

Article

Robust Hammerstein Adaptive Filtering under Maximum Correntropy Criterion

Zongze Wu ¹, Siyuan Peng ¹, Badong Chen ^{2,*} and Haiquan Zhao ³

¹ School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China; E-Mails: zzwu@scut.edu.cn (Z.W.); pengsiyuan9@gmail.com (S.P.)

² School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

³ School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China; E-Mail: hqzhao@home.swjtu.edu.cn

* Author to whom correspondence should be addressed; E-Mail: chenbd@mail.xjtu.edu.cn; Tel.: +86-29-82668672.

Academic Editor: Kevin H. Knuth

Received: 25 June 2015 / Accepted: 12 October 2015 / Published: 22 October 2015

Abstract: The maximum correntropy criterion (MCC) has recently been successfully applied to adaptive filtering. Adaptive algorithms under MCC show strong robustness against large outliers. In this work, we apply the MCC criterion to develop a robust Hammerstein adaptive filter. Compared with the traditional Hammerstein adaptive filters, which are usually derived based on the well-known mean square error (MSE) criterion, the proposed algorithm can achieve better convergence performance especially in the presence of impulsive non-Gaussian (e.g., α -stable) noises. Additionally, some theoretical results concerning the convergence behavior are also obtained. Simulation examples are presented to confirm the superior performance of the new algorithm.

Keywords: Hammerstein adaptive filtering; MCC; nonlinear system identification

MSC Codes: 62B10

1. Introduction

Nonlinear system identification is still an active research area [1]. Although linear systems have established a solid theory [2], most practical systems (e.g., hands-free telephone systems) may be more adequately represented as a nonlinear model. One of the main challenges for nonlinear system identification is the choice of an appropriate nonlinear filtering structure that accurately captures the characteristics of the underlying nonlinear system. A common structure used in nonlinear modeling is the *block-oriented* representation. The Wiener model and the Hammerstein model are two typical block-oriented nonlinear models [3]. Specifically, the Wiener model consists of a cascade of a linear time invariant (LTI) filter followed by a static nonlinear function, indicated as a linear-nonlinear (LN) model [4–6], and the Hammerstein model consists of a cascade of a static nonlinear function followed by a LTI filter, known as a nonlinear-linear (NL) model [7–19]. Other nonlinear models include neural networks (NNs) [20], Volterra adaptive filters (VAFs) [21], kernel adaptive filters (KAF) [22–25], among others.

Hammerstein filters can accurately model many real-world systems and, as a consequence, they have been successfully used in various applications of engineering [26–29]. Due to its simplicity and efficiency, the mean square error (MSE) criterion has been widely applied in Hammerstein adaptive filtering [30]. Adaptive algorithms under MSE usually perform very well when the desired signals are disturbed by Gaussian noises. However, when the desired signals are disturbed by non-Gaussian noises, especially in the presence of large outliers (observations that significantly deviate from the bulk of data), the performance of the MSE based algorithms may deteriorate rapidly. Actually, MSE is rather sensitive to outliers. In most practical situations, heavy-tailed impulsive noises may occur, which often cause large outliers. For instance, different types of artificial noises in electronic devices, atmospheric noises, and lighting spikes in natural phenomena, can be described as an impulsive noise [31,32].

In this work, instead of using the MSE criterion, we apply the maximum correntropy criterion (MCC) to develop a robust Hammerstein adaptive filtering algorithm. Correntropy is a nonlinear similarity measure between two signals [33,34]. The MCC aims at maximizing the similarity (measured by correntropy) between the model output and the desired response such that the adaptive model is as close as possible to the unknown system. It has been shown that, the MCC in terms of the stability and accuracy, is very robust with respect to impulsive noises [33–39]. Compared with the traditional Hammerstein adaptive filtering algorithms based on the MSE criterion, the new algorithm can achieve better performance especially in the presence of impulsive non-Gaussian noises.

The organization of the rest of the paper is as follows. In Section 2, after briefly introducing the correntropy, we derive a Hammerstein adaptive filtering algorithm under MCC criterion. In Section 3, we carry out the convergence analysis. In Section 4, we present simulation examples to demonstrate the superior performance of the proposed algorithm. Finally, we give the conclusion in Section 5.

2. Hammerstein Adaptive Filtering under the Maximum Correntropy Criterion

Figure 1 shows the structure of a Hammerstein adaptive filter under MCC criterion, where the filter consists of a polynomial memoryless nonlinearity followed by a linear FIR filter. This structure has been commonly used in Hammerstein adaptive filtering [8,9,27]. As shown in Figure 1, under the MCC

criterion, the parameters of the linear and nonlinear parts are adjusted to maximize the correntropy between the model output and desired response.

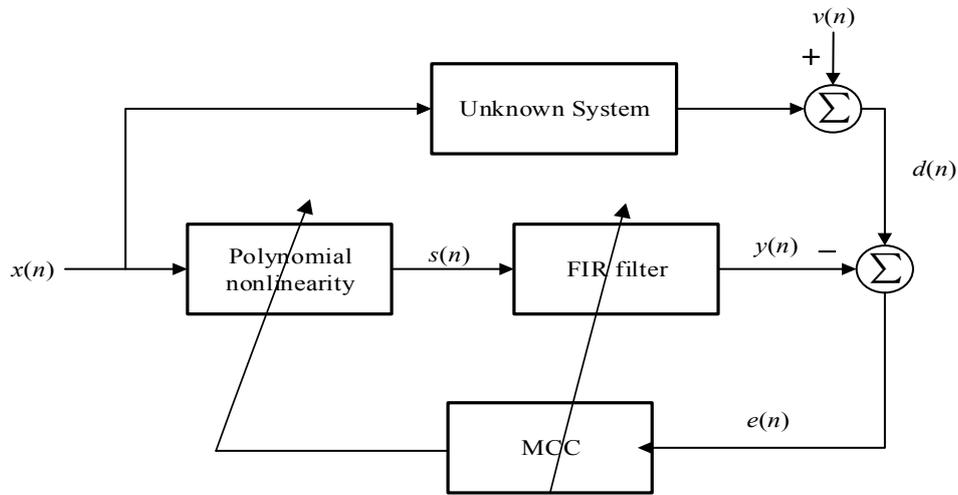


Figure 1. Structure of a Hammerstein adaptive filter under maximum correntropy criterion (MCC) criterion.

2.1. Correntropy

Correntropy is a nonlinear similarity measure between two signals. Given two random variables X and Y , the correntropy is [33–39]

$$V(X, Y) = E[\kappa(X, Y)] = \int \kappa(x, y) f_{XY}(x, y) dx dy \tag{1}$$

where $E[\cdot]$ denotes the expectation operator, $\kappa(\cdot, \cdot)$ is a *shift-invariant* Mercer kernel, and $f_{XY}(x, y)$ stands for the probability density function (PDF) of (X, Y) . The most widely used kernel in correntropy is the Gaussian kernel, given by

$$\kappa_{\sigma}(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{e^2}{2\sigma^2}\right) \tag{2}$$

where $e = x - y$, and σ stands for the kernel bandwidth. In this work, without being mentioned otherwise, the kernel function is a Gaussian kernel. In practical situations, the joint distribution of X and Y is usually unknown and only a finite number of data $\{(d(i), y(i))\}_{i=1}^K$ are available. In these cases, one can use a sample mean estimator of the correntropy:

$$\hat{V}_{N,\sigma}(X, Y) = \frac{1}{K} \sum_{i=1}^K \kappa_{\sigma}(d(i) - y(i)) \tag{3}$$

The optimization cost under MCC is thus

$$\max J_{MCC} = \sum_{i=1}^K \kappa_{\sigma}(e(i)) \tag{4}$$

where $e(i) = d(i) - y(i)$. We can evaluate the sensitivity (derivative) of the MCC cost J_{MCC} with respect to the error $e(i)$,

$$\frac{\partial J_{MCC}}{\partial e(i)} = \frac{-1}{\sqrt{2\pi\sigma^3}} \cdot \exp\left(\frac{-e^2(i)}{2\sigma^2}\right) e(i) \tag{5}$$

The derivative curves of $-J_{MCC}$ for different kernel widths are illustrated in Figure 2. As one can see, when the magnitude of error is very large, the derivative will become rather small especially for a smaller kernel width. Therefore, the MCC training is insensitive (hence robust) to a large error.

