Article

Generalized Maximum Complex Correntropy Augmented Adaptive IIR Filtering

Haotian Zheng and Guobing Qian *

College of Electronic and Information Engineering, Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Southwest University, Chongqing 400715, China; zhenghaotian151@foxmail.com
* Correspondence: guobingqian@swu.edu.cn

Abstract: Augmented IIR filter adaptive algorithms have been considered in many studies, which are suitable for proper and improper complex-valued signals. However, lots of augmented IIR filter adaptive algorithms are developed under the mean square error (MSE) criterion. It is an ideal optimality criterion under Gaussian noises but fails to model the behavior of non-Gaussian noise found in practice. Complex correntropy has shown robustness under non-Gaussian noises in the design of adaptive filters as a similarity measure for the complex random variables. In this paper, we propose a new augmented IIR filter adaptive algorithm based on the generalized maximum complex correntropy criterion (GMCCC-AIIR), which employs the complex generalized Gaussian density function as the kernel function. Stability analysis provides the bound of learning rate. Simulation results verify its superiority.

Keywords: GMCCC; complex; augmented IIR; non-Gaussian noise; system identification

1. Introduction

Complex-valued adaptive filtering algorithm has a wide range of engineering applications in radio systems [1], system identification [2], environment signal processing [3], and other fields. Generally speaking, complex-valued adaptive filtering algorithm is an extension of the real-valued adaptive filtering algorithm. When the complex signal is second-order circular (or proper) [3], the performance of adaptive filter is optimal. For second-order circular signals, the covariance matrix $\mathcal{C}_{\mathbf{xx}} = E\{\mathbf{x}(n)\mathbf{x}^H(n)\}$ is second-order statistics. A complex-valued random variable is second-order circular if its first and second-order statistics are rotation-invariant in the complex plane, but in most cases, complex signals are noncircular (or improper) [3].

In order to suit both proper and improper complex-valued signals, augmented complex statistics are proposed. There are quantities of adaptive filtering algorithms based on augmented complex statistics, such as augmented complex least mean square (ACLMS) [4], augmented complex adaptive infinite impulse response (IIR) algorithm (ACA-IIR) [5], diffusion augmented complex adaptive IIR algorithm (DACA-IIR) [6], and incremental augmented complex adaptive IIR algorithm (IACA-IIR) [7]. These adaptive filtering algorithms are based on the mean square error (MSE) criterion, which is mathematically tractable, computationally simple, and optimal under Gaussian assumptions [8]. However, the MSE-based algorithm may perform poorly or encounter instability problems when the signal is disturbed by non-Gaussian noise [9,10]. From a statistical point of view, the mean square error is not sufficient to capture all possible information in a non-Gaussian signal. In practical applications, non-Gaussian noise is common. For example, some sources of non-Gaussian impulse noise are non-synchronization in digital recording, motor ignition noise in internal combustion engines, and lightning spikes in natural phenomena [11,12].

Entropy generally describes a measure of uncertainty of a real random variable, and as a means of a functional analysis method, entropy of a signal can define the noise without

using a threshold criterion [13]. Correntropy is an extension of entropy, which is a quantity of how similar two random complex variables are in a neighborhood of the joint space controlled by the kernel bandwidth. Compared with the MSE-based algorithms, correntropy algorithm is superior. On the whole, the correntropy uses the Gaussian function as the kernel function [14,15], because it is smooth and strictly positive definite. However, the Gaussian kernel is not always an appropriate choice. Recently, He et al. [16] and Chen et al. [17] extended it to more general cases and proposed the generalized maximum correntropy criterion (GMCC) algorithm, which has strong generality and flexibility, and Qian et al. [18] proposed a GMCCC algorithm based on the generalized maximum complex correntropy criterion, which uses the complex generalized Gaussian density(CGGD) function as a kernel of the complex correntropy. It succeeds in the excellent characteristics of GMCC and can deal with complex signals at the same time. These correntropy algorithms are finite impulse response (FIR) wide linear adaptive filtering algorithms, but when FIR filters need to use a large number of coefficients to obtain satisfactory filtering performance, FIR wide linear models may not always be appropriate.

Unlike the FIR counterpart, the memory depth of an IIR filter is independent of the filter order and the number of coefficients [19]. Alternatively, an IIR filter generally requires considerably fewer coefficients than the corresponding FIR filter to achieve a certain level of performance. Thus, the IIR adaptive filters are suitable for systems with memory, such as autoregressive moving average (ARMA) models. Navarro-Moreno et al. [20] developed an ARMA widely linear model with fixed coefficients. To derive a recursive algorithm for augmented complex adaptive IIR filtering, Took et al. [7] proposed the ACA-IIR to learn the parameters of a widely linear ARMA model.

Based on the generalized maximum complex correntropy criterion (GMCCC) and widely linear ARMA model, we propose a GMCCC algorithm variant, namely the GMCCC augmented adaptive IIR filtering algorithm (GMCCC-AIIR). We show that the GMCCC-AIIR is very flexible, with ACA-IIR, GMCCC, and ACLMS as its special cases. Stability analysis shows that GMCCC-AIIR always converges when the step-size satisfies the theoretical bound. Simulation results demonstrate the superiority of the GMCCC-AIIR algorithm.

The organization of the paper is as follows: Section 2 introduces and describes the augmented IIR system. Section 3 defines the generalized complex correntropy, derives the GMCCC-AIIR algorithm, and introduces a reduced-complexity version of the proposed algorithm. Section 4 provides the analysis on the bounds of the step-size for convergence. The superiority of the GMCCC-AIIR algorithm is verified by simulations in Section 4, and the conclusion is drawn in Section 5.

2. Augmented IIR System

The signals used in communications are usually complex circular, whereas the class of signals made complex by convenience of representation become more general, and such signals are often noncircular. For the stochastic modeling of this kind of signal, Picinbono et al. introduce a widely linear moving average (MA) model, which is given by [21]:

$$y(n) = \sum_{m=0}^{N} b_m x(n-m) + \sum_{m=0}^{N} h_m x^*(n-m) \tag{1}$$

where b_m and h_m are filter coefficients. Based on this widely linear model, an ACLMS algorithm was proposed [22].

