stefan.werner@ntnu.no).

Danilo P. Mandic is with the Imperial College London, London (e-mail: d.mandic@imperial.ac.uk).

the symbol $\otimes$ designates the right Kronecker product operator. Finally, $(\cdot)^{\langle\tau\rangle}$ denotes the elementwise implementation of the function $g(z)=|z|^{\tau}\text{sign}(z)$, where $\text{sign}(\cdot)$ and $|\cdot|$ return the sign and the absolute values of their input, respectively.

II. Preliminaries

In various real-world applications such as underwater acoustics [7], wideband communications [8], financial data modeling [9], audio signal processing [10], [11] and random fluctuations of gravitational fields [12], the simple Gaussian assumption on the signal and noise processes is not reasonable. These signals can be more accurately approximated by α-stable distributions.

In general, α-stable distributions do not have an explicit closed form expression for their probability distribution functions[1] However, the class of real-valued α-stable random processes with elliptically symmetric distributions, which is often referred as symmetric α-stable (SαS), have the characteristic function of the form [33]

$$\Phi_{\mathbf{z}}(\mathbf{s})=\mathrm{E}[\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{z})]=\exp(i\mathbf{s}^{\mathrm{T}}\boldsymbol{\xi})\exp\Big(-(\frac{1}{2}\mathbf{s}^{\mathrm{T}}\boldsymbol{\Gamma}_{\mathbf{z}}\mathbf{s})^{\frac{\alpha}{2}}\Big), \tag{1}$$

where $\Phi_{\mathbf{z}}(\cdot)$ is the characteristic function of $\mathbf{z}$, $i^2=-1$, with $\mathrm{E}[\cdot]$ denoting the statistical expectation operator. The positive definite covariance matrix, $\boldsymbol{\Gamma}_{\mathbf{z}}$, determines the elliptical shape of the distribution of $\mathbf{z}$ that is centered around the mean vector $\boldsymbol{\xi}$. The characteristic exponent, $\alpha\in(0,2]$ in (1) governs the tail heaviness of the density function [33]. Small values of α correspond to strong impulsiveness, resulting in heavier tails.

Excluding the Gaussian case, SαS random processes have only finite statistical moments of orders strictly less than α [14]. Therefore, when it comes to filtering solutions, it is implicitly assumed that $\alpha\in(1,2]$, so that conditional expectations can be established. Without loss of generality, hereafter it is assumed that signals are real-valued SαS random processes with $\alpha\in(1,2]$.

III. Proposed Distributed Solution

Consider a sensor network modeled as a connected graph $\mathcal{G}=\{\mathcal{N},\mathcal{E}\}$, where the node set $\mathcal{N}$ represents the agents of the network and $\mathcal{E}$ is the set of edges that represents bidirectional communication links between the nodes, i.e., $(k,l)\in\mathcal{E}$ if nodes k and l are connected. Furthermore, the set $\mathcal{N}_k$ denotes the neighborhood of node k including itself and the cardinality of the set $\mathcal{N}_k$ is denoted by $|\mathcal{N}_k|$. At time instant n, node k observes the unknown parameter vector $\mathbf{w}^*$ through the input signal vector $\mathbf{x}_{k,n}$ and output signal $y_{k,n}$, which are related via the linear model

$$y_{k,n}=\mathbf{x}_{k,n}^{\mathrm{T}}\ \mathbf{w}^{\star}+\upsilon_{k,n}, \tag{2}$$

where $\upsilon_{k,n}$ represents the observational noise and random sequences $\upsilon_{k,n}$ and $\mathbf{x}_{k,n}$ are assumed zero-mean and SαS. The objective here is to estimate $\mathbf{w}^*$ collaboratively over a network.

[1]Important special cases include the Lévy distribution ($\alpha=0.5$), Cauchy distribution ($\alpha=1$) and Gaussian distribution ($\alpha=2$).

