

Article

A Smoothed Algorithm with Convergence Analysis under Generalized Maximum Correntropy Criteria in Impulsive Interference

Hua Qu ¹, Youwei Shi ^{1,*} and Jihong Zhao ^{1,2}

¹ School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; qh@xjtu.edu.cn (H.Q.); zhaojihong@xjtu.edu.cn (J.Z.)

² School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710061, China

* Correspondence: shiyw21@stu.xjtu.edu.cn; Tel.: +86-159-2998-0762

Received: 18 October 2019; Accepted: 8 November 2019; Published: 11 November 2019



Abstract: The generalized maximum correntropy criterion (GMCC) algorithm is computationally simple and robust against impulsive noise but it suffers from slow convergence speed as it is derived and based on stochastic gradient, which only use the current data sample. In order to deal with this issue, a smoothed GMCC algorithm (SGMCC) is proposed. In the SGMCC algorithm, instead of taking the exponential weighted average of gradient vector to approximate the expectation of the gradient vector, we take the exponential weighted average of the variable step-size so that the SGMCC algorithm can be viewed as a sign GMCC algorithm with smoothed variable step-size. Moreover, convergence performance analyses are derived in terms of variable step-size, mean-square stability and steady-state behavior to demonstrate the robustness of the proposed algorithm. At last, simulation comparisons show that the proposed algorithm is robust against impulsive noise and converges fast with lower computational complexity. Also, for the steady-state behavior, simulation results verify that the simulated value matches well with the theoretical one.

Keywords: generalized correntropy; exponential weighted average; robust adaptive filtering algorithm; impulsive noise

1. Introduction

Information theoretic learning (ITL) [1] methods have been shown to be efficient approaches in non-Gaussian signal processing due to their robustness against impulsive noise. The maximum correntropy criterion (MCC) [2,3] is one of the most popular optimization criteria in ITL due to its simplicity and robustness. Recently, it has been successfully applied in various signal processing scenarios, particularly the adaptive filtering [4–10].

The introduction of the correntropy as a cost function into adaptive filters was proposed in Reference [11]. The theoretical analysis in Reference [12] has shown that the steady-state excess mean square error (EMSE) of the MCC algorithm is controlled by the step-size and the kernel width. Various kernel width selection methods for the MCC algorithm have been investigated. Adaptive kernel width adjusting methods were proposed in References [6,7] to improve the convergence speed of the MCC algorithm.

Just like the least mean square (LMS)-type algorithms, the MCC algorithm with fixed step-size is insufficient to achieve a good tradeoff between fast convergence and low steady-state misadjustment. The adaptive variable step-size technique is a promising way to deal with the conflicting requirements of faster learning speed and lower steady-state misadjustment error. Many varieties of variable step-size LMS-type algorithms have been proposed to improve the convergence performance [13–16].

However, many variable step-size methods based on instantaneous error cannot be directly applied in the MCC algorithm. Of course, just a few variable step-size methods for the MCC algorithm have been developed to improve the convergence performance in recent years. A convex combination of two MCC algorithms with different step-sizes was proposed in Reference [4]. The mixing factor in this method was adaptively updated so that the MCC algorithm with larger step-size can improve the convergence speed at the beginning while the other with smaller step-size can achieve a lower misadjustment at steady state. The combinational approach can achieve desirable convergence performance but its computational complexity was more than two times higher than that of the standard MCC algorithm since two MCC algorithms were adopted in parallel. A curvature based variable step-size method for the MCC algorithm was proposed in Reference [17]. This developed method can improve the convergence speed at the initial stage especially when the weight vector is far away from the optimal solution.