Since the FIR generalized linear model is not always an optimal choice, Moreno et al. introduce a fixed coefficient ARMA generalized linear model [20].

$$\begin{aligned} y(n) &= \sum_{m=1}^{p} a_m y(n-m) + \sum_{m=0}^{q} b_m x(n-m) + \sum_{m=1}^{p} g_m y^*(n-m) \\ &+ \sum_{m=0}^{q} h_m x^*(n-m) \end{aligned} \tag{2}$$

where a_m, g_m are the coefficients of feedback and its conjugation, p and q are the orders of the AR and MA parts, respectively. The model provides a theoretical basis for the proposed recursive algorithm for training adaptive IIR filters.

To introduce a recursive algorithm of augmented complex adaptive IIR filter, Took et al. [5] give the output of the widely linear IIR filter in the following form:

$$
\begin{aligned}
y(n) &= \sum_{m=1}^{M} a_m(n)y(n-m) + \sum_{m=0}^{N} b_m(n)x(n-m) + \sum_{m=1}^{M} g_m(n)y^*(n-m) \\
&+ \sum_{m=0}^{N} h_m(n)x^*(n-m)
\end{aligned}
\tag{3}
$$

where M is the order of the feedback and N is the length of the input. This model can be simplified as follows:

$$
y(n) = \mathbf{w}^{\mathbf{T}}(n)\mathbf{z}(n)
\tag{4}
$$

where:

$$
\mathbf{w}(n) = [a_1(n),\ldots,a_M(n),g_1(n),\ldots,g_m(n),b_0(n),\ldots,b_M(n),h_0(n),\ldots,h_M(n)]^{T}
\tag{5}
$$

$$
\mathbf{z}(n) = \left[\mathbf{y}^{\mathbf{T}}(n),\mathbf{x}^{\mathbf{T}}(n)\right]^{T}
\tag{6}
$$

3. Generalized Complex Correntropy and GMCCC-AIIR Algorithm

3.1. Generalized Complex Correntropy

For two complex variables $C_1 = X + jY$ and $C_2 = Z + jS$, complex Correntropy is defined as [23]:

$$
(C_1, C_2) = E[\kappa(C_1 - C_2)]
\tag{7}
$$

where X, Y, Z, S are real variables, $\kappa(C_1 - C_2)$ is the kernel function.

For the Gaussian kernel in the complex field [23], the kernel function can be expressed as:

$$
\kappa(C_1 - C_2) = G^C(C_1 - C_2) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right)
\tag{8}
$$

where σ is the kernel width.

In this paper, we employ a CGGD function as the kernel function, and it's corresponding correntropy is named generalized complex correntropy [18]:

$$
\begin{aligned}
\kappa(C_1 - C_2) &= G^{L}_{\alpha,\beta}(C_1 - C_2) \\
&= \frac{\alpha}{\pi\beta\Gamma(1/\alpha)}\exp\left\{-\frac{\left[(C_1 - C_2)(C_1 - C_2)^*\right]^{\alpha}}{\beta^{\alpha}}\right\} \\
&= \gamma_{\alpha,\beta}\exp\left\{-\lambda\left[(C_1 - C_2)(C_1 - C_2)^*\right]^{\alpha}\right\}
\end{aligned}
\tag{9}
$$

where α is the shape parameter, $\beta = 2\sigma^2\Gamma(1/\alpha)/\Gamma(2/\alpha)$ is the kernel width, $\lambda = 1/\beta^{\alpha}$, $\gamma_{\alpha,\beta} = \frac{\alpha}{\pi\beta\Gamma(1/\alpha)}$.

In this way, generalized complex correntropy can be written as:

$$
V^C_{\alpha,\beta}(C_1, C_2) = E\left[G^C_{\alpha,\beta}(C_1 - C_2)\right]
\tag{10}
$$

The samples $\{(c_{1i}, c_{2i})\}_{i=1}^{N}$ are finite in reality, so we estimate the generalized complex correntropy by sample mean.

$$\widehat{V_{\alpha,\beta}^C}(C_1, C_2) = \frac{1}{N}\sum_{i=1}^N G_{\alpha,\beta}^C(c_{1i} - c_{2i})$$

$$= \frac{1}{N}\sum_{i=1}^N G_{\alpha,\beta}^C(e_i) \tag{11}$$

where $e_i = c_{1i} - c_{2i}$.

Instead of the correntropy in data analysis, the correntropic loss is often used, so we define the generalized complex correntropic loss as:

$$J_{GC-loss}^C(C_1, C_2) = G_{\alpha,\beta}^C(0) - V_{\alpha,\beta}^C(C_1, C_2) \tag{12}$$

Then, when the sample is finite, the generalized complex correntropic can be expressed as:

$$\widehat{J_{GC-loss}^C}(C_1, C_2) = \gamma_{\alpha,\beta} - \frac{1}{N}\sum_{i=1}^N G_{\alpha,\beta}^C(c_{1i} - c_{2i})$$

$$= \gamma_{\alpha,\beta} - \frac{1}{N}\sum_{i=1}^N G_{\alpha,\beta}^C(e_i) \tag{13}$$

There are some properties of generalized complex correntropy [18].

Property 1. $V_{\alpha,\beta}^C(C_1, C_2)$ *is symmetric, i.e.,* $V_{\alpha,\beta}^C(C_1, C_2) = V_{\alpha,\beta}^C(C_2, C_1)$

Property 2. $V_{\alpha,\beta}^C(C_1, C_2)$ *is bounded with* $0 \leq V_{\alpha,\beta}^C(C_1, C_2) \leq \gamma_{\alpha,\beta}$ *and achieves its maximum when* $C_1 = C_2$

We can get $J_{GC-loss}^C(C_1, C_2)$, it is symmetric and achieves its minimum when $C_1 = C_2$ on the basis of Properties 1 and 2.

Property 3. *Given* $\mathbf{e} = [\mathbf{e_1, e_2} \ldots \mathbf{e_N}]^{\mathbf{T}}$, *the following conclusions about* $\widehat{J_{GC-loss}^C}$ *are true:*

- *When* $\alpha \geq 1/2$, $\widehat{J_{GC-loss}^C}$ *is convex at any* $\mathbf{e}$ *with* $|e_n| \leq [2a - 1/2a\lambda]^{1/2a}$;
- *When* $0 < \alpha < 1/2$, $\widehat{J_{GC-loss}^C}$ *is non-convex at any* $\mathbf{e}$ *with* $|e_n| \neq 0$.