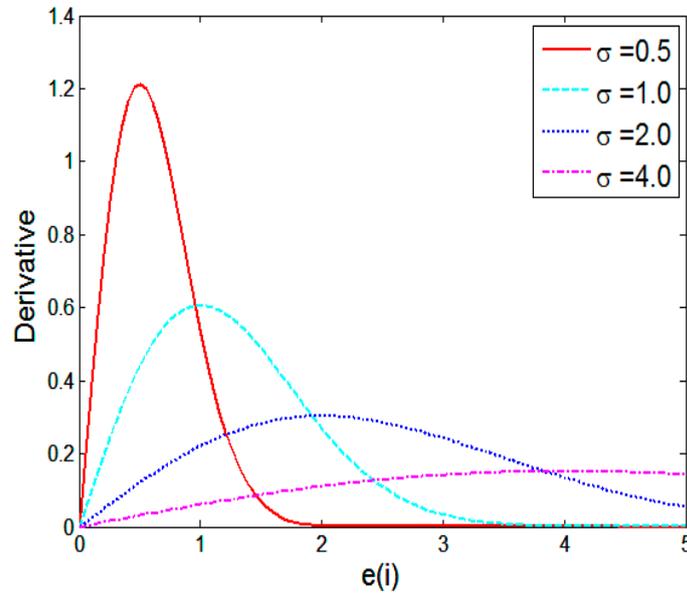


Figure 2. Derivative curves of $-J_{MCC}$ with respect to $e(i)$ for different kernel widths.

2.2. Hammerstein Adaptive Filtering

Assuming that the input-output mapping of the memoryless polynomial nonlinearity is

$$s(n) = p_1x(n) + p_2x^2(n) + \dots + p_Mx^M(n) \tag{6}$$

where M and p_M denote the polynomial order and the m -th order coefficient, Expression (6) can be rewritten as

$$s(n) = \mathbf{p}^T(n)\mathbf{x}_p(n) \tag{7}$$

where $\mathbf{x}_p(n) = [x(n) x^2(n) \dots x^M(n)]^T$ is the polynomial regressor, and $\mathbf{p}(n) = [p_1 p_2 \dots p_M]^T$ is the polynomial coefficient vector. The output of the FIR filter can be expressed as

$$y(n) = \mathbf{w}^T(n)s(n) \tag{8}$$

where $\mathbf{w}(n) = [w_0 w_1 \dots w_{N-1}]^T$ is the FIR weight vector, and $\mathbf{s}(n) = [s(n) s(n-1) \dots s(n-N+1)]^T$ is the FIR input vector, with N being the FIR memory size. Let $\mathbf{X}(n) = [\mathbf{x}_p(n) \mathbf{x}_p(n-1) \dots \mathbf{x}_p(n-N+1)]^T$. Then we have

$$s(n) = \mathbf{X}^T(n)\mathbf{p}(n) \tag{9}$$

Combining Equations (8) and (9) yields

$$y(n) = \mathbf{w}^T(n)s(n) = \mathbf{w}^T(n)\mathbf{X}^T(n)\mathbf{p}(n) \tag{10}$$

Assume that the unknown system that needs to be identified is also a Hammerstein system with parameter vectors $\mathbf{p}^* = [p_1^* \ p_2^* \ \dots \ p_M^*]^T$ and $\mathbf{w}^* = [w_0^* \ w_1^* \ \dots \ w_{N-1}^*]^T$. Then, the desired signal can be expressed as

$$d(n) = \mathbf{w}^{*T} \mathbf{X}^T(n) \mathbf{p}^* + v(n) \tag{11}$$

where $v(n)$ stands for an additive disturbance noise. The error signal can then be calculated as $e(n) = d(n) - \mathbf{w}^T(n) \mathbf{X}^T(n) \mathbf{p}(n)$. In the following, we derive an adaptive algorithm to estimate the Hammerstein parameter vectors using MCC instead of MSE as an optimization criterion. Let us consider the following cost function

$$\begin{aligned} J_{MCC}(\mathbf{p}, \mathbf{w}) &= \sum_{j=n-L+1}^n \kappa_{\sigma}(d(j), y(j)) \\ &= \frac{1}{\sqrt{2\pi\sigma}} \sum_{j=n-L+1}^n \exp\left(\frac{-e^2(j)}{2\sigma^2}\right) \end{aligned} \tag{12}$$

where $e(j) = d(j) - y(j)$, and L denotes the sliding data length. Then, a steepest ascent algorithm for estimating the polynomial coefficient vector can be derived as follows:

$$\begin{aligned} \frac{\partial J_{MCC}(\mathbf{p}, \mathbf{w})}{\partial \mathbf{p}(n)} &= \frac{1}{\sqrt{2\pi\sigma^3}} \cdot \sum_{j=n-L+1}^n \exp\left(\frac{-e^2(j)}{2\sigma^2}\right) e(j) \cdot \frac{\partial e(j)}{\partial \mathbf{p}(j)} \\ &= \frac{1}{\sqrt{2\pi\sigma^3}} \cdot \sum_{j=n-L+1}^n \exp\left(\frac{-e^2(j)}{2\sigma^2}\right) e(j) \cdot \mathbf{X}(j) \mathbf{w}(j) \end{aligned} \tag{13}$$

$$\mathbf{p}(n+1) = \mathbf{p}(n) + \mu_p \frac{\partial J_{MCC}(\mathbf{p}, \mathbf{w})}{\partial \mathbf{p}(n)} = \mathbf{p}(n) + \frac{\mu_p}{\sqrt{2\pi\sigma^3}} \cdot \sum_{j=n-L+1}^n \exp\left(\frac{-e^2(j)}{2\sigma^2}\right) e(j) \cdot \mathbf{X}(j) \mathbf{w}(j) \tag{14}$$

In a similar way, we propose the following weight update equation for the coefficients of the FIR filter:

$$\begin{aligned} \frac{\partial J_{MCC}(\mathbf{p}, \mathbf{w})}{\partial \mathbf{w}(n)} &= \frac{1}{\sqrt{2\pi\sigma^3}} \cdot \sum_{j=n-L+1}^n \exp\left(\frac{-e^2(j)}{2\sigma^2}\right) e(j) \cdot \frac{\partial e(j)}{\partial \mathbf{w}(j)} \\ &= \frac{1}{\sqrt{2\pi\sigma^3}} \cdot \sum_{j=n-L+1}^n \exp\left(\frac{-e^2(j)}{2\sigma^2}\right) e(j) \cdot \mathbf{X}^T(j) \mathbf{p}(j) \end{aligned} \tag{15}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w \frac{\partial J_{MCC}(\mathbf{p}, \mathbf{w})}{\partial \mathbf{w}(n)} = \mathbf{w}(n) + \frac{\mu_w}{\sqrt{2\pi\sigma^3}} \cdot \sum_{j=n-L+1}^n \exp\left(\frac{-e^2(j)}{2\sigma^2}\right) e(j) \cdot \mathbf{X}^T(j) \mathbf{p}(j) \tag{16}$$

