

Article

A Robust Adaptive Filter for a Complex Hammerstein System

Guobing Qian ^{1,2,*} , Dan Luo ¹ and Shiyuan Wang ^{1,*}

¹ College of Electronic and Information Engineering, Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Southwest University, Chongqing 400715, China; luodan_2018@163.com

² School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

* Correspondence: guobingqian@swu.edu.cn (G.Q.); wsy@swu.edu.cn (S.W.)

Received: 22 December 2018; Accepted: 6 February 2019; Published: 9 February 2019



Abstract: The Hammerstein adaptive filter using maximum correntropy criterion (MCC) has been shown to be more robust to outliers than the ones using the traditional mean square error (MSE) criterion. As there is no report on the robust Hammerstein adaptive filters in the complex domain, in this paper, we develop the robust Hammerstein adaptive filter under MCC to the complex domain, and propose the Hammerstein maximum complex correntropy criterion (HMCCC) algorithm. Thus, the new Hammerstein adaptive filter can be used to directly handle the complex-valued data. Additionally, we analyze the stability and steady-state mean square performance of HMCCC. Simulations illustrate that the proposed HMCCC algorithm is convergent in the impulsive noise environment, and achieves a higher accuracy and faster convergence speed than the Hammerstein complex least mean square (HCLMS) algorithm.

Keywords: complex; Hammerstein; adaptive filters; impulsive noise; stability

1. Introduction

Since traditional mean square error (MSE) criterion derived algorithms are sensitive to outliers, they cannot be used to deal with impulsive noise effectively [1,2]. However, impulsive noise commonly exists in practice. To solve this problem, a higher-order statistic, named correntropy, was proposed [3,4]. It has been proven that the maximum correntropy criterion (MCC) algorithm is robust to impulsive noises, and outperforms the traditional MSE algorithms obviously when the noises are non-Gaussian distributed. Thus, the MCC algorithm [5,6] and its variants [7–10], such as generalized maximum correntropy criterion (GMCC) [7], are widely used in practice.

Different from the well-known Wiener adaptive filter [11], the Hammerstein adaptive filter consists of two parts, namely: a nonlinear memoryless polynomial function and a linear finite impulse response (FIR) filter [12–14]. The Hammerstein system has been widely applied to signal processing [15–18] as well as other applications [19,20]. Considering that the performance of the Hammerstein adaptive filter using MSE criterion decreases dramatically when an impulsive noise exists, Wu et al. applied the MCC criterion to the Hammerstein adaptive filter, and developed a robust Hammerstein adaptive filtering algorithm [21]. This novel adaptive filter is insensitive to outliers and behaves better than the traditional Hammerstein adaptive filters, especially in the case of the impulsive noise.

However, the Hammerstein adaptive filters under the traditional MSE criterion and MCC criterion are defined in the field of real numbers. They cannot be directly employed to handle the complex-valued data. In fact, many signals are defined in the complex domain in practical applications [22–25]. Thus, in this work, we put forward a Hammerstein maximum complex correntropy criterion (HMCCC) algorithm, which extends the Hammerstein adaptive filter, using MCC criterion, to the complex domain.

HMCCC can be used to handle complex-valued data directly, while being robust to impulsive noise. We analyzed the stability and provided the steady-state mean square performance of the HMCCC algorithm. Simulations show that the HMCCC is robust to outliers, and achieves a higher accuracy and faster convergence speed than the Hammerstein complex least mean square (HCLMS) algorithm.

The rest of the paper is organized as follows. A complex Hammerstein adaptive filter under MCCC is developed in Section 2. In Section 3, we analyze the stability and provide the steady-state mean square performance of the HMCCC algorithm. In Section 4, several simulations are presented so as to verify the superior performance of the HMCCC algorithm. Finally, a conclusion is drawn in Section 5.

2. Hammerstein Adaptive Filter under the Maximum Complex Correntropy Criterion

2.1. Complex Correntropy

Considering two complex variables that are $C_1 = X_1 + jY_1$ and $C_2 = X_2 + jY_2$, respectively, the complex correntropy is defined by the following [22]:

$$V^c(C_1, C_2) = E[\kappa_\sigma^c(C_1 - C_2)] \quad (1)$$

where $\kappa_\sigma^c(C_1 - C_2)$ represents the kernel function, and X_1, Y_1, X_2 , and Y_2 denote the real variables.

A Gaussian kernel is adopted in this paper, which is expressed as follows:

$$\kappa_\sigma^c(C_1 - C_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(C_1 - C_2)(C_1 - C_2)^*}{2\sigma^2}\right) \quad (2)$$

with σ being the kernel width.

2.2. Adaptive Filter for Complex Hammerstein System

2.2.1. Cost Function

Consider a complex Hammerstein system, the output of polynomial nonlinear part is as follows:

$$s(k) = \mathbf{p}^H \mathbf{x}_p(k) \quad (3)$$

where $\mathbf{p} = [p_1 \ p_2 \ \cdots \ p_M]^T$ denotes the vector of the complex polynomial coefficient, M is the polynomial order, and $\mathbf{x}_p(k) = [x(k) \ x^2(k) \ \cdots \ x^M(k)]^T$ is the complex polynomial regressor vector, $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose, respectively.

The cost function of the complex Hammerstein filtering algorithm under MCCC is as follows:

$$J_{HMCCC} = E[\kappa_\sigma^c(e(k))] \quad (4)$$

where $e(k) = d(k) - \mathbf{w}^H \mathbf{s}(k)$ is the error at the k -th iteration, $\mathbf{w} = [w_1 \ w_2 \ \cdots \ w_N]^T$ denotes the estimated weight vector, $\mathbf{s}(k) = [s(k) \ s(k-1) \ \cdots \ s(k-N+1)]^T$ is the input vector of the complex finite impulse response (FIR) filter, $d(k) = \mathbf{w}_0^H \mathbf{s}_0(k) + v(k)$ is the desired signal at the k -th iteration, $\mathbf{s}_0(k) = [s_0(k) \ s_0(k-1) \ \cdots \ s_0(k-N+1)]^T$, $s_0(k) = \mathbf{p}_0^H \mathbf{x}_p(k)$, N denotes the length of the linear FIR filter, $\mathbf{w}_0$ and $\mathbf{p}_0$ are the unknown system parameters to be estimated, which are the optimum solutions for $\mathbf{w}$ and $\mathbf{p}$, and $v(k)$ is the observation noise.