3.2. GMCCC-AIIR Algorithm

Based on properties of generalized complex correntropy, we define the cost function of the GMCCC-AIIR algorithm as:

$$J_{GC-loss}^C = G_{\alpha,\beta}^C(0) - E\left[G_{\alpha,\beta}^C(e(n))\right]$$

$$= \gamma_{\alpha,\beta}\left\{1 - E\left[\exp\left[-\lambda(e(n)e^*(n))^\alpha\right]\right]\right\} \tag{14}$$

where $e(n) = d(n) - y(n)$.

We can infer from (4) that $e(n) = d(n) - \mathbf{w}^{\mathbf{T}}(n)\mathbf{z}(n)$, so $e^*(n) = d^*(n) - \mathbf{w}^{\mathbf{H}}(n)\mathbf{z}^*(n)$. Then, we search for the optimal solution by stochastic gradient descent method, i.e.,

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \eta\nabla_{\mathbf{w}}\left\{\left[1 - \exp\left[-\lambda(e(n)e^*(n))^\alpha\right]\right]\right\}$$

$$= \mathbf{w}(n) + \eta\alpha\lambda\exp\left[-\lambda|e(n)|^{2\alpha}\right]|e(n)|^{2\alpha-2} \times \nabla_{\mathbf{w}}[e(n)e^*(n)] \tag{15}$$

where the option of the kernel bandwidth alpha is according to *property 3*, so that the cost function is convex and the result of the stochastic gradient descent method is the global optimal solution rather than the local optimal one.

The gradients can be computed as: [24]

$$
\begin{aligned}
\nabla_{\mathbf{w}}[e(n)e^*(n)] &= -\left[e(n)\frac{\partial y^*(n)}{\partial \mathbf{w}(n)} + \frac{\partial y(n)}{\partial \mathbf{w}(n)}e^*(n)\right] \\
&= -[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)]
\end{aligned}
\tag{16}
$$

where:

$$
\mathbf{\Phi_w}(n) = \left(\frac{\partial y^*(n)}{\partial \mathbf{w}_R(n)} + j\frac{\partial y^*(n)}{\partial \mathbf{w}_I(n)}\right)
\tag{17}
$$

$$
\mathbf{\Psi_w}(n) = \left(\frac{\partial y(n)}{\partial \mathbf{w}_R(n)} + j\frac{\partial y(n)}{\partial \mathbf{w}_I(n)}\right)
\tag{18}
$$

The gradient vectors (17) and (18) can be written as:

$$
\mathbf{\Phi_w}(n) = \left[\Phi_{a_1}(n),\ldots,\Phi_{a_M}(n),\Phi_{g_1}(n),\ldots,\Phi_{g_M}(n),\ \Phi_{b_0}(n),\ldots,\Phi_{b_N}(n),\Phi_{h_0}(n),\ldots,\Phi_{h_N}(n)\right]^T
\tag{19}
$$

$$
\mathbf{\Psi_w}(n) = \left[\Psi_{a_1}(n),\ldots,\Psi_{a_M}(n),\Psi_{g_1}(n),\ldots,\Psi_{g_M}(n),\ \Psi_{b_0}(n),\ldots,\Psi_{b_N}(n),\Psi_{h_0}(n),\ldots,\Psi_{h_N}(n)\right]^T
\tag{20}
$$

where **R** and **I** are the real and the imaginary part of complex quantities respectively, and $j = \sqrt{-1}$. To calculate the gradient in (16), items in (17), (18) must be calculated separately, such as:

$$
\frac{\partial y^*(n)}{\partial a_{m_R}(n)} = y^*(n-m) + \sum_{l=1}^{M} a_l^*(n)\frac{\partial y^*(n-l)}{\partial a_{m_R}(n)} + \sum_{l=1}^{M} g_l^*(n)\frac{\partial y(n-l)}{\partial a_{m_R}(n)}
\tag{21}
$$

$$
\frac{\partial y^*(n)}{\partial a_{m_I}(n)} = -jy^*(n-m) + \sum_{l=1}^{M} a_l^*(n)\frac{\partial y^*(n-l)}{\partial a_{m_I}(n)} + \sum_{l=1}^{M} g_l^*(n)\frac{\partial y(n-l)}{\partial a_{m_I}(n)}
\tag{22}
$$

The feedback in the IIR system leads to the recursions on the right side of (21) and (22). These are the derivatives of the past values to present weights, which are impossible to compute. To avoid this problem, for a small step-size, we can approximate that:

$$
\mathbf{w}(n) \approx \mathbf{w}(n-1) \approx \cdots \approx \mathbf{w}(n-\tau) \quad \tau = \max\{M, N+1\}
\tag{23}
$$

Thus, the gradient $\mathbf{\Phi_w}(n)$ can be written in the following forms,

$$
\Phi_{a_m}(n) = y^*(n-m) + \sum_{l=1}^{M} a_l^*(n)\Phi_{a_m}(n-l) + \sum_{l=1}^{M} g_l^*(n)\Psi_{a_m}(n-l)
\tag{24}
$$

$$
\Phi_{b_m}(n) = x^*(n-m) + \sum_{l=1}^{M} a_l^*(n)\Phi_{b_m}(n-l) + \sum_{l=1}^{M} g_l^*(n)\Psi_{b_m}(n-l)
\tag{25}
$$

$$
\Phi_{g_m}(n) = y(n-m) + \sum_{l=1}^{M} a_l^*(n)\Phi_{g_m}(n-l) + \sum_{l=1}^{M} g_l^*(n)\Psi_{g_m}(n-l)
\tag{26}
$$

$$
\Phi_{h_m}(n) = x(n-m) + \sum_{l=1}^{M} a_l^*(n)\Phi_{h_m}(n-l) + \sum_{l=1}^{M} g_l^*(n)\Psi_{h_m}(n-l)
\tag{27}
$$

and for the gradient vector $\mathbf{\Psi}_w(n)$ in (18), similarly, we have,

$$\Psi_{a_m}(n) = \sum_{l=1}^{M} a_l(n)\Psi_{a_m}(n-l) + \sum_{l=1}^{M} g_l(n)\Phi_{a_m}(n-l) \tag{28}$$

$$\Psi_{b_m}(n) = \sum_{l=1}^{M} a_l(n)\Psi_{b_m}(n-l) + \sum_{l=1}^{M} g_l(n)\Phi_{b_m}(n-l) \tag{29}$$

$$\Psi_{g_m}(n) = \sum_{l=1}^{M} a_l(n)\Psi_{g_m}(n-l) + \sum_{l=1}^{M} g_l(n)\Phi_{g_m}(n-l) \tag{30}$$

$$\Psi_{h_m}(n) = \sum_{l=1}^{M} a_l(n)\Psi_{h_m}(n-l) + \sum_{l=1}^{M} g_l(n)\Phi_{h_m}(n-l) \tag{31}$$