In Equations (14) and (16), μ_p and μ_w are, respectively, step-sizes for polynomial nonlinearity subsystem and FIR subsystem. In this work, for simplicity we consider only the stochastic gradient based algorithm (i.e., $L = 1$). In this case, we have

$$\mathbf{p}(n+1) = \mathbf{p}(n) + \eta_p \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n) \cdot \mathbf{X}(n) \mathbf{w}(n) \tag{17}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta_w \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n) \cdot \mathbf{X}^T(n) \mathbf{p}(n) \tag{18}$$

where $\eta_p = \frac{\mu_p}{\sqrt{2\pi\sigma^3}}$, and $\eta_w = \frac{\mu_w}{\sqrt{2\pi\sigma^3}}$. The above update equations are referred to as the Hammerstein adaptive filtering algorithm under MCC criterion, whose pseudocodes are presented in Algorithm 1. The proposed algorithm is in form similar to the traditional Hammerstein adaptive filters under MSE criterion [7], but the step-sizes are different.

Algorithm 1: Hammerstein adaptive filtering Algorithm under MCC.

Parameters setting: μ_p, μ_w, σ

Initialization: $\mathbf{p}(0), \mathbf{w}(0)$

For $n = 1, 2, \dots$ **do**

- (1) $\mathbf{s}(n) = \mathbf{X}^T(n)\mathbf{p}(n)$
- (2) $y(n) = \mathbf{w}^T(n)\mathbf{s}(n)$
- (3) $e(n) = d(n) - y(n)$
- (4) $\mathbf{p}(n+1) = \mathbf{p}(n) + \eta_p \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n) \cdot \mathbf{X}(n)\mathbf{w}(n)$
- (5) $\mathbf{w}(n+1) = \mathbf{w}(n) + \eta_w \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n) \cdot \mathbf{X}^T(n)\mathbf{p}(n)$

End for

3. Convergence Analysis

3.1. Stability Analysis

Using the Taylor series expansion of the error $e(n + 1)$ around the instant n and keeping only the linear term, we have [4,7,40]

$$e(n+1) = e(n) + \left. \frac{\partial e(n)}{\partial \mathbf{p}(n)} \right|_{\mathbf{w}(n)=const} \Delta \mathbf{p}(n) + \left. \frac{\partial e(n)}{\partial \mathbf{w}(n)} \right|_{\mathbf{p}(n)=const} \Delta \mathbf{w}(n) + h.o.t \tag{19}$$

where *h.o.t* denotes higher-order terms. Combining Equations (11), (17) and (18), we can obtain

$$\frac{\partial e(n)}{\partial \mathbf{p}(n)} = -\mathbf{X}(n)\mathbf{w}(n) \tag{20}$$

$$\frac{\partial e(n)}{\partial \mathbf{w}(n)} = -\mathbf{X}^T(n)\mathbf{p}(n) \tag{21}$$

$$\Delta \mathbf{p}(n) = \eta_p \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n) \cdot \mathbf{X}(n)\mathbf{w}(n) \tag{22}$$

$$\Delta \mathbf{w}(n) = \eta_w \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) e(n) \cdot \mathbf{X}^T(n)\mathbf{p}(n) \tag{23}$$

Substituting Equations (20)–(23) in Equation (19), and after simple manipulation, we have

$$e(n+1) = \left[1 - \eta_p \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) \cdot \|X(n)w(n)\|^2 - \eta_w \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) \cdot \|X^T(n)p(n)\|^2 \right] e(n) \quad (24)$$

To ensure the stability of the proposed algorithm, we must assure that $|e(n+1)| \leq |e(n)|$, and hence

$$\left| 1 - \eta_p \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) \cdot \|X(n)w(n)\|^2 - \eta_w \cdot \exp\left(\frac{-e^2(n)}{2\sigma^2}\right) \cdot \|X^T(n)p(n)\|^2 \right| \leq 1 \quad (25)$$

which yields

$$0 < \eta_p \cdot \|X(n)w(n)\|^2 + \eta_w \cdot \|X^T(n)p(n)\|^2 \leq \frac{2}{\exp\left(\frac{-e^2(n)}{2\sigma^2}\right)} \quad (26)$$

Since $\exp\left(\frac{-e^2(n)}{2\sigma^2}\right) \leq 1$, the following condition guarantees convergence:

$$0 < \eta_p \cdot \|X(n)w(n)\|^2 + \eta_w \cdot \|X^T(n)p(n)\|^2 \leq 2 \quad (27)$$

Remark 1. The derived bound on step-sizes is only of theoretical importance as in general, Equation (27) cannot be verified in a practical situation. Similar theoretical results can be found in [7].

3.2. Steady-State Mean Square Performance

We denote $e_{pw}(n)$ the *a priori* error of the whole system, $e_p(n)$ the *a priori* error when only the nonlinear part is adapted while the linear filter is fixed, and $e_w(n)$ the *a priori* error when only the linear filter is adapted while the nonlinear part is fixed. Let $H_p = \lim_{n \rightarrow \infty} E[e_p^2(n)]$, $H_w = \lim_{n \rightarrow \infty} E[e_w^2(n)]$, and $H_{pw} = \lim_{n \rightarrow \infty} E[e_{pw}^2(n)]$ be the steady-state *excess mean square errors* (EMSEs). In addition, we denote

$$f(e(i)) = \exp\left(\frac{-e^2(i)}{2\sigma^2}\right) e(i) \quad (28)$$

Before evaluating the theoretical values of the steady-state EMSEs, we make the following assumptions:

- (A) The noise $v(n)$ is zero-mean, independent, identically distributed, and is independent of the input $X(n)$, $\hat{s}(n)$ and $e(n)$.
- (B) The *a priori* errors $e_p(n)$ and $e_w(n)$ are zero-mean Gaussian, and independent of the noise $v(n)$.
- (C) $\|X(n)w(n)\|^2$ and $\|X^T(n)p(n)\|^2$ are asymptotically uncorrelated with $f^2(e(n))$, that is

$$\lim_{n \rightarrow \infty} E\left[\|X(n)w(n)\|^2 f^2(e(n))\right] = Tr(R_{WX}) \lim_{n \rightarrow \infty} E\left[f^2(e(n))\right] \quad (29)$$

$$\lim_{n \rightarrow \infty} E\left[\|X^T(n)p(n)\|^2 f^2(e(n))\right] = Tr(R_{XP}) \lim_{n \rightarrow \infty} E\left[f^2(e(n))\right] \quad (30)$$

where $R_{WX} = E[(X(n)w(n))(X(n)w(n))^T]$ and $R_{XP} = E[(X^T(n)p(n))(X^T(n)p(n))^T]$ are the covariance matrices, and $Tr(\cdot)$ denotes the trace operator.