2.2.2. Adaptive Algorithm

Based on the Wirtinger Calculus [26–28], we derive the stochastic gradient of J_{HMCCC} with respect to $\mathbf{p}^*$ as follows:

$$\frac{\partial J_{HMCCC}}{\partial \mathbf{p}^*} = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \frac{1}{2\sigma^2} e^*(k) \mathbf{X}(k) \mathbf{w}^*(k) \quad (5)$$

and with respect to $\mathbf{w}^*$ as follows:

$$\frac{\partial J_{HMCCC}}{\partial \mathbf{w}^*} = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \frac{1}{2\sigma^2} e^*(k) \mathbf{X}^T(k) \mathbf{p}^*(k) \quad (6)$$

where $\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}_p(k) & \mathbf{x}_p(k-1) & \cdots & \mathbf{x}_p(k-N+1) \end{bmatrix}$.

Then, the updates for $\mathbf{p}$ and $\mathbf{w}$ are as follows:

$$\mathbf{p}(k+1) = \mathbf{p}(k) + \eta_p f(e(k)) \mathbf{X}(k) \mathbf{w}^*(k) \quad (7)$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta_w f(e(k)) \mathbf{X}^T(k) \mathbf{p}^*(k) \quad (8)$$

where $f(e(k)) = \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] e^*(k)$ and the constant $\frac{1}{2\pi\sigma^2}$ is merged into the step-size parameters η_p and η_w .

Finally, Table 1 summarizes the HMCCC algorithm.

Table 1. The Hammerstein maximum complex correntropy criterion (HMCCC) algorithm.

Input: $\sigma, \eta_p, \eta_w, M, N, d(k), x(k)$
1. Initializations: $\mathbf{p}(0), \mathbf{w}(0)$.
2. While $\{d(k), x(k)\}$ available, do
3. $\mathbf{x}_p(k) = \begin{bmatrix} x(k) & x^2(k) & \cdots & x^M(k) \end{bmatrix}^T$
4. $\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}_p(k) & \mathbf{x}_p(k-1) & \cdots & \mathbf{x}_p(k-N+1) \end{bmatrix}$
5. $\mathbf{s}(k) = \mathbf{X}^T(k) \mathbf{p}^*(k)$
6. $e(k) = d(k) - \mathbf{w}^H(k) \mathbf{s}(k)$
7. $f(e(k)) = \exp\left[-\frac{ e(k) ^2}{2\sigma^2}\right] e^*(k)$
8. $\mathbf{p}(k+1) = \mathbf{p}(k) + \eta_p f(e(k)) \mathbf{X}(k) \mathbf{w}^*(k)$
9. $\mathbf{w}(k+1) = \mathbf{w}(k) + \eta_w f(e(k)) \mathbf{X}^T(k) \mathbf{p}^*(k)$
10. End while
11. $\hat{\mathbf{w}}_0 = \mathbf{w}(k+1)$
Output: Estimated polynomial coefficient $\hat{\mathbf{p}}_0$ and filter weight $\hat{\mathbf{w}}_0$.

3. Convergence Analysis

To begin the derivation of the convergence analysis, some widely used assumptions are adopted, as follows:

(A1) $v(k)$ is independently identically distributed (iid), zero-mean, circular, and independent of $e_p(k)$, $e_w(k)$ and $x(k)$;

(A2) Both $\|\mathbf{X}^T(k) \mathbf{p}^*(k)\|^2$ and $\|\mathbf{X}(k) \mathbf{w}^*(k)\|^2$ are uncorrelated with $|f(e(k))|^2$ when $k \rightarrow \infty$.

Remark 1: (1) A1 is reasonable in a practical case and is widely used in the theoretical analysis of an adaptive filter [21,29,30];

(2) When $k \rightarrow \infty$, $\mathbf{p}(k) \rightarrow \mathbf{p}_0(k)$ and $\mathbf{w}(k) \rightarrow \mathbf{w}_0(k)$. Additionally, $\mathbf{X}(k)$ is independent of $e(k)$, based on A1. Thus, both $\|\mathbf{X}^T(k) \mathbf{p}^*(k)\|^2$ and $\|\mathbf{X}(k) \mathbf{w}^*(k)\|^2$ are uncorrelated with $|f(e(k))|^2$. Thus, A2 is also a reasonable assumption.

3.1. Stability Analysis

As $|e(k)|^2$ is a real-valued function for $\begin{bmatrix} \mathbf{p}^T(k) & \mathbf{w}^T(k) \end{bmatrix}^T$, the following expression can be derived by taking the Taylor series expansion of $|e(k+1)|^2$ at $\begin{bmatrix} \mathbf{p}^T(k) & \mathbf{w}^T(k) \end{bmatrix}^T$,

$$|e(k+1)|^2 = |e(k)|^2 + 2\text{Re} \left\{ \left. \frac{\partial |e(k)|^2}{\partial \mathbf{p}^*(k)} \right|_{\mathbf{w}(k)=\text{cons}} \cdot (\Delta \mathbf{p}(k))^* \right\} + 2\text{Re} \left\{ \left. \frac{\partial |e(k)|^2}{\partial \mathbf{w}^*(k)} \right|_{\mathbf{p}(k)=\text{cons}} \cdot (\Delta \mathbf{w}(k))^* \right\} + h.o.t. \quad (9)$$

where

$$\frac{\partial |e(k)|^2}{\partial \mathbf{p}^*(k)} = -\mathbf{X}(k)\mathbf{w}^*(k)e^*(k) \quad (10)$$

$$\frac{\partial |e(k)|^2}{\partial \mathbf{w}^*(k)} = -\mathbf{X}^T(k)\mathbf{p}^*(k)e^*(k) \quad (11)$$

$$\Delta \mathbf{p}(k) = \eta_p f(e(k))\mathbf{X}(k)\mathbf{w}^*(k) \quad (12)$$

$$\Delta \mathbf{w}(k) = \eta_w f(e(k))\mathbf{X}^T(k)\mathbf{p}^*(k) \quad (13)$$

and *h.o.t* represents the terms of higher order infinitesimal.