So the GMCCC-AIIR can be expressed in the form as:

$$
\begin{aligned}
\mathbf{w}(n+1) &= \mathbf{w}(n) - \mu\alpha\lambda\exp\left[-\lambda|e(n)|^{2\alpha}\right]|e(n)|^{2\alpha-2}[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)] \\
&= \mathbf{w}(n) + \mu\exp\left[-\lambda|e(n)|^{2\alpha}\right]|e(n)|^{2\alpha-2}[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)]
\end{aligned}
\tag{32}
$$

3.3. GMCCC-AIIR as a Generalization of ACA-IIR and ACLMS

When $\lambda \to 0^+$ and $\alpha = 1$ degenerate to:

$$\mathbf{w}(n+1) = w(n) + \mu[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)] \tag{33}$$

i.e., the classical ACA-IIR algorithm. On this basis, when feedback within the GMCCC-AIIR is cancelled, that is, the partial derivatives on the right-hand side of (25), (27), (29) and (31) vanish for the widely linear FIR filter, yielding:

$$\Phi_{b_m}(n) = x^*(n-m) \qquad m = 0,...,N \tag{34}$$

$$\Phi_{h_m}(n) = x(n-m) \qquad m = 0,...,N \tag{35}$$

$$\Psi_{b_m}(n) = \Psi_{g_m}(n) = 0 \qquad m = 0,...,N \tag{36}$$

As desired, the GMCCC-AIIR algorithm (32) now simplifies into the ACLMS algorithm for FIR filters, given by [22]:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu e(n)x(n) \tag{37}$$

Reduce the Computational Complexity of AGMCCC-IIR

The weight update of AGMCCC-IIR has a large amount of calculation, and it requires $4 \times (M+N+1)$ recursions for the sensitivities $\mathbf{\Phi_w}(n)$ and $\mathbf{\Psi_w}(n)$. However, this can be simplified to updating only eight sensitivities by the approximation (23). For example,

$$\mathbf{\Phi_a}(n) = \left[\Phi_{a_1}(n), \Phi_{a_2}(n),..., \Phi_{a_M}(n)\right]^T \tag{38}$$

$$\Phi_{a_2}(n) = y^*(n-2) + \sum_{l=1}^{M} a_l^*(n)\Phi_{a_2}(n-l) + \sum_{l=1}^{M} g_l^*(n)\Psi_{a_2}(n-l) \tag{39}$$

Further, we define $\mathbf{\Phi_a^F}(n)$ as follows:

$$\mathbf{\Phi_a^F}(n) = \left[\Phi_{a_1}^F(n), \Phi_{a_2}^F(n), \ldots, \Phi_{a_M}^F(n) \right]^T \tag{40}$$

$$
\begin{aligned}
\Phi_{a_2}^F(n) &= \Phi_{a_1}(n-1) \\
&= y^*(n-1-1) + \sum_{l=1}^M a_l^*(n-1)\Phi_{a_1}(n-l-1) \\
&\quad + \sum_{l=1}^M g_l^*(n-1)\Psi_{a_2}(n-l-1) \\
&= y^*(n-2) + \sum_{l=1}^M a_l^*(n-1)\Phi_{a_2}^F(n-l) + \sum_{l=1}^M g_l^*(n-1)\Psi_{a_2}^F(n-l) \tag{41}
\end{aligned}
$$

For a small step-size, $a_l^*(n-1)$ and $g_l^*(n-1)$ approximate to $a_l^*(n)$ and $g_l^*(n)$, hence $\Phi_{a_2}^F(n) \approx \Phi_{a_2}(n)$. $\mathbf{\Phi_a}(n)$ can be approximated, i.e.,

$$
\begin{aligned}
\mathbf{\Phi_a}(n) &= \left[\Phi_{a_1}^F(n), \Phi_{a_2}^F(n), \Phi_{a_3}^F(n) \ldots, \Phi_{a_M}^F(n) \right]^T \\
&= \left[\Phi_{a_1}(n), \Phi_{a_1}(n-1), \Phi_{a_2}(n-1), \ldots, \Phi_{a_{M-1}}(n-1) \right]^T \tag{42} \\
&= \left[\Phi_{a_1}(n), \Phi_{a_1}(n-1), \Phi_{a_1}(n-2), \ldots, \Phi_{a_1}(n-M+1) \right]^T
\end{aligned}
$$

We only need to update $\Phi_{a_1}(n)$ for the sensitivity $\mathbf{\Phi_a^F}(n)$. This approximation also applies for all other sensitivities.