In recent years, a generalized maximum correntropy criterion (GMCC) has been proposed, which adopts the generalized Gaussian density [18–20] function as the kernel function (not necessarily a Mercer kernel [21]) and the type of this correntropy is called the generalized correntropy. Similar to the original correntropy with Gaussian kernel, the generalized correntropy can also be used as an optimization cost in the estimation-related problems. Under the GMCC criterion, a stochastic-gradient based adaptive filtering algorithm, called the GMCC algorithm, was developed [21]. We can see that the GMCC algorithm was derived and based on stochastic gradient which only uses the current data sample. Although the GMCC algorithm is computationally simple and robust in the presence of impulsive noise, it suffers from high steady-state mean square deviation (MSD), which also can be verified from the next simulation.

In this paper, a smoothed GMCC algorithm (called SGMCC) is proposed to improve the performance of steady-state MSD in the presence of impulsive noise. To avoid the high steady-state MSD caused by the stochastic gradient in the GMCC algorithm, we take the GMCC algorithm as a sign algorithm with variable step-size. Then, we take the exponential weighted average of the variable step-size rather than the gradient vector, which contributes to reduce the computational complexity and further improve the convergence speed. At last, we present the convergence performance analyses of the proposed algorithm to demonstrate its robustness. The main contributions of our paper can be summarized as follows:

1. We propose a novel SGMCC algorithm, which can improve the performance of steady-state MSD and is robust against impulsive noise.
2. To demonstrate the robustness of the proposed algorithm, we derive the convergence performance analyses from three aspects including the variable step-size, mean-square stability and steady-state behavior.
3. Simulation results demonstrate that the proposed algorithm is robust against impulsive noise and computationally simple. Also, the SGMCC algorithm owns faster convergence speed compared with other robust adaptive algorithms.

The rest of the paper is organized as follows. In Section 2, after a brief review of the original GMCC algorithm, a SGMCC algorithm is proposed. In Section 3, the convergence performances of the SGMCC algorithm are studied. In Section 4, Monte-Carlo simulation results are presented to confirm the desirable performance of the proposed algorithm. Finally, a conclusion is given in Section 5.

2. SGMCC Algorithm

Assume the received signal d_n is obtained from an unknown system \mathbf{W}_*

$$d_n = \mathbf{W}_*^T \mathbf{X}_n + v_n \quad (1)$$

where $y_n = \mathbf{W}_*^T \mathbf{X}_n$ denotes the output of the unknown system, \mathbf{X}_n is the $M \times 1$ dimension input signal vector at n th time index and \mathbf{W}_* denotes the $M \times 1$ dimension unknown system vector that we wish to estimate. Here, $(\cdot)^T$ means the vector transpose operator.

In Equation (1), v_n denotes the Gaussian mixture distribution noise, which is modeled as a random variable of 2-components Gaussian mixture distribution with mixed coefficient p_r . Its probability density function can be expressed as follows:

$$MG(\sigma_1, \sigma_2, p_r) = (1 - p_r)N(0, \sigma_1^2) + p_r N(0, \sigma_2^2) \quad (2)$$

where the first Gaussian distribution $N(0, \sigma_1^2)$ generates the normal Gaussian noise with probability $(1 - p_r)$, the second Gaussian distribution $N(0, \sigma_2^2)$ generates impulsive noise with probability p_r . Usually, $\sigma_1 \ll \sigma_2$ and the impulsive noise occurrence probability p_r is set as a number that is near but larger than 0. σ_1^2 and σ_2^2 denotes the Gaussian noise variance and the impulsive noise variance, respectively.

2.1. Review of the GMCC Algorithm

As it is well-known that MCC has been successfully applied in various non-Gaussian signal processing matters due to its robustness properties [5,6,8,22]. As a non-linear local similarity measure between two random variables x and y , the correntropy is defined by [2,3]:

$$V_\sigma(x, y) = E[\kappa_\sigma(x, y)] \quad (3)$$

where $\kappa_\sigma(\cdot)$ denotes a shift-invariant Mercer kernel. The kernel function is a key factor of the correntropy that dramatically affects the similarity of two random variables. In most case, the Gaussian density function is used as kernel function, which is defined as follows:

$$\kappa_\sigma(x, y) = \exp\left(-\frac{|x - y|^2}{\sigma^2}\right) \quad (4)$$

However, for all non-Gaussian signal processing applications, the Gaussian density function is not always the best kernel function. In order to deal with this issue and take full advantage of the potential robustness properties of the correntropy, the generalized Gaussian density function was proposed as a kernel function of the correntropy in Reference [21]. The generalized Gaussian density function with zero-mean is given by References [18,19] as follows:

$$G_{\alpha, \beta}(x, y) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left|\frac{x - y}{\beta}\right|^\alpha\right) \quad (5)$$

where $\Gamma(\cdot)$ is the gamma function, $\alpha > 0$ denotes the shape parameter and $\beta > 0$ denotes the scale parameter. For simplicity, α is usually set as an integer value. Note that the Gaussian kernel is a special case of the generalized Gaussian density function when $\alpha = 2$.

In practice, the distribution of the data is usually unknown and only a finite number of samples are available. It is common to use these samples of the data to approximate its expectation, so the generalized correntropy can be estimated by [21]:

$$\hat{V}_{\alpha, \beta}(X, Y) = \frac{1}{N} \sum_{i=1}^N G_{\alpha, \beta}(x_i, y_i) \quad (6)$$

Just like the correntropy, the generalized correntropy can be used as a cost function in adaptive filters. Hence, the generalized correntropy between the desired signal d_n and the filter output y_n can be shown as a cost function by [21]:

$$J_{GMCC} = E [G_{\alpha,\beta} (e_n)] = \gamma_{\alpha,\beta} E [\exp (-\lambda |e_n|^\alpha)] \quad (7)$$

where $e_n = d_n - y_n$ is the instantaneous prediction error, $\gamma_{\alpha,\beta} = \alpha / (2\beta\Gamma(1/\alpha))$ represents the normalization constant and $\lambda = \beta^{-\alpha}$ denotes the kernel parameter.

The goal of the adaptive filtering algorithm is to find an estimator of the unknown system vector which maximizes the generalized correntropy of e_n . The optimal solution of this cost maximization problem can be solved by a stochastic-gradient based adaptive algorithm, called GMCC algorithm [21]. The update expression of the weight vector in the GMCC algorithm can be derived as [21]

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta \exp (-\lambda |e_n|^\alpha) |e_n|^{\alpha-1} \text{sign} (e_n) \mathbf{X}_n \quad (8)$$

where η is the step-size parameter and $\text{sign}(\cdot)$ is a universal sign function. It is obvious that the GMCC algorithm can be viewed as a sign algorithm with variable step-size $\eta \exp (-\lambda |e_n|^\alpha) |e_n|^{\alpha-1}$.

2.2. SGMCC Algorithm

As mentioned above, although the GMCC algorithm is robust and computationally simple, it suffers from high steady-state MSD as it is derived and based on stochastic gradient which only uses the current data sample. The instantaneous prediction error e_n fluctuates randomly with the background noise v_n and the input vector \mathbf{X}_n . The random fluctuations of e_n and \mathbf{X}_n can disturb the update of the weight vector and thus lead to a slow convergence speed.

To deal with this issue, it is common practice to take the average of the latest N sample data to approximate the expectation of the gradient vector, which can statistically reduce the adverse effects caused by the randomness of e_n and \mathbf{X}_n . Hence, the update expression of the weight vector in Equation (8) can be rewritten as follows:

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \frac{\eta}{N} \sum_{i=n-N+1}^n \exp (-\lambda |e_i|^\alpha) |e_i|^{\alpha-1} \text{sign} (e_i) \mathbf{X}_i, \quad (9)$$

where $e_i = d_i - \mathbf{W}_n^T \mathbf{X}_i$ is the prediction error at n th iteration for the input vector \mathbf{X}_i . It is worth noting that the calculation of each instantaneous prediction error e_i ($i = n - N + 1, \dots, n$) is based on the present weight vector \mathbf{W}_n , so the weight update in Equation (9) needs N times of computational cost and storage cost than that of Equation (8).