Then,

$$\begin{aligned} E\{|e(k+1)|^2\} &= E\left\{ \left[1 - 2\eta_p \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \|\mathbf{X}(k)\mathbf{w}^*(k)\|^2 - 2\eta_w \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \|\mathbf{X}^T(k)\mathbf{p}^*(k)\|^2 \right] |e(k)|^2 \right\} \\ &\approx E\left\{ \left[1 - 2\eta_p \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \|\mathbf{X}(k)\mathbf{w}^*(k)\|^2 - 2\eta_w \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \|\mathbf{X}^T(k)\mathbf{p}^*(k)\|^2 \right] \right\} E\{|e(k)|^2\} \end{aligned} \quad (14)$$

Thus, the sequence $|e(k)|$ will decrease in the mean sense if

$$E\left\{ \left[1 - 2\eta_p \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \|\mathbf{X}(k)\mathbf{w}^*(k)\|^2 - 2\eta_w \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \|\mathbf{X}^T(k)\mathbf{p}^*(k)\|^2 \right] \right\} \leq 1 \quad (15)$$

that is,

$$0 \leq E\left\{ \left| \eta_p \|\mathbf{X}(k)\mathbf{w}^*(k)\|^2 + \eta_w \|\mathbf{X}^T(k)\mathbf{p}^*(k)\|^2 \right| \right\} \leq \frac{1}{E\left\{ \exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \right\}} \quad (16)$$

Considering the fact that $\exp\left[-\frac{|e(k)|^2}{2\sigma^2}\right] \leq 1$, we can obtain that the sequence $|e(k)|$ will decrease in the mean sense, if

$$0 \leq E\left[\left| \eta_p \|\mathbf{X}(k)\mathbf{w}^*(k)\|^2 + \eta_w \|\mathbf{X}^T(k)\mathbf{p}^*(k)\|^2 \right| \right] \leq 1 \quad (17)$$

In this case, the HMCCC algorithm will converge in the mean sense.

3.2. Steady Excess Mean Square Error

We define $H_p = \lim_{k \rightarrow \infty} E[|e_p(k)|^2]$ as the steady excess mean square error (EMSE) for the nonlinear part, $H_w = \lim_{k \rightarrow \infty} E[|e_w(k)|^2]$ as the steady EMSE for the linear filter, and $H_{pw} = \lim_{k \rightarrow \infty} E[|e_{pw}(k)|^2]$ as the steady EMSE for the whole Hammerstein system, where $e_p(k) = \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}_0^* - \mathbf{w}_0^H(k) \mathbf{X}^T(k) \mathbf{p}^*(k)$, $e_w(k) = \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}_0^*(k) - \mathbf{w}^H(k) \mathbf{X}^T(k) \mathbf{p}_0^*(k)$, $e_{pw}(k) = \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}_0^* - \mathbf{w}^H(k) \mathbf{X}^T(k) \mathbf{p}^*(k)$.

When the algorithm reaches the steady, the error for the whole Hammerstein system can be approximately divided into two parts, as follows:

$$\begin{aligned} e_{pw}(k) &= \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}_0^* - \mathbf{w}^H(k) \mathbf{X}^T(k) \mathbf{p}^*(k) \\ &= \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}_0^* - \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}^*(k) + \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}^*(k) - \mathbf{w}^H(k) \mathbf{X}^T(k) \mathbf{p}^*(k) \\ &\approx \mathbf{w}_0^H(k) \mathbf{X}^T(k) \tilde{\mathbf{p}}^*(k) + \tilde{\mathbf{w}}^H(k) \mathbf{X}^T(k) \mathbf{p}_0^*(k) \\ &= e_p(k) + e_w(k) \end{aligned} \quad (18)$$

When only the nonlinear part is taken into consideration, we have

$$\begin{aligned} e_p(k) &= \mathbf{w}_0^H \mathbf{X}^T(k) \mathbf{p}_0^* - \mathbf{w}_0^H(k) \mathbf{X}^T(k) \mathbf{p}^*(k) \\ &\approx \mathbf{w}_0^H(k) \mathbf{X}^T(k) \tilde{\mathbf{p}}^*(k) \end{aligned} \quad (19)$$

and

$$\tilde{\mathbf{p}}(k+1) = \tilde{\mathbf{p}}(k) - \eta_p f(e(k)) \mathbf{X}(k) \mathbf{w}_0^*(k) \quad (20)$$

Multiplying each side of Equation (20) by its conjugate transpose and taking the expectation, we obtain the following:

$$\begin{aligned} E\{\|\tilde{\mathbf{p}}(k+1)\|^2\} &= E\{\|\tilde{\mathbf{p}}(k)\|^2\} - 2\eta_p E\{\text{Re}[e_p(k)f(e(k))]\} \\ &\quad + \eta_p^2 E\{\|\mathbf{X}(k)\mathbf{w}_0^*(k)\|^2 |f(e(k))|^2\} \end{aligned} \quad (21)$$

Since $\lim_{k \rightarrow \infty} E\{\|\tilde{\mathbf{p}}(k+1)\|^2\} = \lim_{k \rightarrow \infty} E\{\|\tilde{\mathbf{p}}(k)\|^2\}$, we further obtain the following:

$$2 \lim_{k \rightarrow \infty} E\{\text{Re}[e_p(k)f(e(k))]\} = \lim_{k \rightarrow \infty} \eta_p E\{\|\mathbf{X}(k)\mathbf{w}_0^*(k)\|^2 |f(e(k))|^2\} \quad (22)$$