4. Convergence of AGMCCC-IIR

For convenience, we write the algorithm (32) as:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu f(e(n))[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)] \tag{43}$$

$\mathbf{w}_0$ is defined as the unknown system parameter, and $\widetilde{\mathbf{w}}(n) = \mathbf{w}_0 - \mathbf{w}(n)$. In this way,

$$\widetilde{\mathbf{w}}(n+1) = \widetilde{\mathbf{w}}(n) - \mu f(e(n))[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)] \tag{44}$$

Thus,

$$
\begin{aligned}
E\left\{ \|\widetilde{\mathbf{w}}(n+1)\|^2 \right\} &= E\left\{ \|\widetilde{\mathbf{w}}(n)\|^2 \right\} - 2\mu E\left\{ \mathrm{Re}[\widetilde{\mathbf{w}}(n)f(e(n))[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)]] \right\} \tag{45} \\
&\quad + \mu^2 E\left\{ \|[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)]\|^2 |f(e(n))|^2 \right\}
\end{aligned}
$$

We know that the step-size μ is a small positive constant. If the system converges when $n \to \infty$, we can approximate $E\left\{ \|\widetilde{\mathbf{w}}(n+1)\|^2 \right\} \approx E\left\{ \|\widetilde{\mathbf{w}}(n)\|^2 \right\}$. It can be inferred that:

$$0 < \mu < \frac{2E\{\mathrm{Re}[\mathbf{w}(n)f(e(n))[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)]]\}}{E\left\{ \|[e(n)\mathbf{\Phi_w}(n) + \mathbf{\Psi_w}(n)e^*(n)]\|^2 |f(e(n))|^2 \right\}} \tag{46}$$

5. Simulation

In this section, we present simulation results to confirm the theoretical conclusions drawn in previous sections. We demonstrate the superiority of the GMCCC-AIIR algorithm compared with the ACA-IIR algorithm in non-Gaussian noise. All the system parameters, signal noise, and input signals are complex valued. The unknown augmented IIR system is given by:

$$
\begin{aligned}
y(0) &= 0 \\
y(n) &= 0.125jy(n-1) + 0.25jy(n-2) + (-0.3+0.7j)x(n) + (0.5-0.8j)x(n-1) \\
&\quad + (0.2+0.5j)x(n-2) + 0.25jy^*(n-1) + 0.21jy^*(n-2) + (0.32+0.21j)x^*(n) \\
&\quad + (-0.3+0.7j)x^*(n-1) + (0.5-0.8j)x^*(n-2)
\end{aligned}
\tag{47}
$$

The real part and the imaginary part of the input signal x are Gaussian distributed, with zero mean and unit variance. One hundred Monte Carlo simulations were ran. To evaluate estimation accuracy, the mean square deviation (MSD) is defined by $\mathrm{MSD} = ||w_0 - w(n)||^2$.

5.1. Complex Non-Gaussian Noise Models

Unlike Gaussian noise, non-Gaussian noise is a random process which the probability distribution function (pdf) of non-Gaussian noise does not satisfy the normal distribution. Generally speaking, the non-Gaussian noise distributions can be divided into two categories: light-tailed (e.g., binary, uniform, etc.) and heavy-tailed (e.g., Cauchy, mixed Gaussian, alphastable, etc.). In the following experiments, four common non-Gaussian noise models, including Cauchy noise, mixed Gaussian noise, alpha-stable noise, and Student's t noise, are selected for performance evaluation, and the additive complex noise can be written as: $v = v_{re} + jv_{im}$, where v_{re} and v_{im} are obedient to different distributions in different non-Gaussian noise models. The descriptions of these non-Gaussian noise are the following.

5.1.1. Mixed Gaussian Noise

The mixed Gaussian noise model is given by [25]:

$$
(1-\theta)\mathcal{N}\left(\lambda_1, v_1^2\right) + \theta\mathcal{N}\left(\lambda_2, v_2^2\right)
\tag{48}
$$

where $\mathcal{N}\left(\lambda_i, v_i^2\right)(i=1,2)$ denotes the Gaussian distributions with mean values λ_i and variances v_i^2, and θ is the mixture parameter. Usually, one can set θ to a small value and $v_2^2 \gg v_1^2$ to represent the impulsive noise. Thus, we define the mixed Gaussian noise parameter vector as $V_{\mathrm{mix}} = \left(\lambda_1, \lambda_2, v_1^2, v_2^2, \theta\right)$.

5.1.2. Alpha-Stable Noise

The alpha-stable distribution is often used to model the probability distribution of heavy-tailed impulse noise. It is a more generalized Gaussian distribution, or Gaussian distribution is a special case of alpha-stable distribution. It is compatible with many signals in practice, such as noise in telephone lines, atmospheric noise, and backscattering echos in radar systems; even the modeling of economic time series is very successful. The characteristic function of the alpha-stable noise is defined as [26,27]:

$$
\psi(t) = \exp\{j\delta t - \gamma|t|^\alpha[1 + j\beta\,\mathrm{sgn}(t)S(t,\alpha)]\} \quad (t \neq 0)
\tag{49}
$$

in which:

$$
S(t,\alpha) = \begin{cases} \tan(\alpha\pi/2) & \text{if } \alpha \neq 1 \\ 2/\pi\log|t| & \text{if } \alpha = 1 \end{cases}
\tag{50}
$$

From (49) and (50), one can observe that a stable distribution is completely determined by four parameters: (1) the characteristic factor α; (2) the symmetry parameter β; (3) the dispersion parameter γ; and (4) the location parameter δ. Both v_{re} and v_{im} obey the alpha-stable distribution, so we define the alpha-stable noise parameter vector as $V_{\text{alpha}} = (\alpha,\ \beta,\ \gamma,\ \delta)$.

It is worth mentioning that, in the case of $\alpha = 2$, the alpha-stable distribution coincides with the Gaussian distribution, while $\alpha = 1, \delta = 0$ is the same as the Cauchy distribution.

5.1.3. Cauchy Noise

The PDF of the Cauchy noise is [28]:

$$p(v) = \frac{1}{\pi(1 + v^2)} \tag{51}$$

5.1.4. Student's T Noise

The PDF of the Student's t noise is [29]:

$$p(v, n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{v^2}{n}\right)^{\frac{n+1}{2}}, -\infty < v < +\infty \tag{52}$$

where n is the degree of freedom, $\Gamma(\cdot)$ denotes the Gamma function.

5.2. *Augmented Linear System Identification*

Figure 1 is the block diagram of the system identification, and the length of the adaptive filter is equal to the unknown system impulse response.
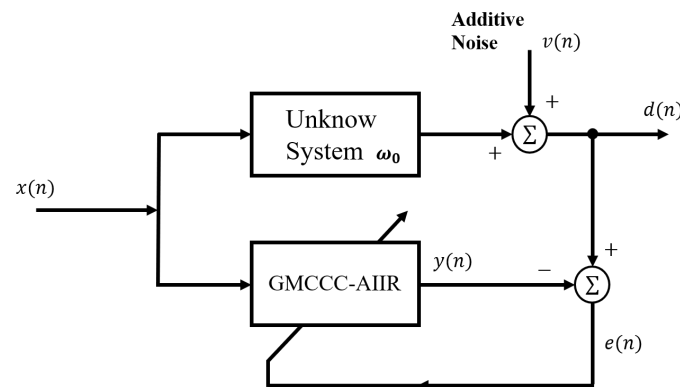

Figure 1. System Identification Configuration.