Recently, correntropy criterion based adaptive filters are shown to be effective against presence of impulsive noise [23], [25], [26]. However, since SαS random signals have finite statistical moments of order less than α, MCC and GMCC based distributed adaptive filters [30], [31] are not a reasonable choice when both the signal and noise are SαS signals. The lack of second and higher-order moments of the error measure becomes the main hurdle in establishing the convergence criterion of the mentioned techniques. Furthermore, the usage of ordinary calculus makes these techniques unstable [19]. In order to overcome these issues, we use the *fractional-order correntropy* criterion which is defined over the fractional-order of the error measure [34]. Therefore, the estimates $\{\mathbf{w}_n: n=1,2,\ldots\}$ are chosen so that they maximize

$$\mathcal{J}_n=\frac{1}{|\mathcal{N}|}\sum_{l\in\mathcal{N}}\mathrm{E}\Big[\exp\big(-\frac{\epsilon_{l,n}\epsilon_{l,n}^{\alpha'-1}}{2\beta^2}\big)\Big], \tag{3}$$

where $\epsilon_{l,n}=y_{l,n}-\hat{y}_{l,n}$, with $\hat{y}_{l,n}$ denoting the estimated filter output, with $\beta>0$ specifying the bandwidth of the kernel. The parameter $\alpha'\in(1,\alpha)$ is a real-valued constant that guarantees a concave shape to the cost function (3). It also ensures that the $(\alpha'-1)$-order error measure at each agent, i.e., $\epsilon_{l,n}\epsilon_{l,n}^{\alpha'-1}$, has finite statistical expectations.

At every time instant n, at node k, the estimate of $\mathbf{w}^*$, i.e., $\mathbf{w}_n$, can be updated in a steepest ascent manner as

$$\mathbf{w}_{n+1}=\mathbf{w}_n+\eta\nabla^{\alpha'-1}\mathcal{J}_n, \tag{4}$$

where $\nabla^{\alpha'-1}$ denotes the $(\alpha'-1)$-order gradient operator and η is the positive real-valued gain. Using the fractional differentials [16]–[18] to evaluate $\nabla^{\alpha'-1}$ and absorbing the extra multiplicative terms into gain η, the adaptation rule for $\mathbf{w}_n$ can be obtained as

$$\mathbf{w}_{n+1}=\mathbf{w}_n+\eta\sum_{l\in\mathcal{N}}g(\epsilon_{l,n})\epsilon_{l,n}\mathbf{x}_{l,n}^{\langle\alpha'-1\rangle}=\frac{1}{|\mathcal{N}|}\sum_{l\in\mathcal{N}}\boldsymbol{\psi}_{l,n+1}, \tag{5}$$

where

$$\boldsymbol{\psi}_{l,n+1}=\mathbf{w}_n+\mu\ g(\epsilon_{l,n})\ \epsilon_{l,n}\ \mathbf{x}_{l,n}^{\langle\alpha'-1\rangle}, \tag{6}$$

is the intermediate estimate of $\mathbf{w}^*$ at node k and time index n. The term $g(\epsilon_{l,n})=\exp(-(\epsilon_{l,n}\epsilon_{l,n}^{\alpha'-1})/(2\beta^2))$, is a function of the fractional-order of the estimation error and $\mu=\eta|\mathcal{N}|$ is the adaptation gain. The average in (5) can be evaluated in a distributed manner using an average consensus filter (ACF) [35], [36]. Operation of the ACF at its mth iteration is given by

$$\mathbf{h}_{k,(m)}=\mathbf{h}_{k,(m-1)}+\sum_{l\in\mathcal{N}_k}a_{lk}\big(\mathbf{h}_{l,(m-1)}-\mathbf{h}_{k,(m-1)}\big), \tag{7}$$

where $\mathbf{h}_{k,m}$ is the estimate of the ACF at node k after m iterations and the combiner coefficients a_{lk} are non-negative and satisfy $\sum_{l\in\mathcal{N}_k}a_{lk}=1$ [35]. If matrix $\mathbf{A}$ with $[\mathbf{A}]_{l,k}=a_{lk}$, is doubly stochastic and satisfies the conditions stated in [35], we then have

$$\lim_{m\to\infty}\mathbf{h}_{k,(m)}=\frac{1}{|\mathcal{N}|}\sum_{l\in\mathcal{N}}\mathbf{h}_{l,(0)}. \tag{8}$$