Based on the results of Equations (38) and (46) in the literature [23], we similarly obtain the following expressions by replacing α and λ with 1 and $1/2\sigma^2$, respectively, as follows:

$$E\{\text{Re}[e_p(k)f(e(k))]\} = H_p E\left\{\exp\left[-|v(k)|^2/2\sigma^2\right] \left[1 - \left(|v(k)|^2/2\sigma^2\right)\right]\right\} \quad (23)$$

$$E\{|f(e(k))|^2\} = E\left\{\exp\left[-|v(k)|^2/\sigma^2\right] |v(k)|^2\right\} + H_p \times R_1 \quad (24)$$

where

$$R_1 = E\left\{\exp\left[-|v(k)|^2/\sigma^2\right] \left[|v(k)|^4/\sigma^4 - 3|v(k)|^2/\sigma^2 - 1\right]\right\} \quad (25)$$

Furthermore, based on the result of Equation (47) in the literature [23], we obtain the EMSE for the nonlinear part by replacing $\text{Tr}(\mathbf{R}_{\mathbf{x}\mathbf{x}^H})$ (i.e., $E\{\|\mathbf{x}(i)\|^2\}$) with $E\{\|\mathbf{X}(k)\mathbf{w}_0^*\|^2\}$,

$$H_p = \frac{\eta_p E\{\|\mathbf{X}(k)\mathbf{w}_0^*\|^2\} E\left\{\exp\left[-|v(k)|^2/\sigma^2\right] |v(k)|^2\right\}}{2E\left\{\exp\left[-|v(k)|^2/2\sigma^2\right] \left[1 - \left(|v(k)|^2/2\sigma^2\right)\right]\right\} - \eta_p E\{\|\mathbf{X}(k)\mathbf{w}_0^*\|^2\} R_1} \quad (26)$$

When only the FIR filter is taken into consideration, we similarly derive

$$H_w = \frac{\eta_w E\{\|\mathbf{X}^T(k)\mathbf{p}_0^*\|^2\} E\{\exp[-|v(k)|^2/\sigma^2] |v(k)|^2\}}{2E\{\exp[-|v(k)|^2/2\sigma^2] [1 - (|v(k)|^2/2\sigma^2)]\} - \eta_w E\{\|\mathbf{X}^T(k)\mathbf{p}_0^*\|^2\} R_1} \quad (27)$$

Furthermore, when both the nonlinear part and FIR filter are considered, we have

$$\begin{aligned} H_{pw} &= \lim_{k \rightarrow \infty} E[|e_{pw}(k)|^2] \approx \lim_{k \rightarrow \infty} E[|e_p(k) + e_w(k)|^2] \\ &= \lim_{k \rightarrow \infty} E[|e_p(k)|^2] + \lim_{k \rightarrow \infty} E[|e_w(k)|^2] + \lim_{k \rightarrow \infty} E[2\text{Re}(e_p^*(k)e_w(k))] \\ &= H_p + H_w + H_{\text{cross}} \end{aligned} \quad (28)$$

Remark 2: $H_{\text{cross}} = \lim_{k \rightarrow \infty} E[2\text{Re}(e_p^*(k)e_w(k))]$ is the cross EMSE of the Hammerstein system, and equals zero when both $e_p(k)$ and $e_w(k)$ are zero mean and independent.

4. Simulation

In this part, we provide some simulations to illustrate the superior performance of HMCCC. We chose the weight vector as $\mathbf{w}_0 = [1 + 0.6j \quad 0.6 + j \quad 0.1 + 0.2j \quad 0.2 + 0.1j \quad 0.06 + 0.04j \quad 0.04 + 0.06j]^T$ and the complex polynomial coefficient vector as $\mathbf{p}_0 = [1 + 0.6j \quad 0.6 + j]^T$. An additive complex noise $v = v_R + jv_I$ was considered in the simulations, with v_R and v_I being the real and imaginary parts, respectively. We compared the performance of HMCCC with HCLMS (HCLMS is the extension of HLMS [17] to complex domain, which is summarized in the Appendix A), and chose the parameters of both algorithms by trial, in order to ensure a desirable solution. Simulation results were obtained by averaging 100 Monte Carlo trials. The input signal $x(k)$ was generated by a first-order autoregressive process, as follows:

$$x(k) = ax(k-1) + \sqrt{1-a^2}\zeta(k) \quad (29)$$

where $x(k) = x_R(k) + jx_I(k)$, $x_R(k)$, and $x_I(k)$ are the real and imaginary parts of $x(k)$, $a = 0.95$, $\zeta(k) = \zeta_R(k) + j\zeta_I(k)$, $\zeta_R(k)$, and $\zeta_I(k)$ are the real and imaginary parts of $\zeta(k)$, $\zeta_R(k), \zeta_I(k) \sim N(0, 1)$, and $N(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2 , respectively.

First, the superiority of HMCCC was verified in the complex alpha stable noise environment. The noise parameters were $v_R, v_I \in \sigma_v \cdot v_{\text{alpha}}(\alpha, \beta, \gamma, \delta)$, where $\sigma_v^2 = 0.1$, $\alpha = 1.2$ is the characteristic factor, $\beta = 0$ is the symmetry parameter, $\gamma = 0.6$ is the dispersion parameter, and $\delta = 0$ is the location parameter, respectively. Figure 1 shows the time sequence and histogram for the real and imaginary parts of the complex alpha stable noise. It is noted that HCLMS may diverge in the complex alpha stable noise environment. Thus, we omitted the trials for HCLMS if $\|\mathbf{w}\|_2 \geq 100$. The simulation shows that HCLMS diverged twice in the 100 trials, while HMCCC did not diverge. The performances of the different algorithms in terms of the normalized testing mean square error (MSE) are shown in Figure 2, where the testing MSE was obtained from a test set of 100 samples, and the trials of the divergence were omitted for HCLMS. It is clear that compared with HCLMS, HMCCC has a better filtering performance in the presence of complex alpha stable noise.