In order to reduce the computational cost and the storage cost, the exponential weighted average of the gradient vector can be adopted to approximate the expectation of the gradient vector, which is shown as follows:

$$\bar{\mathbf{p}}_n = \theta \bar{\mathbf{p}}_{n-1} + (1 - \theta) \exp (-\lambda |e_n|^\alpha) |e_n|^{\alpha-1} \text{sign} (e_n) \mathbf{X}_n, \quad (10)$$

where $\bar{\mathbf{p}}_n$ is the exponential weighted average of the GMCC gradient vector with a smoothing factor θ ($0 < \theta < 1$). In this case, the update expression of the weight vector can be further expressed as

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta \bar{\mathbf{p}}_n. \quad (11)$$

Although computational cost of Equation (11) is much lower than that of Equation (9), it still needs extra N multiplications and additions than that of Equation (8). In order to reduce the computational cost derived from gradient vector, we take the GMCC algorithm as a variable step-size sign algorithm by taking the exponential weighted average of the variable step-size instead of that of gradient vector. The exponential weighted average of the variable step-size, μ_n , can be expressed as follows:

$$\mu_n = \theta \mu_{n-1} + (1 - \theta) \exp (-\lambda |e_n|^\alpha) |e_n|^{\alpha-1}. \quad (12)$$

The replacement of the variable step-size from gradient vector can further reduce the computational cost and lead to a smoothed GMCC algorithm, namely SGMCC. The weight vector update of the SGMCC algorithm can be expressed as follows:

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta \mu_n \text{sign}(e_n) \mathbf{X}_n. \quad (13)$$

For convenience, the proposed SGMCC algorithm is specifically summarized in Algorithm 1.

Algorithm 1 SGMCC Algorithm.

Input: $\mathbf{X}_n, d_n, \alpha, \lambda, \theta, \eta$.

Initialize: $\mathbf{W}_1 = \mathbf{0}, e_0 = d_1, \mu_0 = 0$.

while $\{\mathbf{X}_n, d_n\}$ available

$$e_n = d_n - \mathbf{W}_n^T \mathbf{X}_n,$$

$$\mu_n = \theta \mu_{n-1} + (1 - \theta) \exp(-\lambda |e_n|^\alpha) |e_n|^{\alpha-1} \text{ according to Equation (12),}$$

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta \mu_n \text{sign}(e_n) \mathbf{X}_n \text{ according to Equation (13),}$$

end while

$$\mathbf{W}_* = \mathbf{W}_{n+1},$$

Output: estimated unknown system \mathbf{W}_* .

2.3. Computational Complexity

The complexity of the adaptive filtering algorithm is one of the important factors to measure its performance. The recursive weight updates of adaptive filtering algorithm generally include addition and multiplication operations. Since the complexity of multiplication operations is much higher than that of additive operations, in the iterative update process, the number of multiplication operations is usually used to calculate the computational complexity of the adaptive filtering algorithm. Next, we will make a comparison of the computational complexity between SGMCC and several other robust algorithms. The detail is summarized in Table 1.

Table 1. Comparison of computational complexity.

Algorithm	Computational Complexity
CMCC [4]	$6M + 10$
VSSA [16]	$5M + 4$
GMCC [21]	$2M + \alpha + 4$
SGMCC	$2M + \alpha + 6$

As one can see from Table 1, the computational complexity of the proposed SGMCC algorithm and the GMCC algorithm are much lower than that of the other algorithms. The SGMCC algorithm owns almost the same the computational complexity as the GMCC algorithm. Those two algorithms require about $2M + \alpha$ (usually $\alpha < M$) multiplications, which is just about half of that required by other robust algorithms. The CMCC algorithm requires the most multiplication operations since it has to calculate two MCC algorithms with different step-sizes and update the combination parameter to combine those two MCC algorithms. The VSSA algorithm still requires more multiplication operations than the proposed SGMCC algorithm as this algorithm involves the calculation of the weighted average of the gradient vector and its square norm.