Hereafter, the ACF at node k after m iterations is depicted in the schematic form as

$$\mathbf{h}_{k,(m)} \leftarrow \boxed{\text{ACF}} \leftarrow \{\forall l \in \mathcal{N} : \mathbf{h}_{l,(0)}\}.$$

The ACF described in (7)-(8) is used to perform the update operation (5) in a distributed manner. The proposed distributed fractional-order correntropy adaptive filter (DFCAF) is summarized in Algorithm 1, where $\mathbf{w}_{k,n}$ denotes the weight vector estimate local to node k.

Algorithm 1: DFCAF

For nodes $k = 1, 2, \cdots, K$:

Estimated Filter Output: $\hat{y}_{k,n} = \mathbf{x}_{k,n}^{\mathrm{T}} \mathbf{w}_{k,n}$,

Error: $\epsilon_{k,n} = y_{k,n} - \hat{y}_{k,n}$,

Local Update:

$$\boldsymbol{\psi}_{k,n+1} = \mathbf{w}_{k,n} + \mu\, g(\epsilon_{k,n})\, \epsilon_{k,n}\, \mathbf{x}_{k,n}^{\langle \alpha'-1\rangle}, \quad (9)$$

Average Consensus Update:

$$\mathbf{w}_{k,n+1} \leftarrow \boxed{\text{ACF}} \leftarrow \{\forall l \in \mathcal{N} : \boldsymbol{\psi}_{l,n+1}\}. \quad (10)$$

Remark 1. Given that at each node k, $g(\epsilon_{k,n}) \in (0, 1]$ and $(\alpha' - 1) \in (0, 1)$, the sharp spikes present both in signal and noise processes are effectively regulated during the local update stage. This circumvents the problem of sharp jitters propagating over the network. Hence, Algorithm 1 always achieves lower steady-state error than the conventional distributed fractional-order adaptive filter (DFAF) [19].

Remark 2. For the case of $\alpha' \to 2$ and the ACF only iterated once, Algorithm 1 operates akin to the distributed maximum correntropy criterion filter (DMCC) [30]. Moreover, for the case that $\beta \to \infty$, the Algorithm 1 reduces to the DFAF.

Remark 3. A more robust update term than that in (9), is attainable if input is normalized with respect to the input vector, which is given by

$$\boldsymbol{\psi}_{k,n+1} = \mathbf{w}_{k,n} + \mu\, g(\epsilon_{k,n})\, \epsilon_{k,n}\, \frac{\mathbf{x}_{k,n}^{\langle \alpha'-1\rangle}}{\|\mathbf{x}_{k,n}\|_{\alpha'}^{\alpha'}}, \quad (11)$$

where $\|\mathbf{x}_{k,n}\|_{\alpha'}^{\alpha'} = \mathbf{x}_{k,n}^{\mathrm{T}} \mathbf{x}_{k,n}^{\langle \alpha'-1\rangle}$.

IV. Convergence Analysis

A. Network-Wide Model

Before proceeding to the network-level analysis, we define the optimal filter coefficient vector $\mathbf{w}_{net}^{\star} = \mathbf{1}_{|\mathcal{N}|} \otimes \mathbf{w}^{\star}$, estimated filter coefficient vector $\mathbf{w}_{net,n} = \mathrm{col}\{\mathbf{w}_{1,n}, \mathbf{w}_{2,n}, \ldots, \mathbf{w}_{|\mathcal{N}|,n}\}$, input data matrix $\mathbf{X}_n = \mathrm{blockdiag}\{\mathbf{x}_{1,n}, \mathbf{x}_{2,n}, \ldots, \mathbf{x}_{|\mathcal{N}|,n}\}$ and observation noise vector $\boldsymbol{\upsilon}_{net,n} = \mathrm{col}\{\upsilon_{1,n}, \upsilon_{2,n}, \ldots, \upsilon_{|\mathcal{N}|,n}\}$, where $\mathrm{col}\{\cdot\}$ and $\mathrm{blockdiag}\{\cdot\}$ denote the columnwise stacking operator and block diagonalization operator, respectively. The symbol $\mathbf{1}_{|\mathcal{N}|}$ is a column vector of size $|\mathcal{N}| \times 1$ with every element taking the value one. Using above definitions, the network-level data model and error vector are given by