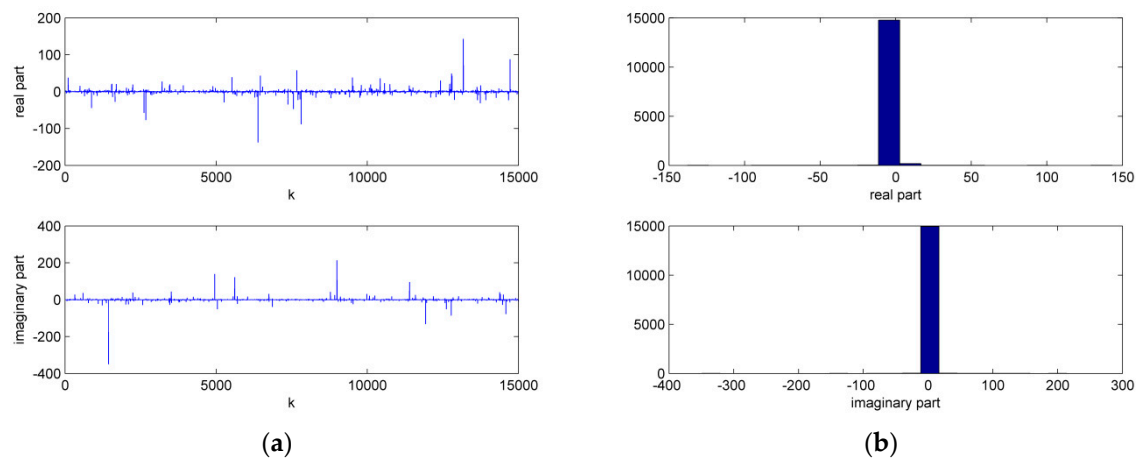


Figure 1. Time sequence and histogram for the complex alpha stable noise. (a) time sequence; (b) histogram.

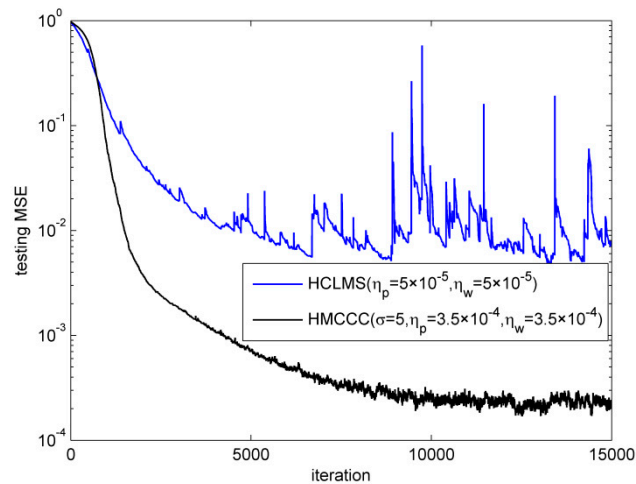


Figure 2. Learning curves of different algorithms. HCLMS—Hammerstein complex least mean square; HMCCC—Hammerstein maximum complex correntropy criterion.

Then, we compared the steady testing MSE of HMCCC under different noise parameters. We ran 15,000 iterations to make sure the HMCCC reaches steady, and calculated the steady testing MSE with the average of next 1000 iterations. Figure 3 shows the steady normalized testing MSEs under different characteristic factors and dispersion parameters, respectively. It illustrates that HMCCC can perform well under different parameters of alpha stable noise.

Next, the superiority of HMCCC is verified in the contaminated Gaussian (CG) noise environment, where $v(k) = (1 - c(k))v_1(k) + c(k)v_2(k)$, $v_1(k) = v_{1R}(k) + jv_{1I}(k)$, $v_2(k) = v_{2R}(k) + jv_{2I}(k)$, $v_{1R}, v_{1I} \in N(0, 0.1)$, and $v_{2R}, v_{2I} \in N(0, 20)$ represent outliers. Additionally, $P(c(k) = 1) = 0.06$, $P(c(k) = 0) = 1 - 0.06$. Figure 4 shows the time sequence and histogram for the real and imaginary parts of the CG noise. The performances of the different algorithms on the basis of normalized testing MSE are shown in Figure 5, where the testing MSE was also obtained from a test set of 100 samples. One can clearly see that, compared with HCLMS, HMCCC has a better filtering performance in the presence of CG noise.

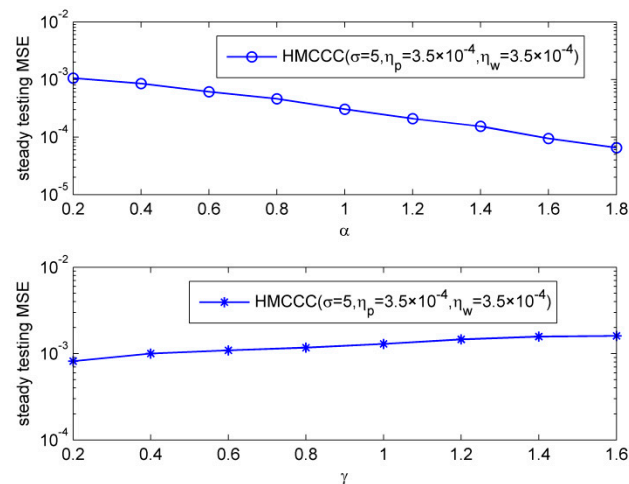


Figure 3. Steady testing mean square errors (MSEs) under different characteristic factors and dispersion parameters.