Remark 2. For the assumption (A), it is very common to assume that the noise is independent of the regression vector [41–43]. In addition, the noise is often restricted to be zero-mean, identically

distributed [33–35]. As discussed in [44,45], the assumption (B) is reasonable for long adaptive filters. Since $e_p(n)$ is the *a priori* error when only the nonlinear part is adapted while the linear filter is fixed, we have the approximation $\mathbf{w}^* \approx \mathbf{w}(n)$ such that $\mathbf{w}(n)$ is asymptotically uncorrelated with $f^2(e(n))$. Due to the independent assumption (A), $\mathbf{X}(n)$ is also asymptotically uncorrelated with $f^2(e(n))$. So $\|\mathbf{X}^T(n)\mathbf{w}(n)\|^2$ is asymptotically uncorrelated with $f^2(e(n))$. Similarly, $\|\mathbf{X}^T(n)\mathbf{p}(n)\|^2$ is asymptotically uncorrelated with $f^2(e(n))$. Therefore, the assumption (C) is rational.

When only the polynomial part with parameter vector \mathbf{p} is adapted, the error $e_p(n)$ is

$$e_p(n) = \mathbf{w}^{*T} \mathbf{X}^T(n) \mathbf{p}^* - \mathbf{w}^T(n) \mathbf{X}^T(n) \mathbf{p}(n) \approx \mathbf{w}^T(n) \mathbf{X}^T(n) \tilde{\mathbf{p}}(n) \tag{31}$$

where $\tilde{\mathbf{p}}(n) = \mathbf{p}^* - \mathbf{p}^T(n)$. In Equation (31), we use the approximation $\mathbf{w}^* \approx \mathbf{w}(n)$ at steady-state. From Equation (17), it follows easily that

$$\tilde{\mathbf{p}}(n+1) = \tilde{\mathbf{p}}(n) - \eta_p \cdot f(e(n)) \cdot \mathbf{X}(n) \mathbf{w}(n) \tag{32}$$

Squaring both sides of Equation (32), we have

$$\|\tilde{\mathbf{p}}(n+1)\|^2 = \|\tilde{\mathbf{p}}(n)\|^2 - 2\eta_p \cdot f(e(n)) \cdot e_p(n) + \eta_p^2 \cdot f^2(e(n)) \cdot \|\mathbf{X}(n) \mathbf{w}(n)\|^2 \tag{33}$$

Taking the expectations of the both sides of Equation (33) yields

$$E[\|\tilde{\mathbf{p}}(n+1)\|^2] = E[\|\tilde{\mathbf{p}}(n)\|^2] - 2\eta_p \cdot E[f(e(n)) \cdot e_p(n)] + \eta_p^2 \cdot E[f^2(e(n)) \cdot \|\mathbf{X}(n) \mathbf{w}(n)\|^2] \tag{34}$$

Assuming the filter is stable and attains the steady state, it holds

$$\lim_{n \rightarrow \infty} E[\|\tilde{\mathbf{p}}(n+1)\|^2] = \lim_{n \rightarrow \infty} E[\|\tilde{\mathbf{p}}(n)\|^2] \tag{35}$$

Combining Equations (34) and (35) and the above assumptions, we obtain

$$2 \cdot \lim_{n \rightarrow \infty} E[f(e(n)) \cdot e_p(n)] = \eta_p \text{Tr}(R_{wX}) \lim_{n \rightarrow \infty} E[f^2(e(n))] \tag{36}$$

In order to derive a theoretical value of the steady-state EMSE, we consider two cases below.

Case A. Gaussian Noise

Recalling that $e(n) = e_p(n) + v(n)$, and assuming that the noise $v(n)$ is zero-mean Gaussian, with variance ζ_v^2 , we get [34]

$$\lim_{n \rightarrow \infty} E[f(e(n)) \cdot e_p(n)] = \frac{\sigma^3 H_p}{(\sigma^2 + \zeta_v^2 + H_p)^{3/2}} \tag{37}$$

where σ_e^2 denotes the variance of the error, and $\sigma_e^2 = E[e_p^2(n)] + \zeta_v^2$. Similarly, we obtain [29]

$$\lim_{n \rightarrow \infty} E[f^2(e(n))] = \frac{\sigma^3 (H_p + \zeta_v^2)}{(\sigma^2 + 2\zeta_v^2 + 2H_p)^{3/2}} \tag{38}$$

Substituting Equations (37) and (38) into Equation (36), we have

$$2 \cdot \frac{\sigma^3 H_p}{(\sigma^2 + \zeta_v^2 + H_p)^{3/2}} = \eta_p \text{Tr}(R_{wX}) \frac{\sigma^3 (H_p + \zeta_v^2)}{(\sigma^2 + 2\zeta_v^2 + 2H_p)^{3/2}} \tag{39}$$

Therefore, the steady-state EMSE H_p satisfies

$$H_p = \frac{\eta_p \text{Tr}(R_{WX})}{2} \frac{(H_p + \zeta_v^2)(\sigma^2 + \zeta_v^2 + H_p)^{3/2}}{(\sigma^2 + 2\zeta_v^2 + 2H_p)^{3/2}} \tag{40}$$

Theorem 1. In a Gaussian noise environment and with the same step-size, the proposed nonlinear Hammerstein adaptive filter under MCC criterion has a smaller steady-state EMSE than under MSE criterion. As the kernel width increases, their values of the steady-state EMSE will become almost identical.

Proof. It can be shown that [34]

$$H_{p-MSE} = \frac{\eta_p \text{Tr}(R_{WX}) \zeta_v^2}{2 - \eta_p \text{Tr}(R_{WX})} \tag{41}$$

where H_{p-MSE} denotes the steady-state EMSE under MSE criterion. From Equation (40), we have

$$H_p = \frac{\varepsilon \eta_p \text{Tr}(R_{WX}) \zeta_v^2}{2 - \varepsilon \eta_p \text{Tr}(R_{WX})} \tag{42}$$

where $\varepsilon = \frac{(\sigma^2 + \zeta_v^2 + H_p)^{3/2}}{(\sigma^2 + 2\zeta_v^2 + 2H_p)^{3/2}}$. Since $\varepsilon < 1$, it holds

$$H_p < H_{p-MSE} \tag{43}$$

Further, as $\sigma \rightarrow \infty$, we have $H_p \rightarrow H_{p-MSE}$.

Case B. Non-Gaussian Noise

Taking the Taylor series expansion of $f(e(n))$ around $v(n)$ yields

$$\begin{aligned} f(e(n)) &= f(e_p(n) + v(n)) \\ &= f(v(n)) + f'(v(n))e_p(n) + \frac{1}{2} f''(v(n))e_p^2(n) + h.o.t \end{aligned} \tag{44}$$

with

$$\begin{aligned} f'(v(n)) &= \exp\left(\frac{-v^2(n)}{2\sigma^2}\right) \left(1 - \frac{v^2(n)}{\sigma^2}\right) \\ f''(v(n)) &= \exp\left(\frac{-v^2(n)}{2\sigma^2}\right) \left(\frac{v^3(n)}{\sigma^4} - \frac{3v(n)}{\sigma^2}\right) \end{aligned} \tag{45}$$

Under the assumptions (A) and (B), we get [34]

$$\lim_{n \rightarrow \infty} E[f(e(n)) \cdot e_p(n)] \approx E[f'(v(n))] H_p \tag{46}$$

$$\lim_{n \rightarrow \infty} E[f^2(e(n))] \approx E[f^2(v(n))] + E\left[f(v(n))f''(v(n)) + |f'(v(n))|^2\right] H_p \tag{47}$$