$$\begin{aligned} \mathbf{y}_n &= \mathrm{col}\{y_{1,n}, y_{2,n}, \ldots, y_{|\mathcal{N}|,n}\} = \mathbf{X}_n^{\mathrm{T}} \mathbf{w}_{net}^{\star} + \boldsymbol{\upsilon}_n, \\ \boldsymbol{\epsilon}_n &= \mathrm{col}\{\epsilon_{1,n}, \epsilon_{2,n}, \ldots, \epsilon_{|\mathcal{N}|,n}\} = \mathbf{y}_n - \mathbf{X}_n^{\mathrm{T}} \mathbf{w}_{net,n}. \end{aligned} \quad (12)$$

Using the above definitions, from (9) and (10), the global model of the proposed DFCAF can be stated as

$$\mathbf{w}_{net,n+1} = \boldsymbol{\mathcal{A}}\big(\mathbf{w}_{net,n} + \mu\, \mathbf{G}_n^{\epsilon}\, \mathbf{X}_n^{\langle \alpha'-1\rangle}\, \boldsymbol{\epsilon}_n\big), \quad (13)$$

where

$$\begin{aligned} \boldsymbol{\mathcal{A}} &= \mathbf{A}^m \otimes \mathbf{I}, \\ \mathbf{G}_n^{\epsilon} &= \mathrm{diag}\{g(\epsilon_{1,n}), g(\epsilon_{2,n}), \cdots, g(\epsilon_{|\mathcal{N}|,n})\} \otimes \mathbf{I}. \end{aligned} \quad (14)$$

B. Mean Convergence Analysis

Denote the global weight deviation vector of the DFCAF at n-th index as $\widetilde{\mathbf{w}}_{net,n} = \mathbf{w}_{net}^{\star} - \mathbf{w}_{net,n}$, and recall that $\boldsymbol{\mathcal{A}}\mathbf{w}_{net}^{\star} = \mathbf{w}_{net}^{\star}$ (since $\mathbf{A}$ is doubly stochastic); then from (13), $\widetilde{\mathbf{w}}_{net,n}$ can be recursively expressed as

$$\widetilde{\mathbf{w}}_{net,n+1} = \boldsymbol{\mathcal{B}}_n \widetilde{\mathbf{w}}_{net,n} - \mu\, \mathbf{G}_n^{\epsilon}\, \mathbf{X}_n^{\langle \alpha'-1\rangle}\, \boldsymbol{\upsilon}_n, \quad (15)$$

where $\boldsymbol{\mathcal{B}}_n = \boldsymbol{\mathcal{A}}\big(\mathbf{I} - \mu\, \mathbf{G}_n^{\epsilon}\, \boldsymbol{\mathcal{X}}_n\big)$ with $\boldsymbol{\mathcal{X}}_n = \mathbf{X}_n^{\langle \alpha'-1\rangle}\, \mathbf{X}_n^{\mathrm{T}} = \mathrm{diag}(\boldsymbol{\mathcal{X}}_{1,n}, \boldsymbol{\mathcal{X}}_{2,n}, \ldots, \boldsymbol{\mathcal{X}}_{K,n})$, with $\boldsymbol{\mathcal{X}}_{k,n} = \mathbf{x}_{k,n}^{\langle \alpha'-1\rangle}\, \mathbf{x}_{k,n}^{\mathrm{T}}$. In the following, we establish the condition of convergence of the proposed algorithm. For this, we assume the following:

A1: For all $k \in \mathcal{N}$, the input signal vectors $\mathbf{x}_{k,n}$ are assumed to be temporally independent,

A2: The noise process $\upsilon_{k,n}$ is assumed to be zero-mean i.i.d. and independent of all input and output data,

A3: For all $k \in \mathcal{N}$, the quantity $g(\epsilon_{k,n})$ is assumed to be independent of other quantities.