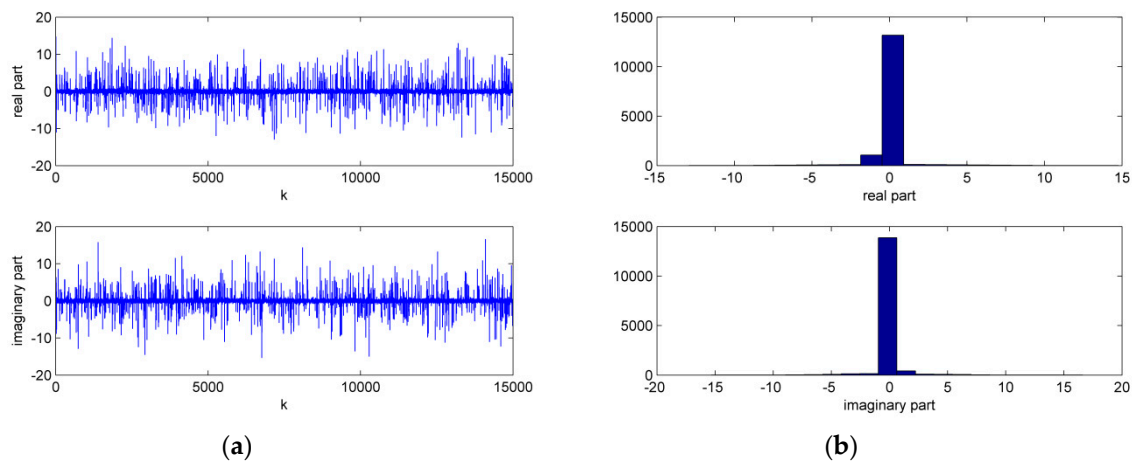


Figure 4. Time sequence and histogram for the contaminated Gaussian (CG) noise. (a) time sequence; (b) histogram.

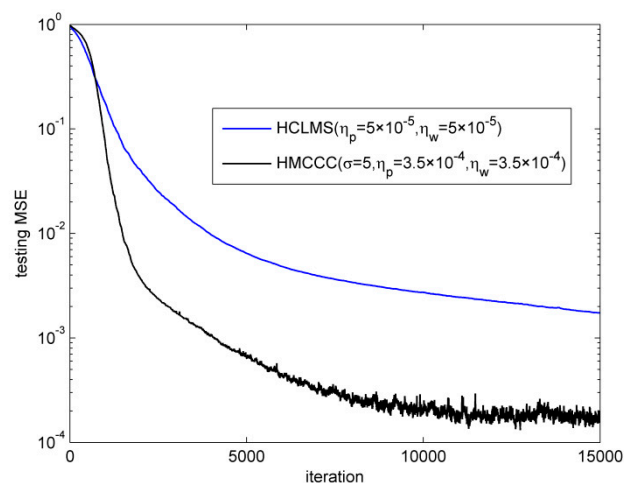


Figure 5. Learning curves of different algorithms.

Furthermore, we tested the robustness of the HMCCC algorithm to the outlier. The CG noise was also used in this simulation, where $v_{1R}, v_{1I} \in N(0, 0.1)$, $P(c(k) = 1) = p$, $v_{2R}, v_{2I} \in N(0, \sigma_B^2)$, and $P(c(k) = 0) = 1 - p$. Figure 6 depicts the steady normalized testing MSE of the HMCCC algorithm

in the case of different probabilities of outlier, p ($\sigma_B^2 = 20$), and variances of outlier, σ_B^2 , where 15,000 iterations were run to make sure HMCCC reached steady, and the steady normalized testing MSEs were calculated with the average of the next 1000 iterations. One can observe that the proposed HMCCC algorithm is robust to the outlier, and behaves well under different p and σ_B^2 . Moreover, HMCCC has a slightly smaller steady testing MSE with a bigger σ_B^2 , which is a little surprising, but is consistent with Chen's work [9]. This is due to the fact that the convergence rates are slightly different under different σ_B^2 , even with the same learning rate.

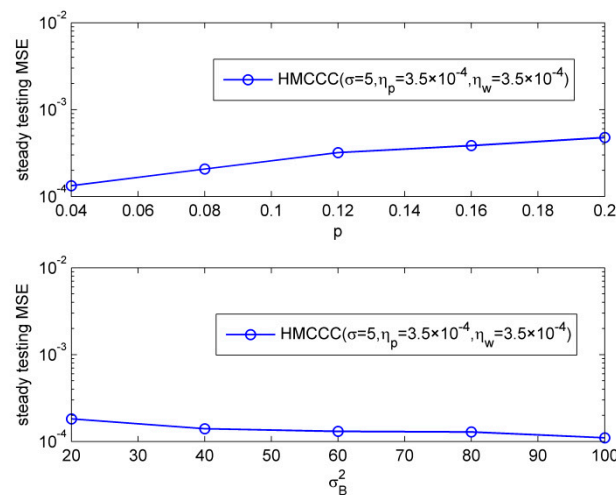


Figure 6. Influence of the probability and variance of outlier.

Afterward, we investigated the influences of the kernel width σ on the performance of HMCCC. The CG noise was also employed in this simulation, where $v_{1R}, v_{1I} \in N(0, 0.1)$, $P(c(k) = 1) = 0.06$, $v_{2R}, v_{2I} \in N(0, 20)$, and $P(c(k) = 0) = 1 - 0.06$. Figure 7 presents the normalized testing MSE of HMCCC in the case of three different kernel widths σ . It can be seen that kernel width σ has a vital role on the learning rate and steady value of HMCCC. With a small kernel width, HMCCC converges slowly, but achieves a small steady value. On the contrary, with a large kernel width, HMCCC converges quickly, but achieves a high steady value.