Substituting Equations (46) and (47) into Equation (36), we have

$$H_p = \frac{\eta_p Tr(R_{wX}) E[f^2(v(n))]}{2E[f'(v(n))] - \eta_p Tr(R_{wX}) E[f(v(n))f''(v(n)) + |f'(v(n))|^2]} \tag{48}$$

Further, substituting Equation (45) into Equation (48), we obtain

$$H_p = \frac{\eta_p Tr(R_{wX}) E\left[\exp\left(\frac{-v^2(n)}{\sigma^2}\right)v^2(n)\right]}{2E\left[\exp\left(\frac{-v^2(n)}{2\sigma^2}\right)\left(1 - \frac{v^2(n)}{\sigma^2}\right)\right] - \eta_p Tr(R_{wX}) E\left[\exp\left(\frac{-v^2(n)}{\sigma^2}\right)\left(1 + \frac{2v^4(n)}{\sigma^4} - \frac{5v^2(n)}{\sigma^2}\right)\right]} \tag{49}$$

When only the linear filter with parameter vector $w(n)$ is adapted, we get

$$2 \cdot \lim_{n \rightarrow \infty} E[f(e(n)) \cdot e_w(n)] = \eta_p Tr(R_{XP}) \lim_{n \rightarrow \infty} E[f^2(e(n))] \tag{50}$$

where $e_w(n) \approx \tilde{w}^T(n)X^T(n)p(n)$, $\tilde{w}(n) = w^* - w(n)$. For Gaussian noise case, we obtain

$$H_w = \frac{\eta_w Tr(R_{XP}) (\sigma^2 + \zeta_v^2)(\sigma^2 + \zeta_v^2 + H_w)^{3/2}}{2(\sigma^2 + 2\zeta_v^2 + 2H_w)^{3/2}} \tag{51}$$

In non-Gaussian environments, we have

$$H_w = \frac{\eta_w Tr(R_{XP}) E\left[\exp\left(\frac{-v^2(n)}{\sigma^2}\right)v^2(n)\right]}{2E\left[\exp\left(\frac{-v^2(n)}{2\sigma^2}\right)\left(1 - \frac{v^2(n)}{\sigma^2}\right)\right] - \eta_w Tr(R_{XP}) E\left[\exp\left(\frac{-v^2(n)}{\sigma^2}\right)\left(1 + \frac{2v^4(n)}{\sigma^4} - \frac{5v^2(n)}{\sigma^2}\right)\right]} \tag{52}$$

Theorem 2. The H_{pw} satisfies the following condition

$$H_{pw} \geq H_p + H_w \tag{53}$$

Proof. Using Equation (31), we derive

$$\begin{aligned} e_{pw}(n) &= w^{*T}X^T(n)p^* - w^T(n)X^T(n)p(n) \\ &= (w^{*T}X^T(n)p^* - w^{*T}(n)X^T(n)p(n)) + (w^{*T}X^T(n)p(n) - w^T(n)X^T(n)p(n)) \\ &\approx w^T(n)X^T(n)\tilde{p}(n) + \tilde{w}^T(n)X^T(n)p(n) \\ &= e_p(n) + e_w(n) \end{aligned} \tag{54}$$

It follows that

$$\begin{aligned} H_{pw} &= \lim_{n \rightarrow \infty} E[e_{pw}^2(n)] \approx \lim_{n \rightarrow \infty} E[(e_p(n) + e_w(n))^2] \\ &= \lim_{n \rightarrow \infty} E[e_p^2(n)] + 2 \lim_{n \rightarrow \infty} E[e_p(n)e_w(n)] + \lim_{n \rightarrow \infty} E[e_w^2(n)] \\ &= H_p + H_w + 2H_{cross} \end{aligned} \tag{55}$$

where $H_{cross} = \lim_{n \rightarrow \infty} E[e_p(n)e_w(n)]$ stands for the *cross-EMSE* and $H_{cross} \geq 0$ ($H_{cross} = 0$ when $e_p(n)$ and $e_w(n)$ are statistically independent and zero mean) [7]. Therefore, $H_{pw} \geq H_p + H_w$, which completes the proof.

4. Simulation Results

Now, we present simulation results to demonstrate the performance of the Hammerstein adaptive filtering under MCC. In order to show the performance of the proposed algorithm in non-Gaussian noises, we adopt the alpha-stable distribution to generate the disturbance noise, whose characteristic function is [32,46]

$$f(t) = \exp\{j\delta t - \gamma |t|^\alpha [1 + j\beta \operatorname{sgn}(t)S(t, \alpha)]\} \quad (56)$$

in which

$$S(t, \alpha) = \begin{cases} \tan \frac{\alpha\pi}{2} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} \log |t| & \text{if } \alpha = 1 \end{cases} \quad (57)$$

where $\alpha \in (0, 2]$ denotes the characteristic factor, $-\infty < \delta < +\infty$ is the location parameter, $\beta \in [-1, 1]$ stands for the symmetry parameter, and $\gamma > 0$ is the dispersion parameter. The characteristic factor α measures the tail heaviness of the distribution. The smaller α is, the heavier the tail is. In addition, γ measures the dispersion of the distribution. The distribution is symmetric about its location δ when $\beta = 0$. Such a distribution is called a *symmetric alpha-stable (SaS)* distribution. The parameters vector of the noise model is defined as $V = (\alpha, \beta, \gamma, \delta)$.

In the simulations below, the input signal considered is a colored signal obtained from the following equation:

$$x(n) = ax(n-1) + \sqrt{1-a^2} \xi(n) \quad (58)$$

with $a = 0.95$, and $\xi(n)$ being a white Gaussian signal of unit variance. In addition, the coefficient vectors are initialized with the first coefficient equal to 1 and the others equal to zero [7].

4.1. Experiment 1

First, we consider an unknown Hammerstein system with parameter vectors $\mathbf{p}^* = [1, 0.6]$, $\mathbf{w}^* = [1, 0.6, 0.1, -0.2, -0.06, 0.04, 0.02, -0.03, -0.02, 0.01]$. Thus, $M = 2$, $N = 10$. The kernel width σ is 1.0. The noise vector V is set at $(1.2, 0, 0.6, 0)$, and the noise signal is shown in Figure 3. Simulation results are averaged over 100 independent Monte Carlo runs, and in each simulation, 15,000 iterations are run to ensure the algorithm will reach the steady state, and the steady-state MSE is obtained as an average over the last 2000 iterations. The step-sizes are set at $\mu_p = \mu_w = 0.005$ and $\mu_p = 0.01$, $\mu_w = 0.01$ for MSE and MCC, respectively. Figure 4 shows the average convergence curves under MCC and MSE. As we can see, the Hammerstein adaptive filtering under MCC criterion achieves faster convergence speed and lower steady-state testing MSE than under MSE criterion. Here the testing MSE is evaluated on a test set with 100 samples.