3. Convergence Performance Analysis

3.1. Analysis of the Variable Step-Size

Since the proposed SGMCC algorithm can be viewed as a variable step-size sign algorithm, it is necessary to evaluate the variable step-size theoretically. For the proposed SGMCC algorithm,

the initial value of the variable step size is usually set as $\mu_0 = 0$, so the variable step size can be expressed as

$$\mu_n = (1 - \theta) \sum_{i=1}^n \theta^{n-i} \exp(-\lambda |e_i|^\alpha) |e_i|^{\alpha-1}. \quad (14)$$

Equation (14) shows the transient value of the variable step-size. If the transient variable step-size is positive and bounded, the adaptive algorithm can converge to steady state. It is obvious that μ_n is positive, namely

$$\mu_n = (1 - \theta) \sum_{i=1}^n \theta^{n-i} \exp(-\lambda |e_i|^\alpha) |e_i|^{\alpha-1} > 0. \quad (15)$$

In order to get the upper bound of the variable step size, we take the derivative of $\phi(e_i) = \exp(-\lambda |e_i|^\alpha) |e_i|^{\alpha-1}$ with respect to e_i and set it to 0, namely

$$\frac{\partial \phi(e_i)}{\partial e_i} = \exp(-\lambda |e_i|^\alpha) |e_i|^{\alpha-2} [(\alpha - 1) - \lambda \alpha |e_i|^\alpha] = 0, \quad (16)$$

then we can obtain the variable e_i which maximizes $\phi(e_i)$:

$$|e_i|^\alpha = \frac{\alpha - 1}{\lambda \alpha} \quad (17)$$

so the maximum of $\phi(e_i)$ is

$$\max \phi(e_i) = \left(\frac{\alpha - 1}{\lambda \alpha} \right)^{\frac{\alpha-1}{\alpha}} \exp\left(-\frac{\alpha - 1}{\alpha}\right). \quad (18)$$

Usually the shape parameter α is an integer (larger than 1) and the kernel parameter $\lambda < 1$, so we have

$$\max \phi(e_i) = \left(\frac{\alpha - 1}{\lambda \alpha} \right)^{\frac{\alpha-1}{\alpha}} \exp\left(-\frac{\alpha - 1}{\alpha}\right) < \lambda^{-1}. \quad (19)$$

Above all, we can get the range of the variable step-size

$$0 < \mu_n < \lambda^{-1}. \quad (20)$$

We can see that the variable step size of the proposed algorithm is positive and bounded, so the proposed SGMCC algorithm can converge to steady state when a suitable step-size parameter η is selected. When the adaptive algorithm converges to steady state, the variable step-size converges to a constant, which determines the steady-state accuracy of the adaptive filtering algorithm.

After taking the expectation for both sides of Equation (14), we obtain

$$\mathbb{E}[\mu_n] = (1 - \theta) \mathbb{E} \left[\sum_{i=1}^n \theta^{n-i} \exp(-\lambda |e_i|^\alpha) |e_i|^{\alpha-1} \right]. \quad (21)$$

When the step-size parameter η is small enough, the weight vector \mathbf{W}_n can approach very close to the optimal weight vector \mathbf{W}_* and the instantaneous prediction error e_n is dominated by the background noise v_n . Assume the proposed SGMCC algorithm converges to steady state when $n > N$, Equation (21) can be approximated as

$$\lim_{n \rightarrow N} \mathbb{E}[\mu_n] = (1 - \theta) \mathbb{E} \left[\lim_{n \rightarrow N} \sum_{i=1}^n \theta^{n-i} \exp(-\lambda |v_i|^\alpha) |v_i|^{\alpha-1} \right]. \quad (22)$$