First, we demonstrate how kernel bandwidth α affects the convergence performance of GMCCC-AIIR. Figure 2 shows the convergence curves of GMCCC-AIIR with different α, in which the noise chooses mixed Gaussian noise and $\mu = 0.013, \lambda = 0.3$. Obviously, the choice of kernel bandwidth has a significant effect on the convergence. In this example, the convergence performance and convergence speed of the proposed algorithm get better when α decreases. Generally speaking, small bandwidth is more robust to impulse noise without considering the convergence rate, and the performance of the algorithm is optimal when $\alpha = 1$.
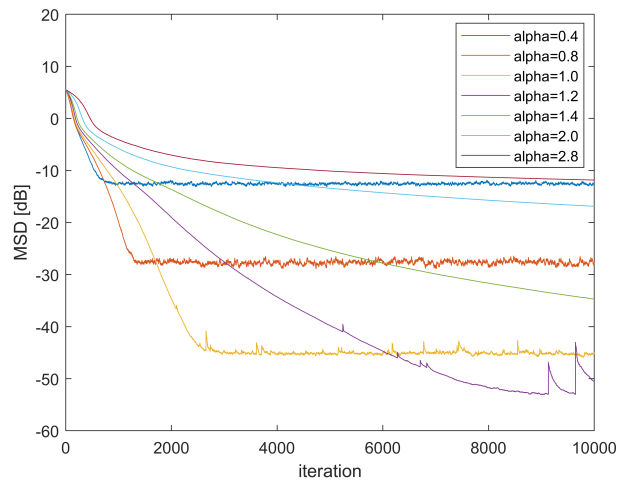
Figure 2. Learning Curve in different α.

Second, the stability of GMCCC-AIIR at different step sizes is investigated. Figure 3 shows the convergence performance for different step sizes. The noise is still the mixed Gaussian noise and $\alpha = 1, \lambda = 0.3$. The simulation results show that when the step size is large, such as $\mu = 0.025$, the convergence performance gets worse, and the GMCCC-AIIR will diverge if step size continues to increase, which confirms the correctness of the stability analysis in Section 4.
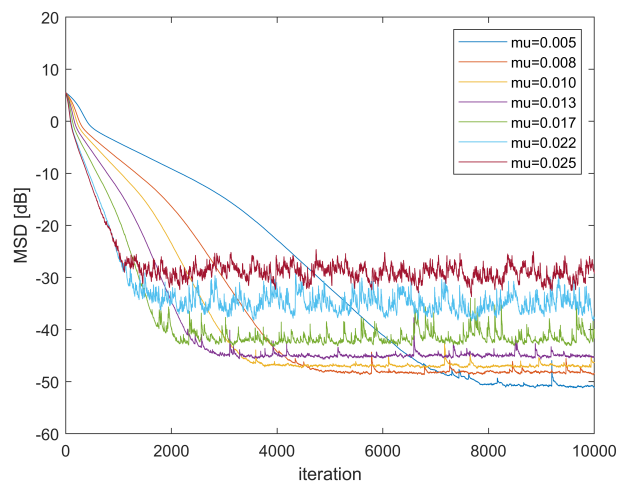
Figure 3. Learning Curve in different μ.

Third, we introduce how the parameter λ will affect the performance of the algorithm. Figure 3 shows the learning curve of GMCCC-AIIR with different λ. The noise is mixed Gaussian noise and $\alpha = 1, \mu = 0.013$. We can see from Figure 4 that when the parameter λ increases, the convergence speed will slow down. However, when λ is too small, the GMCCC-AIIR algorithm is approximate to the ACA-IIR, and the convergence performance is poor in the non-Gaussian noise model. Therefore, we should choose the appropriate parameter λ according to different situations.
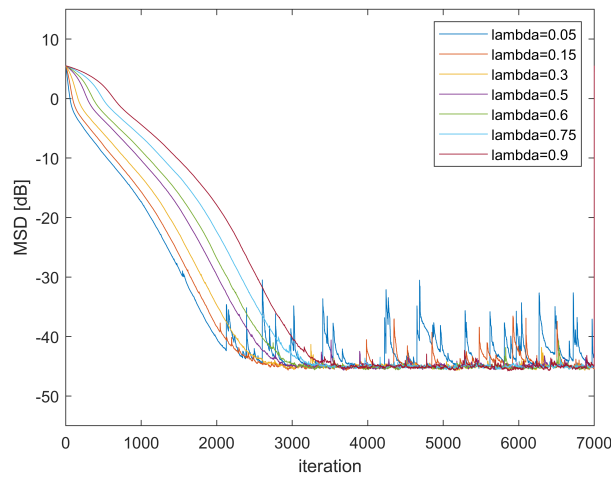
Figure 4. Learning Curve in different λ.