Remark 4. At each node k and time instant n, we always have $0 < g(\epsilon_{k,n}) \le 1$. Furthermore, in worst case scenario, i.e., when $g(\epsilon_{k,n}) = 1$, the proposed algorithm reduces to DFAF [19]. Therefore, **A3** is a reasonable assumption and it does not alter the convergence behavior of the proposed algorithm.

Theorem 1. Assume the data model (12) and the assumptions **A1-3** hold. Then, a sufficient condition for the proposed DFCAF to converge in mean is

$$0 < \mu < \frac{1}{\max\limits_{\forall k \in \mathcal{N}} \{\max\limits_{\forall i} \{\mathrm{E}[g(\epsilon_{k,n})] \lambda_i(\mathrm{E}[\boldsymbol{\mathcal{X}}_{k,n}])\}\}}, \quad (16)$$

where $\lambda_i(\cdot)$ denotes the ith eigenvalue of its argument matrix.

Proof. Taking the statistical expectation $\mathrm{E}[\cdot]$ on both sides of (15) and using the assumptions **A1-3**, we obtain

$$\mathrm{E}[\widetilde{\mathbf{w}}_{net,n+1}] = \mathrm{E}[\boldsymbol{\mathcal{B}}_n]\mathrm{E}[\widetilde{\mathbf{w}}_{net,n}], \quad (17)$$

where $\mathrm{E}[\boldsymbol{\mathcal{B}}_n] = \boldsymbol{\mathcal{A}}\big(\mathbf{I} - \mu\, \mathrm{E}[\mathbf{G}_n^{\epsilon}]\, \mathrm{E}[\boldsymbol{\mathcal{X}}_n]\big)$. A sufficient condition for $\lim_{n\to\infty} \mathrm{E}\big[\widetilde{\mathbf{w}}_{net,n}\big]$ to attain a finite value is that $\|\mathrm{E}[\boldsymbol{\mathcal{B}}_n]\| < 1$ for all n, where $\|\cdot\|$ is any matrix norm. To derive a convergence condition in terms of μ, we use the block maximum norm [37] of the matrix $\mathrm{E}[\boldsymbol{\mathcal{B}}_n]$, i.e., $\|\mathrm{E}[\boldsymbol{\mathcal{B}}_n]\|_{b,\infty}$. From the properties of the block maximum norm, we have

$$\begin{aligned} \|\mathrm{E}[\boldsymbol{\mathcal{B}}_n]\|_{b,\infty} &= \|\boldsymbol{\mathcal{A}}\big(\mathbf{I} - \mu\, \mathrm{E}[\mathbf{G}_n^{\epsilon}]\, \mathrm{E}[\boldsymbol{\mathcal{X}}_n]\big)\|_{b,\infty} \\ &\le \|\boldsymbol{\mathcal{A}}\|_{b,\infty} \|\mathbf{I} - \mu\, \mathrm{E}[\mathbf{G}_n^{\epsilon}]\, \mathrm{E}[\boldsymbol{\mathcal{X}}_n]\|_{b,\infty} \\ &= \|\mathbf{I} - \mu\, \mathrm{E}[\mathbf{G}_n^{\epsilon}]\, \mathrm{E}[\boldsymbol{\mathcal{X}}_n]\|_{b,\infty} \end{aligned} \quad (18)$$

In (18), we used the result $\|\boldsymbol{\mathcal{A}}\|_{b,\infty} = (\|\mathbf{A}^T\|_{\infty})^m = 1$. Using [37, Lemma D. 5], it is seen that $\mathrm{E}\big[\widetilde{\mathbf{w}}_{net,n}\big]$ converges under $\rho\big(\mathbf{I} - \mu\,\mathrm{E}[\mathbf{G}_n^{\epsilon}]\, \mathrm{E}[\boldsymbol{\mathcal{X}}_n]\big) < 1$, or, equivalently,