The background noise v_n is usually assumed to be independent and identically distributed, so Equation (22) can be denoted as

$$\lim_{n \rightarrow N} E[\mu_n] = (1 - \theta) \lim_{n \rightarrow N} \sum_{i=1}^n \theta^{n-i} E \left[\exp(-\lambda |v_i|^\alpha) |v_i|^{\alpha-1} \right]. \quad (23)$$

When N is large enough, Equation (23) can be further simplified as

$$E[\mu_N] = E \left[\exp(-\lambda |v_N|^\alpha) |v_N|^{\alpha-1} \right]. \quad (24)$$

3.2. Mean-Square Stability

The energy conservation relation (ECR) is the most widely used method to evaluate the convergence performance analysis of an adaptive filtering algorithm [12,16,21,23–25]. In this paper, we also use the ECR analysis to successively derive the mean-square stability and the steady-state behavior of the proposed SGMCC algorithm.

For the simplicity of the ECR analysis, the recursive weight update of an adaptive filtering algorithm can be denoted as

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \eta f(e_n) \mathbf{X}_n. \quad (25)$$

For the proposed SGMCC algorithm, $f(e_n)$ can be expressed as

$$f(e_n) = \mu_n \text{sign}(e_n), \quad (26)$$

where the instantaneous prediction error e_n can be denoted as

$$e_n = d_n - \mathbf{W}_{n-1}^T \mathbf{X}_n = e_{a,n} + v_n, \quad (27)$$

where $e_{a,n} = \tilde{\mathbf{W}}_n^T \mathbf{X}_n$ is an a priori error. Here, $\tilde{\mathbf{W}}_n = \mathbf{W}_* - \mathbf{W}_n$ represents the weight-error vector.

The performance of the adaptive filtering algorithm is usually measured by the MSD value of the weight-error vector $\tilde{\mathbf{W}}_n$, namely

$$MSD = E \left[\|\tilde{\mathbf{W}}_n\|^2 \right]. \quad (28)$$

The weight-error vector at n th iteration can be derived as

$$\tilde{\mathbf{W}}_{n+1} = \tilde{\mathbf{W}}_n + \eta f(e_n) \mathbf{X}_n. \quad (29)$$

For both sides of Equation (29), we successively take the squared-norms and expectations, then we obtain the ECR expression as follows:

$$E \left[\|\tilde{\mathbf{W}}_{n+1}\|^2 \right] = E \left[\|\tilde{\mathbf{W}}_n\|^2 \right] - 2\eta E \left[f(e_n) e_{a,n} \right] + \eta^2 E \left[f^2(e_n) \|\mathbf{X}_n\|^2 \right]. \quad (30)$$

To facilitate the convergence analysis of the SGMCC algorithm based on the energy conservation relationship, some commonly-used assumptions [12,16,21,23–25] are listed:

Assumption 1 (A1). The background noise sequence $\{v_n\}$ is independent and identically distributed (i.i.d.). Also it is independent of the input vector sequence $\{\mathbf{X}_n\}$.

Assumption 2 (A2). The filter is long enough such that $e_{a,n}$ is zero-mean Gaussian and independent of the background noise v_n .

Remark 1. A1 is a valid assumption for most practical applications and is a basic assumption often used in signal processing. Moreover, it should be noted that, unlike most conventional signal processing methods that assume the background noise v_n is Gaussian distribution, A1 does not impose any restrictions on the statistical distribution of the background noise. Based on A1, it is easy to conclude that the a priori error $e_{a,n}$ and

background noise v_n are independent, which is assumed by A2. According to the Central Limit Theorem, it is reasonable that the a priori error $e_{a,n}$ satisfies the Gaussian distribution.