Fourth, we introduce how pairwise parameters α, μ, and λ affect the steady-state excess MSD (EMSD) and give 3D diagrams of EMSD. The number of iterations is increased to 15,000 to ensure the convergence of the algorithm, and the additive noise is still mixed Gaussian noise. EMSD equals the average MSD of the last 1000 iterations. Figure 5 shows that EMSD of GMCCC-AIIR mainly depends on α and μ. The performance of the algorithm get worse when μ increases and α approaches 1, the algorithm performs best. Combined with Figures 4 and 5b,c, λ mainly affects the convergence speed of the GMCCC. When λ approaches 0, the robustness of the algorithm to non-Gaussian noise gets worse, the outliers of EMSD increase, and the algorithm may even diverge.
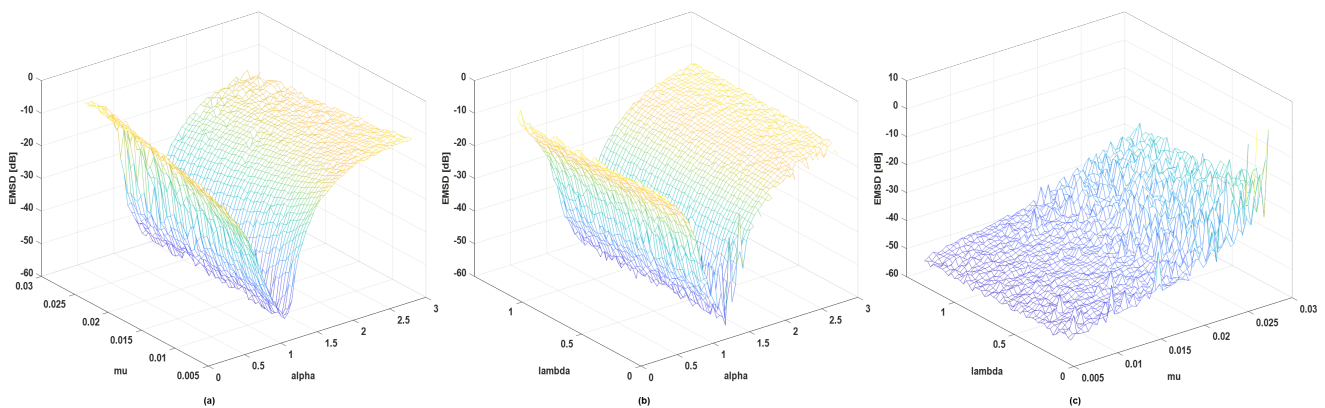

Figure 5. EMSD with different pairwise parameters (**a**) α and μ ($\lambda = 0.3$) (**b**) α and λ ($\mu = 0.013$) (**c**) μ and λ ($\alpha = 1$).

Fifth, we compare the performance of GMCCC-AIIR and ACA-IIR under four noise distributions. In the simulation, the mixed Gaussian noise and alpha-stable noise parameters are set separately at $V_{\mathrm{mix}} = (0, 0, 0.01, 100, 0.03)$ and $V_{\mathrm{alpha}} = (1.4, 0, 0.3, 0)$, the freedom parameter of student noise n is set to 2, and the Cauchy noise is the standard form. The step-sizes are chosen such that both algorithms have almost the same initial convergence speed. The simulation results are shown in Figure 6. As expected, the proposed GMCCC-AIIR algorithm can achieve better steady-state performance than ACA-IIR significantly in these non-Gaussian noise models. The ACA-IIR algorithm diverges after encountering impulse and the MSD approaches infinity at this time. The convergence process cannot be well observed when trying to display the ACA-IIR and the GMCCC-AIIR learning curve in the same coordinate system. Therefore, we limit the height of all simulation results.
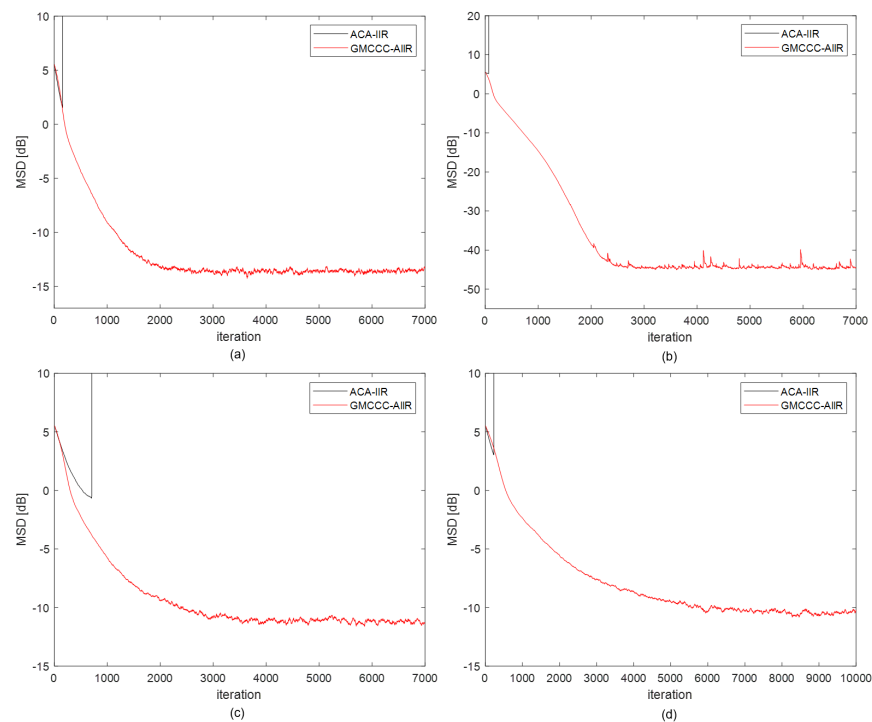
Figure 6. Learning Curve with different noise: (**a**) alpha-stable noise, (**b**) mixed Gaussian noise, (**c**) Cauchy noise, (**d**) Student's t noise.

6. Conclusions

In this paper, we propose an adaptive algorithm for augmented IIR filter based on generalized maximum complex correlation entropy criterion. We study the convergence performance, providing the bound for the step size. Moreover, computational complexity is reduced by making use of the redundancy in the state vector of the filter. We also prove that ACA-IIR and ACLMS are special cases of GMCCC-AIIR. The simulation results verify the theoretical conclusion and show how parameters affect the convergence performance of GMCCC-AIIR and superiority of the GMCCC-IIR algorithm compared with the MSE-based algorithm ACA-IIR when the noise is non-Gaussian distribution.

Author Contributions: Conceptualization, H.Z. and G.Q.; methodology, H.Z.; software, H.Z. and G.Q.; validation, G.Q.; formal analysis, H.Z.; investigation, G.Q.; resources, H.Z. and G.Q.; data curation, H.Z.; writing—original draft preparation, H.Z; writing—review and editing, G.Q.; visualization, H.Z. and G.Q.; supervision, H.Z. and G.Q.; project administration, G.Q.; funding acquisition, G.Q. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pascual Campo, P.; Anttila, L.; Korpi, D.; Valkama, M. Cascaded Spline-Based Models for Complex Nonlinear Systems: Methods and Applications. *IEEE Trans. Signal Process.* **2021**, *69*, 370–384. [CrossRef]
2. Huang, F.; Zhang, J.; Zhang, S. Complex-valued proportionate affine projection Versoria algorithms and their combined-step-size variants for sparse system identification under impulsive noises. *Digit. Signal Process.* **2021**, *118*, 103209. [CrossRef]

3. Mandic, D.P.; Goh, V.S.L. Complex valued nonlinear adaptive filters: Noncircularity, widely linear and neural models. In *Adaptive and Learning Systems for Signal Processing, Communications, and Control*; Wiley: Chichester, UK, 2009.