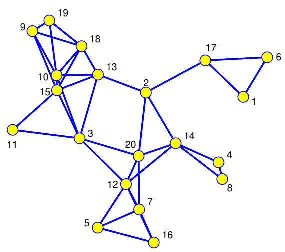
$\forall k, i : |1 - \mu \mathrm{E}[g(\epsilon_{k,n})]\lambda_i(\mathrm{E}[\boldsymbol{\mathcal{X}}_{k,n}])| < 1$, where $\rho(\cdot)$ denotes the spectral radius of the argument matrix. After solving the above convergence condition, we arrive at (16). $\square$

Remark 5. The condition on α' (i.e., $\alpha' \in (1, \alpha)$) ensures the existence of $\mathrm{E}[\boldsymbol{\mathcal{X}}_{k,n}]$, so that the bounds on μ can be evaluated. Upon recalling that $0 < g(\epsilon_{k,n}) \leq 1$, which implies $0 < \mathrm{E}[g(\epsilon_{k,n})] \leq 1$, it can be seen that the convergence conditions of the DFAF are also sufficient for the convergence of the proposed algorithm.

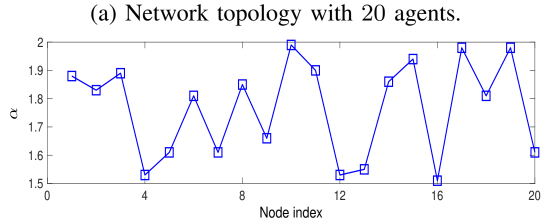
Remark 6. Analysis similar to that in the proof of Theorem 1 shows that the convergence of the normalized version is guaranteed for $0 < \mu < 1$.

V. Numerical Simulations

In this section, we conduct simulations to demonstrate the effectiveness of the proposed DFCAF in the context of system identification. For this, we considered a randomly generated network with the topology shown in Fig. 1(a).



(a) Network topology with 20 agents.



(b) α value at each node

Fig. 1: Network topology and node profile statistics.

The goal is to collaboratively estimate the 16-tap optimal parameter vector $\mathbf{w}^*$ which was generated from a standard Gaussian distribution. In all simulations, the input signal and the observation noise were taken to be SαS signals. The value of the characteristic exponent, α, varied from node to node and its distribution against the node index k is shown in Fig. 1(b). At each node, the parameter α' was taken to be 1.3. The non-negative coefficients in the ACF were obtained through the Metropolis rule [35] and the ACF was iterated for 5 times to approximate the required averages. The adaptation gain μ of the proposed unnormalized and the normalized DFCAF algorithms was set to 0.065 and 0.2, respectively. The kernel bandwidth parameter β for both proposed filtering approaches was fixed at 1.6. The network-level mean absolute deviation (MAD) (given by $\frac{1}{|\mathcal{N}|}\mathrm{E}[\|\widetilde{\mathbf{w}}_{net,n}\|_1]$) was considered as the performance metric. The proposed class of DFCAF algorithms was simulated and the corresponding learning curves, i.e., network-level MAD (in dB) vs iteration index n, obtained by averaging over 500 independent experiments, are plotted in Fig. 2. For comparison, Fig. 2 also includes MAD performance curves of the distributed LMS (DLMS), DMCC [30] and DFAF [19]. To provide a fair comparison, the adaptation gain of these filters was adjusted to match the convergence rate (top graph of Fig. 2) and steady-state MAD (in bottom graph of Fig. 2) of the proposed filters.