Mean-square stability analysis is carried out to determine the upper bound of step-size parameter η that makes sure the SGMCC algorithm is convergent. The adaptive algorithm is convergent implies that MSD is monotonously decreasing. That is to say that the following condition exists:

$$E \left[\|\tilde{\mathbf{W}}_{n+1}\|^2 \right] \leq E \left[\|\tilde{\mathbf{W}}_n\|^2 \right] \tag{31}$$

Based on the ECR relation, the following inequality holds:

$$2E [f(e_n)e_{a,n}] \geq \eta E \left[f^2(e_n)\|\mathbf{X}_n\|^2 \right]. \tag{32}$$

So, the step-size parameter η that makes adaptive filtering algorithm converges to steady state would be

$$\eta \leq 2 \inf_{n \geq 0} \frac{E [f(e_n)v_n]}{E \left[|f(e_n)|^2 \|\mathbf{X}_n\|_2^2 \right]}. \tag{33}$$

In our case, the scale parameter η that ensures the stability of the SGMCC algorithm should meet

$$\eta \leq 2 \inf_{n \geq 0} \frac{E [\mu_n \text{sign}(e_n)e_{a,n}]}{E \left[\mu_n^2 \|\mathbf{X}_n\|_2^2 \right]}. \tag{34}$$

Because μ_n is an exponential weighted average of $\exp(-\lambda|e_i|^\alpha) |e_i|^{\alpha-1}$, so the correlation between μ_n and $e_n, e_{a,n}$ is negligible when the smoothing factor β is close to 1. Therefore, the inequality Equation (34) can be further rewritten as follows:

$$\eta \leq 2 \inf_{n \geq 0} \frac{E [\mu_n] E [\text{sign}(e_n)e_{a,n}]}{E [\mu_n^2] E \left[\|\mathbf{X}_n\|_2^2 \right]}. \tag{35}$$

Next, considering the above mentioned condition $0 < \mu_n < \lambda^{-1}$, we obtain a smaller lower limit of the step-size parameter η :

$$\eta \leq \frac{2\lambda}{\text{Tr}(\mathbf{R}\mathbf{x})} \inf_{n \geq 0} E [\text{sign}(e_n)e_{a,n}], \tag{36}$$

where $\mathbf{R}\mathbf{x} = E [\mathbf{X}_n\mathbf{X}_n^H]$ is the covariance matrix of the input vector and $\text{Tr}(\cdot)$ denotes the trace operator.

In this paper, the impulsive background noise v_n is modeled as a random variable that follows a 2-component Gaussian mixture distribution $MG(\sigma_1, \sigma_2, p_r)$. Based on the assumption A2, the Price's theorem [26] and References [16,27], we have

$$E [\text{sign}(e_{a,n} + v_n)e_{a,n}] = E [e_{a,n}^2] \sqrt{\frac{2}{\pi}} \left\{ \frac{1 - p_r}{\sqrt{E [e_{a,n}^2] + \sigma_1^2}} + \frac{p_r}{\sqrt{E [e_{a,n}^2] + \sigma_2^2}} \right\}. \tag{37}$$

The Cramer-Rao [28] bound c is the minimum mean square error that the a priori error $e_{a,n}$ of an adaptive filtering algorithm can reach, so the step-size parameter η for the SGMCC algorithm would be

$$\eta \leq \frac{2\lambda c}{\text{Tr}(\mathbf{R}\mathbf{x})} \sqrt{\frac{2}{\pi}} \left\{ \frac{1 - p_r}{\sqrt{c + \sigma_1^2}} + \frac{p_r}{\sqrt{c + \sigma_2^2}} \right\}. \tag{38}$$

3.3. Steady-State Behavior

The steady state behavior is usually measured by $E[e_{a,n}^2]$ when the adaptive filtering algorithm converges to steady state. The steady-state value of $E[e_{a,n}^2]$ is generally known as EMSE. Assume the adaptive filtering algorithm converges to steady state when $n > N$, then EMSE can be denoted as

$$S = \lim_{n \rightarrow N} E[e_{a,n}^2]. \quad (39)$$

MSD would converge to a constant when the algorithm reaches steady-state, so we have

$$\lim_{n \rightarrow N} E[\|\tilde{\mathbf{W}}_{n+1}\|^2] = \lim_{n \rightarrow N} E[\|\tilde{\mathbf{W}}_n\|^2]. \quad (40)$$