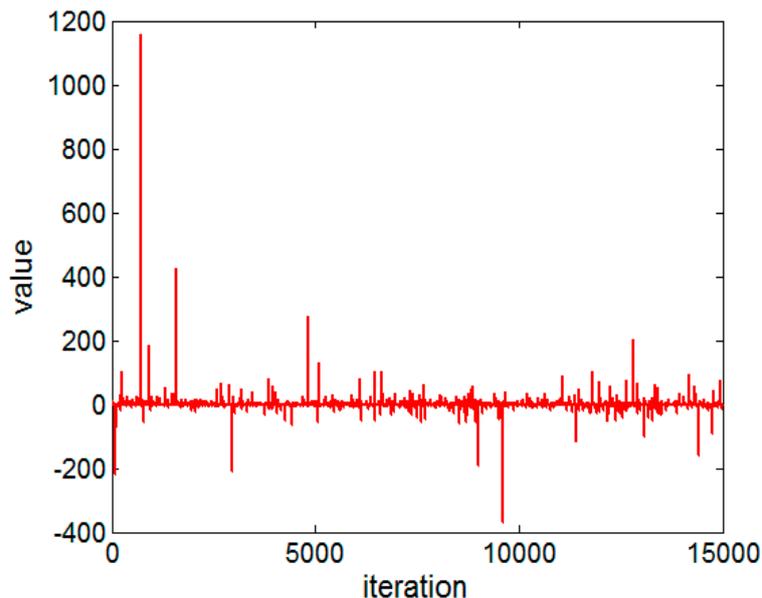


Figure 3. A typical sequence of the alpha-stable noise with $V = (1.2, 0, 0.6, 0)$.

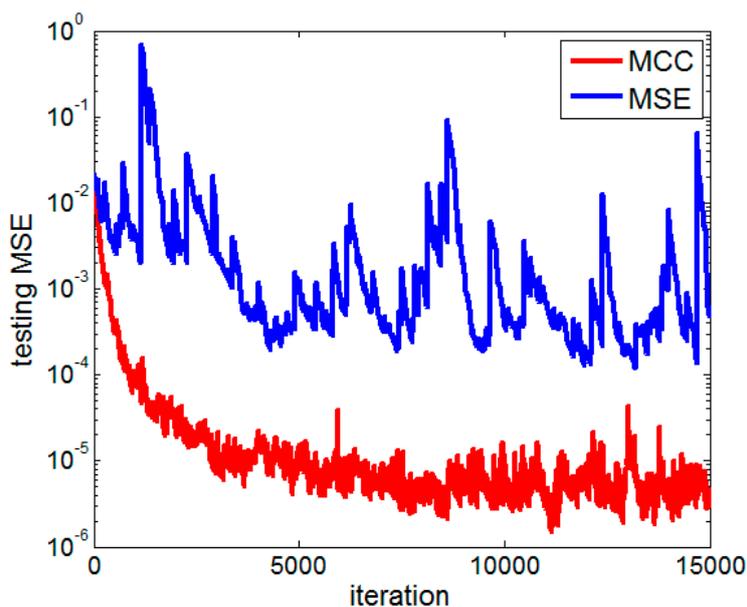


Figure 4. Convergence curves under maximum correntropy criterion (MCC) and mean square error (MSE) (for unknown system with polynomial nonlinearity).

Second, we investigate the performance of the algorithms with different noise parameters. The steady-state MSEs with different γ (0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6) and different α (0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2.0) are shown in Figures 5 and 6, respectively. We observe: (1) In most cases, the new algorithm performs better and achieves a lower steady-state MSE compared with the Hammerstein adaptive filtering under MSE criterion; (2) When α is close to 2.0, the Hammerstein adaptive filtering under MSE criterion can achieve better performance than under MCC criterion. The main reason for this is that, when $\alpha \approx 2.0$, the noise will be approximately Gaussian. Simulation results suggest that the proposed algorithm is particularly useful for identifying a Hammerstein system in non-Gaussian noises.

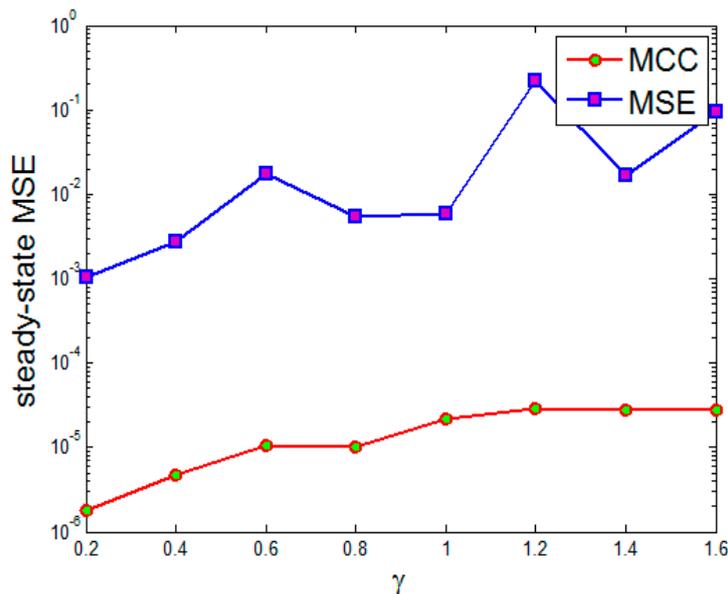


Figure 5. Steady-state mean square error (MSE) with different γ ($\alpha = 1.2$).

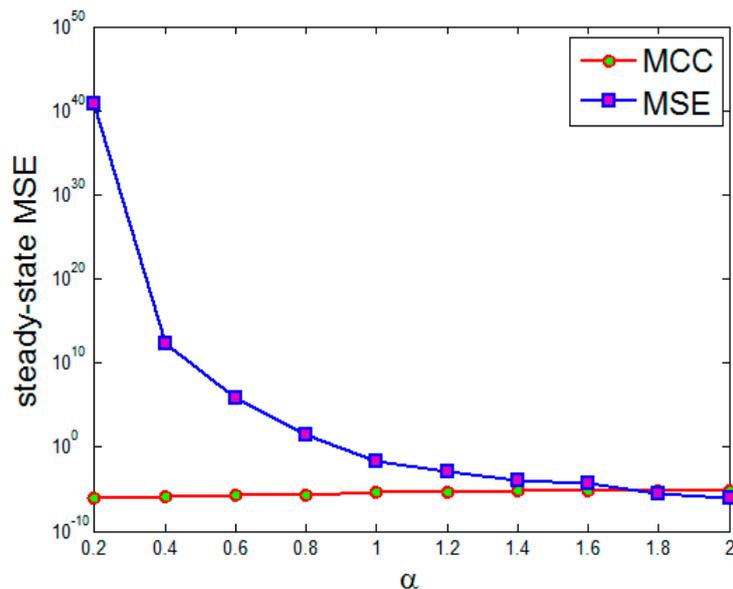


Figure 6. Steady-state mean square error (MSE) with different α ($\gamma = 0.6$).

4.2. Experiment 2

The second experiment is drawn from [47]. The nonlinear dynamic system is composed of two blocks. The first block is a non-polynomial nonlinearity

$$s(n) = \sqrt[3]{x(n)} \tag{59}$$

while the second block is an FIR filter with weight vector

$$h = [1 \ 0.75 \ 0.5 \ 0.25 \ 0 \ -0.25]^T \tag{60}$$

The noise vector V is set at (1.0, 0, 0.8, 0) (see Figure 7 for a typical sequence of the noise), and the polynomial order M and the FIR memory size N are set at 3 and 6, respectively. Simulation results are

averaged over 50 independent Monte Carlo runs, and in each simulation, 30,000 iterations are run to ensure the algorithm will reach the steady state, and the steady-state MSE is obtained as an average over the last 2000 iterations. The testing MSE is evaluated on a test set with 100 samples. Figure 8 demonstrates the convergence curves under MCC and MSE. For both adaptive filtering algorithms, the step-sizes are set at $\mu_p = 0.005$, $\mu_w = 0.015$. It can be seen that, the Hammerstein adaptive filter under MCC criterion performs better (say, with faster convergence speed and smaller mismatch error) than under MSE criterion.