4. Javidi, S.; Pedzisz, M.; Goh, S.L.; Mandic, D. *The Augmented Complex Least Mean Square Algorithm with Application to Adaptive Prediction Problems*; Citeseer: Princeton, NJ, USA, 2008; p. 4.

5. Took, C.C.; Mandic, D.P. Adaptive IIR Filtering of Noncircular Complex Signals. *IEEE Trans. Signal Process.* **2009**, *57*, 4111–4118. [CrossRef]

6. Khalili, A. Diffusion augmented complex adaptive IIR algorithm for training widely linear ARMA models. *Signal Image Video Process.* **2018**, *12*, 1079–1086. [CrossRef]

7. Khalili, A.; Rastegarnia, A.; Bazzi, W.M.; Rahmati, R.G. Incremental augmented complex adaptive IIR algorithm for training widely linear ARMA model. *Signal Image Video Process.* **2017**, *11*, 493–500. [CrossRef]

8. Sayed, A.H. *Fundamentals of Adaptive Filtering*; IEEE Press Wiley-Interscience: New York, NY, USA, 2003.

9. Chen, B.; Zhu, Y.; Hu, J.; Príncipe, J.C. *System Parameter Identification: Information Criteria and Algorithms*, 1st ed.; Elsevier: London, UK; Waltham, MA, USA, 2013.

10. Principe, J.C.; Xu, D.; Iii, J.W.F. *Information-Theoretic Learning*; Springer: Berlin/Heidelberg, Germany, 2010; p. 62.

11. Plataniotis, K.N.; Androutsos, D.; Venetsanopoulos, A.N. Nonlinear Filtering of Non-Gaussian Noise. *J. Intell. Robot. Syst.* **1997**, *19*, 207–231. [CrossRef]

12. Weng, B.; Barner, K. Nonlinear system identification in impulsive environments. *IEEE Trans. Signal Process.* **2005**, *53*, 2588–2594. [CrossRef]

13. Schimmack, M.; Mercorelli, P. An on-line orthogonal wavelet denoising algorithm for high-resolution surface scans. *J. Frankl. Inst.* **2018**, *355*, 9245–9270. [CrossRef]

14. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and Applications in Non-Gaussian Signal Processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [CrossRef]

15. Liu, X.; Chen, B.; Zhao, H.; Qin, J.; Cao, J. Maximum Correntropy Kalman Filter With State Constraints. *IEEE Access* **2017**, *5*, 25846–25853. [CrossRef]

16. He, Y.; Wang, F.; Yang, J.; Rong, H.; Chen, B. Kernel adaptive filtering under generalized Maximum Correntropy Criterion. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 1738–1745. [CrossRef]

17. Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Prı́ncipe, J.C. Generalized Correntropy for Robust Adaptive Filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [CrossRef]

18. Qian, G.; Wang, S. Generalized Complex Correntropy: Application to Adaptive Filtering of Complex Data. *IEEE Access* **2018**, *6*, 19113–19120. [CrossRef]

19. Li, K.; Principe, J.C. Functional Bayesian Filter. *IEEE Trans. Signal Process.* **2022**, *70*, 57–71. [CrossRef]

20. Navarro-Moreno, J. ARMA Prediction of Widely Linear Systems by Using the Innovations Algorithm. *IEEE Trans. Signal Process.* **2008**, *56*, 3061–3068. [CrossRef]

21. Picinbono, B.; Chevalier, P. Widely linear estimation with complex data. *IEEE Trans. Signal Process.* **1995**, *43*, 2030–2033. [CrossRef]

22. Mandic, D.; Javidi, S.; Goh, S.; Kuh, A.; Aihara, K. Complex-valued prediction of wind profile using augmented complex statistics. *Renew. Energy* **2009**, *34*, 196–201. [CrossRef]

23. Guimarães, J.P.F.; Fontes, A.I.R.; Rego, J.B.A.; de M. Martins, A.; Príncipe, J.C. Complex Correntropy: Probabilistic Interpretation and Application to Complex-Valued Data. *IEEE Signal Process. Lett.* **2017**, *24*, 42–45. [CrossRef]

24. Shynk, J. A complex adaptive algorithm for IIR filtering. *IEEE Trans. Acoust. Speech Signal Process.* **1986**, *34*, 1342–1344. [CrossRef]

25. Zhao, S.; Chen, B.; Príncipe, J.C. Kernel adaptive filtering with maximum correntropy criterion. In Proceedings of the 2011 International Joint Conference on Neural Networks, San Jose, CA, USA, 31 July–5 August 2011; pp. 2012–2017. [CrossRef]

26. Zhang, J.; Qiu, T.; Song, A.; Tang, H. A novel correntropy based DOA estimation algorithm in impulsive noise environments. *Signal Process.* **2014**, *104*, 346–357. [CrossRef]

27. Wu, Z.; Peng, S.; Chen, B.; Zhao, H. Robust Hammerstein Adaptive Filtering under Maximum Correntropy Criterion. *Entropy* **2015**, *17*, 7149–7166. [CrossRef]

28. Chen, B.; Xing, L.; Liang, J.; Zheng, N.; Príncipe, J.C. Steady-State Mean-Square Error Analysis for Adaptive Filtering under the Maximum Correntropy Criterion. *IEEE Signal Process. Lett.* **2014**, *21*, 880–884. [CrossRef]

29. Wang, J.; Dong, P.; Shen, K.; Song, X.; Wang, X. Distributed Consensus Student-t Filter for Sensor Networks With Heavy-Tailed Process and Measurement Noises. *IEEE Access* **2020**, *8*, 167865–167874. [CrossRef]