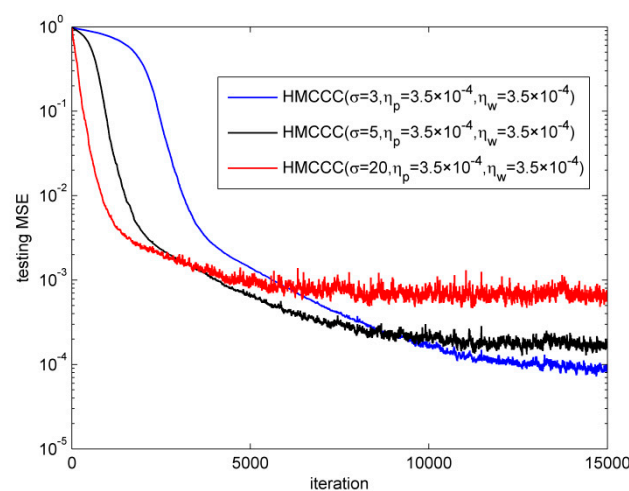


Figure 7. Influence of σ .

Finally, we compared the simulated steady testing MSEs with the theoretical ones. The Gaussian noise is used in this simulation, where $v(k) = v_R(k) + jv_I(k)$ and $v_R, v_I \in N(0, \sigma_v^2)$. It is noted that the testing MSEs were not normalized in this simulation, and were obtained from a test set of 1000 samples. In addition, the theoretical values for the nonlinear part and FIR part were calculated by Equations (26) and (27), respectively. Figure 8 shows the simulated steady testing MSEs and the theoretical ones under different σ_v^2 , where 40,000 iterations were run to make sure the algorithm reached steady, and the steady testing MSEs were calculated with the average of the next 1000 iterations. One can see that the simulated values almost matched with the theoretical ones for the nonlinear part and FIR part. Moreover, there is a little gap between the simulated whole system and the sum of theoretical nonlinear and FIR parts, which is the value of cross EMSE.

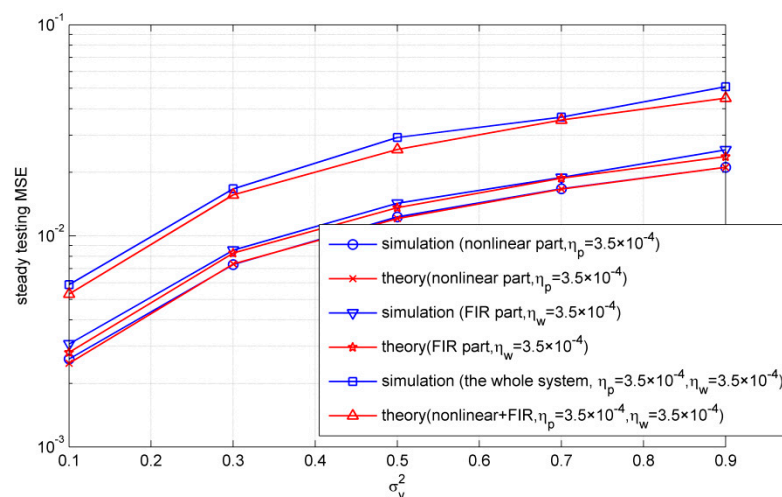


Figure 8. Steady-state testing MSEs with different noise variances ($\sigma = 5$).

5. Conclusions

Since the Hammerstein adaptive filter can only be used to deal with real-valued data, in this paper, we extended the Hammerstein filter under maximum correntropy criterion (MCC) to the complex domain and developed a new algorithm, named the Hammerstein maximum complex correntropy criterion (HMCCC). Simultaneously, we analyzed the stability and derived some theoretical results for the HMCCC algorithm. The simulation illustrated that HMCCC is always convergent and performs better than the traditional Hammerstein complex LMS (HCLMS) algorithm in the presence of impulsive noises. Additionally, the kernel width has an important impact on the performance of HMCCC, and the novel algorithm behaves well with different probabilities and variances of outliers.

Author Contributions: Conceptualization, G.Q., D.L., and S.W.; methodology, G.Q. and S.W.; software, D.L.; validation, G.Q.; formal analysis, G.Q.; investigation, G.Q., D.L., and S.W.; resources, G.Q.; data curation, D.L.; writing (original draft preparation), G.Q. and D.L.; writing (review and editing), S.W.; visualization, D.L.; supervision, G.Q., D.L., and S.W.; project administration, G.Q.; funding acquisition, G.Q.

Funding: This research was funded by the China Postdoctoral Science Foundation Funded Project under grant 2017M610583, and the Fundamental Research Funds for the Central Universities under grant XDJK2018C020.

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

Appendix A

The summary of the HCLMS algorithm:

Input: $\eta_p, \eta_w, M, N, d(k), x(k)$
1. Initializations: $\mathbf{p}(0), \mathbf{w}(0)$.
2. While $\{ d(k) \quad x(k) \}$ available, do
3. $\mathbf{x}_p(k) = \begin{bmatrix} x(k) & x^2(k) & \cdots & x^M(k) \end{bmatrix}^T$
4. $\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}_p(k) & \mathbf{x}_p(k-1) & \cdots & \mathbf{x}_p(k-N+1) \end{bmatrix}$
5. $\mathbf{s}(k) = \mathbf{X}^T(k) \mathbf{p}^*(k)$
6. $e(k) = d(k) - \mathbf{w}^H(k) \mathbf{s}(k)$
7. $\mathbf{p}(k+1) = \mathbf{p}(k) + \eta_p e^*(k) \mathbf{X}(k) \mathbf{w}^*(k)$
8. $\mathbf{w}(k+1) = \mathbf{w}(k) + \eta_w e^*(k) \mathbf{X}^T(k) \mathbf{p}^*(k)$
9. End while
10. $\hat{\mathbf{w}}_0 = \mathbf{w}(k+1)$
Output: Estimated polynomial coefficient $\hat{\mathbf{p}}_0$ and filter weight $\hat{\mathbf{w}}_0$.