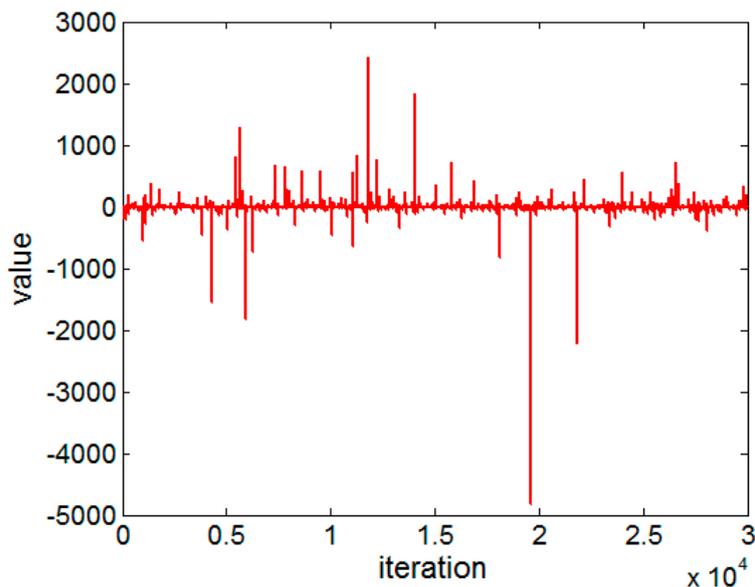


Figure 7. A typical sequence of the alpha-stable noise with $V = (1.0, 0, 0.8, 0)$.

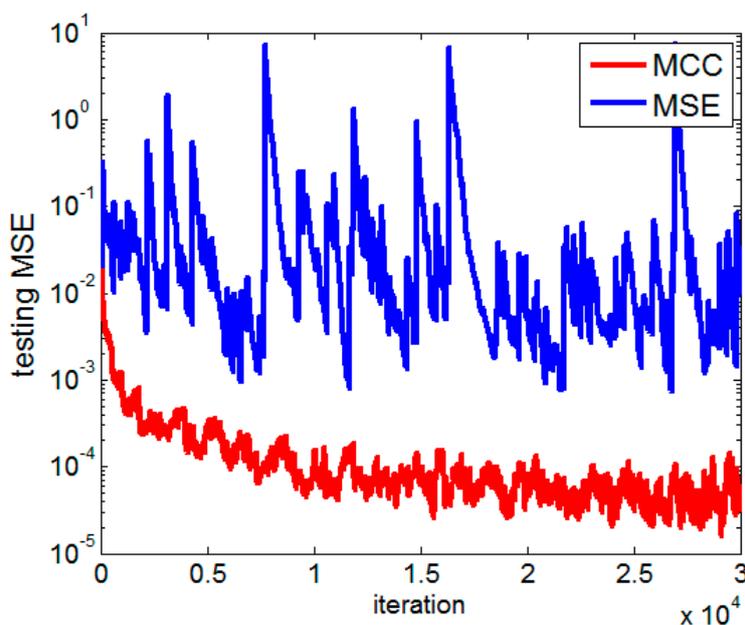


Figure 8. Convergence curves under maximum correntropy criterion (MCC) and mean square error (MSE) (for unknown system with non-polynomial nonlinearity).

Finally, we show the steady-state performance of the algorithms with different kernel widths σ (0.01, 1.0, 2.0, 3.0, 4.0, 5.0). Simulation results are shown in Figure 9. As we can see, the kernel width

has significant influence on the performance of the proposed algorithm. In this example, the lowest steady-state MSE is obtained when $\sigma = 1.0$.

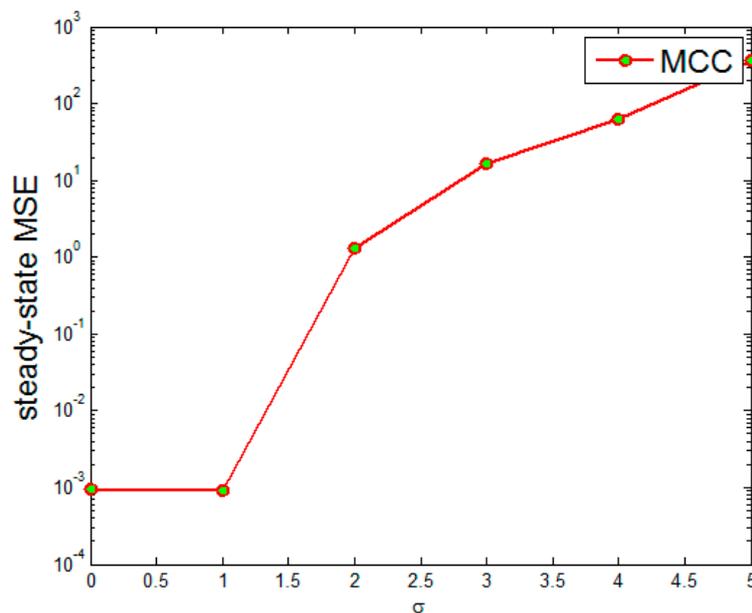


Figure 9. Steady-state mean square error (MSE) with different kernel widths.

5. Conclusions

The MCC has been successfully applied in domains of machine learning and signal processing due to its strong robustness in impulsive non-Gaussian situations. In this work, we develop a robust Hammerstein adaptive filter under MCC criterion. Different from the traditional Hammerstein adaptive filtering algorithms, the new algorithm use the MCC instead of the well-known MSE as the adaptation criterion, which can achieve desirable performance especially in impulsive noises. Based on [7,31], we carry out the convergence analysis, and obtain some important theoretical results. Simulation examples confirm the excellent performance of the proposed algorithm. How to verify the derived theoretical results is an interesting topic for future study.

Acknowledgments

This work was supported by 973 Program (No. 2015CB351703) and National Natural Science Foundation of China (No. 61372152, No. 61271210).