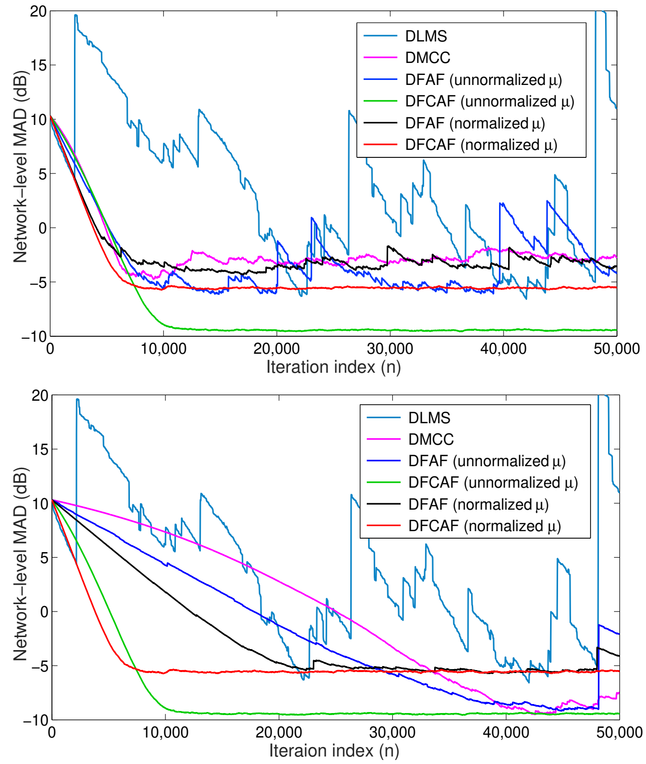


Fig. 2: MAD curves of the proposed distributed filters. MAD curves of the DLMS, DMCC [30] and DFAF [19] are also included for comparison.

From Fig. 2, we see that the proposed distributed filters efficiently identified the system and clearly outperformed the other approaches. In contrast, the DLMS, DMCC and DFAF approaches performed poorly and their steady-state estimates still exhibit the sharp spikes. On the other hand, since the fractional-order correntropy is insensitive to the jittery behavior of the SαS signals, the proposed distributed approaches do not exhibit any sharp spikes in the MAD performance. Also note that when the ACF was iterated for only once, the other approaches exhibited degradation in their MAD performance compared to the case of iterating the ACF five times, however, the proposed class of distributed filters exhibited similar performance.

VI. Conclusions

Distributed adaptive learning in the presence of real-valued SαS signals has been considered, and a novel class of distributed adaptive filters, based on maximizing the fractional-order correntropy criterion, has been derived. The performance of the proposed algorithms has been analyzed and the conditions for their convergence have been established. Simulations results confirmed the superiority of the proposed algorithms over state-of-the-art.

REFERENCES

[1] A. H. Sayed, *Adaptation, Learning, and Optimization over Networks.*, Boston-Delft: NOW, 2014.

[2] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks: cooperation and sensing in networked systems," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99-109, May 2013.

[3] S. Xie and L. Guo, "Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition," *SIAM J. on Control and Optimization*, vol. 56, no. 5, pp. 3404-3431, Sep. 2018.

[4] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460-497, Apr. 2014.

[5] M. J. Piggott and V. Solo, "Diffusion LMS with correlated regressors I: Realization-wise stability," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5473-5484, 1 Nov., 2016

[6] S. Xie and L. Guo, "Analysis of distributed adaptive filters based on diffusion strategies over sensor networks," *IEEE Trans. Auto. Control*, vol. 63, no. 11, pp. 3643-3658, Nov. 2018,

[7] M. Bouvet and S. C. Schwartz, "Comparison of adaptive and robust receivers for signal detection in ambient underwater noise," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 37, no. 5, pp. 621-626, May 1989.

[8] N. Azzaoui and L. Clavier, "Statistical channel model based on α-stable random processes and application to the 60 GHz ultra wide band channel," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1457-1467, May 2010.

[9] J. P. Nolan, "Modeling financial data with stable distributions," *Handbook of heavy tailed distributions in finance*, S. T. Rachev, Ed., vol. 1 of *Handbooks in Finance*, pp. 105-130, 2003.

[10] P. Kidmose, "Alpha-stable distributions in signal processing of audio signals," in *Proc. Conf. Simulat. Model.*, Sep. 2000, pp. 87-94

[11] S. Leglaive, U. Simşekli, A. Liutkus, R. Badeau and G. Richard, "Alpha-stable multichannel audio source separation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Mar. 2017, pp. 576-580.

[12] R. Adler, R. E. Feldman, and M. S. Taqqu, Eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkhauser, Boston, 1998.

[13] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*, Wiley-Interscience, 1995.

[14] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, no. 7, pp. 986-1010, Jul. 1993.