References

1. Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer: New York, NY, USA, 2010.
2. Chen, B.; Zhu, Y.; Hu, J.; Principe, J.C. *System Parameter Identification: Information Criteria and Algorithms*; Newnes: Oxford, UK, 2013.
3. Liu, W.; Pokharel, P.P.; Principe, J. Correntropy: A localized similarity measure. In Proceedings of the 2006 IEEE International Joint Conference on Neural Network (IJCNN), Vancouver, BC, Canada, 16–21 July 2006; pp. 4919–4924.
4. Liu, W.; Pokharel, P.P.; Principe, J. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [[CrossRef](#)]
5. Singh, A.; Principe, J.C. Using correntropy as a cost function in linear adaptive filters. In Proceedings of the 2009 International Joint Conference on Neural Networks (IJCNN), Atlanta, GA, USA, 14–19 June 2009; pp. 2950–2955.
6. Singh, A.; Principe, J.C. A loss function for classification based on a robust similarity metric. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–6.
7. Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Principe, J.C. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [[CrossRef](#)]
8. Zhao, J.; Zhang, H. Kernel recursive generalized maximum correntropy. *IEEE Signal Process. Lett.* **2017**, *24*, 1832–1836. [[CrossRef](#)]
9. Chen, B.; Xing, L.; Xu, B.; Zhao, H.; Zheng, N.; Principe, J.C. Kernel Risk-Sensitive Loss: Definition, Properties and Application to Robust Adaptive Filtering. *IEEE Trans. Signal Process.* **2017**, *65*, 2888–2901. [[CrossRef](#)]
10. Wang, S.; Dang, L.; Wang, W.; Qian, G.; Tse, C.K. Kernel Adaptive Filters With Feedback Based on Maximum Correntropy. *IEEE Access.* **2018**, *6*, 10540–10552. [[CrossRef](#)]
11. Diniz, P.S.R. *Adaptive Filtering: Algorithms and Practical Implementation*, 4th ed.; Springer: New York, NY, USA, 2013.
12. Bai, E.W.; Li, D. Convergence of the iterative Hammerstein system identification algorithm. *IEEE Trans. Autom. Control.* **2004**, *49*, 1929–1940. [[CrossRef](#)]
13. Jeraj, J.; Mathews, V.J. Stochastic mean-square performance analysis of an adaptive Hammerstein filter. *IEEE Trans. Signal Process.* **2006**, *54*, 2168–2177. [[CrossRef](#)]
14. Ding, F.; Liu, X.P.; Liu, G. Identification methods for Hammerstein nonlinear systems. *Dig. Signal Process.* **2011**, *21*, 215–238. [[CrossRef](#)]
15. Scarpiniti, M.; Communiello, D.; Parisi, R.; Uncini, A. Comparison of Hammerstein and Wiener systems for nonlinear acoustic echo cancelers in reverberant environments. In Proceedings of the 17th International Conference on Digital Signal Processing (DSP), Corfu, Greece, 6–8 July 2011; pp. 1–6.

16. Stenger, A.; Kellermann, W. Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling. *Signal Process.* **2000**, *80*, 1747–1760. [[CrossRef](#)]
17. Scarpiniti, M.; Comminiello, D.; Parisi, R.; Uncini, A. Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties. *Signal Process.* **2014**, *100*, 112–123. [[CrossRef](#)]
18. Dang, L.; Wang, W.; Sun, Q.; Long, Z.; Qian, G.; Wang, S. Hammerstein adaptive filter with single feedback under minimum mean square error. In Proceedings of the 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017; pp. 1–7.
19. Hunt, K.J.; Munih, M.; Donaldson, N.N.; Barr, F.M.D. Investigation of the Hammerstein hypothesis in the modeling of electrically stimulated muscle. *IEEE Trans. Biomed. Eng.* **1998**, *45*, 998–1009. [[CrossRef](#)] [[PubMed](#)]
20. Su, S.W.; Wang, L.; Celler, B.G.; Savkin, A.V.; Guo, Y. Identification and control of heart rate regulation during treadmill exercise. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 1238–1246. [[CrossRef](#)] [[PubMed](#)]
21. Wu, Z.; Peng, S.; Chen, B.; Zhao, H. Robust Hammerstein adaptive filtering under maximum correntropy criterion. *Entropy* **2015**, *17*, 7149–7166. [[CrossRef](#)]
22. Guimaraes, J.P.F.; Fontes, A.I.R.; Rego, J.B.A.; Martins, A.M.; Príncipe, J.C. Complex correntropy: Probabilistic interpretation and application to complex-valued data. *IEEE Signal Process. Lett.* **2017**, *24*, 42–45. [[CrossRef](#)]
23. Qian, G.; Wang, S. Generalized Complex Correntropy: Application to Adaptive Filtering of Complex Data. *IEEE Access.* **2018**, *6*, 19113–19120. [[CrossRef](#)]
24. Guimaraes, J.P.F.; Fontes, A.I.R.; Rego, J.B.A.; Martins, A.M.; Príncipe, J.C. Complex Correntropy Function: Properties, and application to a channel equalization problem. *Expert Syst. Appl.* **2018**, *107*, 173–181. [[CrossRef](#)]
25. Qian, G.; Luo, D.; Wang, S. Recursive Minimum Complex Kernel Risk Sensitive Loss Algorithm. *Entropy* **2018**, *20*, 902. [[CrossRef](#)]
26. Wirtinger, W. Zur formalen theorie der funktionen von mehr complexen veränderlichen. *Math. Ann.* **1927**, *97*, 357–375. (In German) [[CrossRef](#)]
27. Bouboulis, P.; Theodoridis, S. Extension of Wirtinger's calculus to reproducing Kernel Hilbert spaces and the complex kernel LMS. *IEEE Trans. Signal Process.* **2011**, *59*, 964–978. [[CrossRef](#)]
28. Zhang, X. *Matrix Analysis and Application*, 2th ed.; Tsinghua University Press: Beijing, China, 2013.
29. 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.
30. Picinbono, B. On circularity. *IEEE Trans. Signal Process.* **1994**, *42*, 3473–3482. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).