Author Contributions

Zongze Wu derived the algorithm and wrote the draft; Siyuan Peng proved the convergence properties and performed the simulations; Badong Chen proposed the main idea and polished the language; Haiquan Zhao was in charge of technical checking. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Ljung, L. *System Identification—Theory for the User*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999.
2. Sayed, A.H. *Fundamentals of Adaptive Filtering*; Wiley: Hoboken, NJ, USA, 2003.
3. Wiener, N. *Nonlinear Problems in Random Theory*; MIT Press: Cambridge, MA, USA, 1958.
4. Scarpiniti, M.; Comminiello, D.; Parisi, R.; Uncini, A. Nonlinear spline adaptive filtering. *Signal Process.* **2013**, *93*, 772–783.
5. Bai, E.W. Frequency domain identification of Wiener models. *Automatica* **2003**, *39*, 1521–1530.
6. Ogunfunmi, T. *Adaptive Nonlinear System Identification: The Volterra and Wiener Model Approaches*; Springer: Berlin/Heidelberg, Germany, 2007.
7. Scarpiniti, M.; Comminiello, D.; Parisi, R.; Uncini, A. Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties. *Signal Process.* **2014**, *100*, 112–123.
8. Ortiz Batista, E.L.; Seara, R. A new perspective on the convergence and stability of NLMS Hammerstein filters. In Proceedings of the IEEE 8th International Symposium on Image and Signal Processing and Analysis (ISPA), Trieste, Italy, 4–6 September 2013; pp. 343–348.
9. Bai, E.W.; Li, D. Convergence of the iterative Hammerstein system identification algorithm. *IEEE Trans. Autom. Control* **2004**, *49*, 1929–1940.
10. Umoh, I.; Ogunfunmi, T. An Affine-Projection-Based Algorithm for Identification of Nonlinear Hammerstein Systems. *Signal Process.* **2010**, *90*, 2020–2030.
11. Umoh, I.; Ogunfunmi, T. An Adaptive Nonlinear Filter for System Identification. *EURASIP J. Adv. Signal Process.* **2009**, *2009*, doi:10.1155/2009/859698.
12. Umoh, I.; Ogunfunmi, T. An adaptive algorithm for Hammerstein filter system identification. In Proceedings of the EURASIP European Signal Processing Conference, Lausanne, Switzerland, 25–29 August 2008; pp. 1–5.
13. Greblicki, W. Continuous time Hammerstein system identification. *IEEE Trans. Autom. Control* **2000**, *45*, 1232–1236.
14. Jeraj, J.; Mathews, V.J. Stochastic mean-square performance analysis of an adaptive Hammerstein filter. *IEEE Trans. Signal Process.* **2006**, *54*, 2168–2177.
15. Voros, J. Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities. *IEEE Trans. Autom. Control* **1999**, *44*, 2145–2149.
16. Greblicki, W. Stochastic approximation in nonparametric identification of Hammerstein systems. *IEEE Trans. Autom. Control* **2002**, *47*, 1800–1810.
17. Ding, F.; Liu, X.P.; Liu, G. Identification methods for Hammerstein nonlinear systems. *Dig. Signal Process.* **2011**, *21*, 215–238.
18. Jeraj, J.; Matthews, V.J. A stable adaptive Hammerstein filter employing partial orthogonalization of the input signals. *IEEE Trans. Signal Process.* **2006**, *54*, 1412–1420.
19. Nordsjo, A.E.; Zetterberg, L.H. Identification of certain time-varying nonlinear Wiener and Hammerstein systems. *IEEE Trans. Signal Process.* **2001**, *49*, 577–592.
20. Atiya, A.; Parlos, A. Nonlinear system identification using spatiotemporal neural networks. In Proceedings of the International Joint Conference on Neural Networks, Baltimore, MD, USA, 7–11 June 1992; pp. 504–509.

21. Volterra, V. *Theory of Functionals and of Integral and Integro-Differential Equations*; Courier Corporation: North Chelmsford, MA, USA, 2005.
22. Principe, J.C.; Liu, W.; Haykin, S. *Kernel Adaptive Filtering: A Comprehensive Introduction*; Wiley: Hoboken, NJ, USA, 2011.
23. Chen, B.; Zhao, S.; Zhu, P.; Principe, J.C. Quantized kernel least mean square algorithm. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 22–32.
24. Chen, B.; Zhao, S.; Zhu, P.; Principe, J.C. Quantized kernel recursive least squares algorithm. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 1484–1491.
25. Chen, B.; Zhao, S.; Zhu, P.; Principe, J.C. Mean square convergence analysis of the kernel least mean square algorithm. *Signal Process.* **2012**, *92*, 2624–2632.
26. Bai, E.W. A blind approach to Hammerstein model identification. *IEEE Trans. Signal Process.* **2002**, *50*, 1610–1619.
27. Stenger, A.; Kellermann, W. Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling. *Signal Process.* **2000**, *80*, 1747–1760.
28. Shi, K.; Ma, X.; Zhou, G.T. An efficient acoustic echo cancellation design for systems with long room impulses and nonlinear loudspeakers. *Signal Process.* **2009**, *89*, 121–132.
29. Scarpiniti, M.; Comminiello, 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.
30. Kailath, T.; Sayed, A.H.; Hassibi, B. *Linear Estimation*; Prentice Hall: Upper Saddle River, NJ, USA, 2000.
31. Plataniotis, K.N.; Androustos, D.; Venetsanopoulos, A.N. Nonlinear filtering of non-Gaussian noise. *J. Intell. Robot. Syst.* **1997**, *19*, 207–231.
32. Weng, B.; Barner, K.E. Nonlinear system identification in impulsive environments. *IEEE Trans. Signal Process.* **2005**, *53*, 2588–2594.
33. Chen, B.; Zhu, Y.; Hu, J.; Principe, J.C. *System Parameter Identification: Information Criteria and Algorithms*; Elsevier: Amsterdam, The Netherlands, 2013.
34. Chen, B.; Xing, L.; Liang, J.; Zheng, N.; Principe, J.C. Steady-state Mean-square Error Analysis for Adaptive Filtering under the Maximum Correntropy Criterion. *IEEE Signal Process. Lett.* **2014**, *21*, 880–884.
35. Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer: New York, NY, USA, 2010.
36. Singh, A.; Principe, J.C. Using Correntropy as a cost function in linear adaptive filters. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), Atlanta, GA, USA, 14–19 June 2009; pp. 2950–2955.
37. Chen, B.; Principe, J.C. Maximum correntropy estimation is a smoothed MAP estimation. *IEEE Signal Process. Lett.* **2012**, *19*, 491–494.
38. Zhao, S.; Chen, B.; Principe, J.C. Kernel adaptive filtering with maximum correntropy criterion. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 31 July–5 August 2011; pp. 2012–2017.
39. Ogunfunmi, T.; Paul, T. The Quaternion Maximum Correntropy Algorithm. *IEEE Trans. Circuits Syst. -II (TCAS-II)* **2015**, *62*, 598–602.

40. Mandic, D.P.; Hanna, A.I.; Razaz, M. A normalized gradient descent algorithm for nonlinear adaptive filters using a gradient adaptive step size. *IEEE Signal Process. Lett.* **2001**, *8*, 295–297.
41. Al-Naffouri, T.Y.; Sayed, A.H. Adaptive filters with error non-linearities: Mean-square analysis and optimum design. *EURASIP J. Appl. Signal Process.* **2001**, *4*, 192–205.
42. Lin, B.; He, R.; Wang, X.; Wang, B. The steady-state mean-square error analysis for least mean p-order algorithm. *IEEE Signal Process. Lett.* **2009**, *16*, 176–179.
43. Yousef, N.R.; Sayed, A.H. A unified approach to the steady-state and tracking analysis of adaptive filters. *IEEE Trans. Signal Process.* **2001**, *49*, 314–324.
44. Duttweiler, D.L. Adaptive filter performance with nonlinearities in the correlation multiplier. *IEEE Trans. Acoust. Speech Signal Process.* **1982**, *30*, 578–586.
45. Mathews, V.J.; Cho, S.H. Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm. *IEEE Trans. Acoust. Speech Signal Process.* **1987**, *35*, 450–454.
46. Shao, M.; Nikias, C.L. Signal processing with fractional lower order moments: Stable processes and their applications. *Proc. IEEE* **1993**, *81*, 986–1010.
47. Hasiewicz, Z.; Pawlak, M.; Sliwinski, P. Nonparametric identification of nonlinearities in block-oriented systems by orthogonal wavelets with compact support. *IEEE Trans. Circuits Syst. I Regul.* **2005**, *52*, 427–442.

© 2015 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 license (<http://creativecommons.org/licenses/by/4.0/>).