[15] S. P. Talebi and D. P. Mandic, "Distributed particle filtering of α-stable signals," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1862-1866, Dec. 2017.

[16] G. Jumarie, "On the derivative chain-rules in fractional calculus via fractional difference and their application to systems modelling," *Central European Journal of Physics*, vol. 11, no. 6, pp. 617-633, jun. 2013.

[17] G. Jumarie,, "Modified Riemann-Liouville derivative and fractional Taylor series of nondifferentiable functions further results," *Computers & Mathematics with Applications*, vol. 51, no. 9, pp. 1367-1376, 2006.

[18] G. Cottone, M. D. Paola and R. Metzler, "Fractional calculus approach to the statistical characterization of random variables and vectors," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 5, pp. 909-920, 2010.

[19] S. P. Talebi, S. Werner and D. P. Mandic, "Distributed adaptive filtering of α-stable signals," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1450-1454, Oct. 2018.

[20] S. P. Talebi, S. Werner and D. P. Mandic, "Complex-valued nonlinear adaptive filters with applications in α-stable environments," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1315-1319, Sep. 2019.

[21] S. P. Talebi, S. Werner, S. Li and D. P. Mandic, "Tracking dynamic systems in α-stable environments," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, May 2019, pp. 4853-4857.

[22] 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.

[23] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *Proc. Int. Joint Conf. on Neural Networks*, Jun. 2009, pp. 2950-2955.

[24] 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.

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

[26] F. Wang, Y. He, S. Wang, B. Chen, "Maximum total correntropy adaptive filtering against heavy-tailed noises," *Signal Process.*, vol. 141, pp. 84-95, 2017.

[27] K. Xiong, W. Shi and S. Wang, "Robust multikernel maximum correntropy filters," *IEEE Trans.Circuits and Syst. II: Express Briefs*, doi: 10.1109/TCSII.2019.2931409.

[28] B. S. Dees, Y. Xia, S. C. Douglas and D. P. Mandic, "Correntropy-based adaptive filtering of noncircular complex data," *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Apr. 2018, pp. 4339-4343.

[29] V. C. Gogineni and S. Mula, "Improved proportionate-type sparse adaptive filtering under maximum correntropy criterion in impulsive noise environments," *Digital Signal Process.*, vol. 79, pp. 190-198, Aug. 2018.

[30] W. Ma, B. Chen, J. Duan and H. Zhao, "Diffusion maximum correntropy criterion algorithms for robust distributed estimation," *Digital Signal Process.*, vol. 58, pp. 10-19, 2016.

[31] F. Chen, X. Li, S. Duan, L. Wang and J. Wu, "Diffusion generalized maximum correntropy criterion algorithm for distributed estimation over multitask network," *Digital Signal Process.*, vol. 81, pp. 16-25, 2018.

[32] Y. He, F. Wang, S. Wang, P. Ren and B. Chen, "Maximum total correntropy diffusion adaptation over networks with noisy links," *IEEE Trans. Circuits and Syst. II: Express Briefs*, vol. 66, no. 2, pp. 307-311, Feb. 2019.

[33] J. P. Nolan, "Multivariate elliptically contoured stable distributions: Theory and estimation," *Computational Statistics*, vol. 28, no. 5, pp. 2067-2089, Oct. 2013.

[34] V. C. Gogineni, S. P. Talebi and S.Werner, "Correntropy criterion based fractional-order adaptive filters for tracking dynamic systems in α-stable environments," *IEEE Trans. Circuits Syst. II: Express Briefs*, (accepted).

[35] L. Xiao, S. Boyd and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. Int. Conf. Info. Process. in Sensor Networks*, 2005, pp. 63–70.

[36] L. Xiao, Stephen Boyd, "Fast linear iterations for distributed averaging," *Syst. & Control Lett.*, Vol. 53, no. 1, pp. 65-78, 2004.

[37] A. H . Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing,* R. Chellappa and S. Theodoridis, Eds., pp. 322-454, Elsevier, 2013. Also available as arXiv:1205.4220 [cs.MA], May 2012.