Based on the ECR relation, we have

$$2 \lim_{n \rightarrow N} E[f(e_n)e_{a,n}] = \eta \lim_{n \rightarrow N} E[f^2(e_n)\|\mathbf{X}_n\|^2]. \quad (41)$$

For the proposed SGMCC algorithm, Equation (41) can be expressed as

$$2 \lim_{n \rightarrow N} E[\mu_n \text{sign}(e_n) e_{a,n}] = \eta \lim_{n \rightarrow N} E[\mu_n^2 \|\mathbf{X}_n\|^2]. \quad (42)$$

With the same reason mentioned above, the correlation between μ_n and e_n , $e_{a,n}$ is negligible. Therefore, the following expression holds

$$2 \lim_{n \rightarrow N} E[\mu_n] E[\text{sign}(e_n) e_{a,n}] = \eta \text{Tr}(\mathbf{R}\mathbf{x}) \lim_{n \rightarrow N} E[\mu_n^2]. \quad (43)$$

As is derived in Equation (24), $E[\mu_n]$ would converge to a constant when the η is small enough. Therefore, Equation (43) becomes

$$2 \lim_{n \rightarrow N} E[\text{sign}(e_n) e_{a,n}] = \eta \text{Tr}(\mathbf{R}\mathbf{x}) E[\mu_N]. \quad (44)$$

The impulsive background noise is assumed to be Gaussian mixture noise that follows $MG(\sigma_1, \sigma_2, p_r)$. Based on the assumption A2, the Price's theorem [26] and References [16,27], we have

$$\lim_{n \rightarrow N} E[\text{sign}(e_{a,n} + v_n)e_{a,n}] = S \sqrt{\frac{2}{\pi}} \left\{ \frac{1 - p_r}{\sqrt{S + \sigma_1^2}} + \frac{p_r}{\sqrt{S + \sigma_2^2}} \right\}. \quad (45)$$

After substituting Equations (44) and (45) into Equation (43), we obtain

$$S = \sqrt{\frac{\pi}{8}} \eta \text{Tr}(\mathbf{R}\mathbf{x}) E[\mu_N] \left\{ \frac{1 - p_r}{\sqrt{S + \sigma_1^2}} + \frac{p_r}{\sqrt{S + \sigma_2^2}} \right\}^{-1}. \quad (46)$$

When the step-size parameter η is small enough, EMSE or S would converge to a smaller number, which is negligible comparing with the power of background noises. In this case, Equation (46) can be approximated by

$$S \approx \eta \text{Tr}(\mathbf{R}\mathbf{x}) E[\mu_N] \sqrt{\frac{\pi}{8}} \frac{\sigma_2 \sigma_1}{\sigma_2 (1 - p_r) + \sigma_1 p_r}. \quad (47)$$

Note that the theoretical EMSE value computed by (Equation (47)) is an approximate value that the SGMCC algorithm can achieve at steady state. The accuracy of the approximation is largely affected by the step-size parameter η . A valid approximation of EMSE can be secured with a small-enough η .

Usually, $\sigma_1 \ll \sigma_2$ and $p_r \ll 1 - p_r$, Equation (47) can be simplified as

$$S \approx \eta \text{Tr}(\mathbf{R}\mathbf{x}) E[\mu_N] \sqrt{\frac{\pi}{8}} \frac{\sigma_1}{1 - p_r}. \tag{48}$$

4. Simulation Results

To validate the theoretical analysis and evaluate the performance of the proposed SGMCC algorithm, we conduct simulations in a channel estimation [29–32] with Gaussian mixture background noise. An unknown time-varying channel with 20 multipath subchannels is randomly generated and then normalized by $\mathbf{W}_{opt}^H \mathbf{W}_{opt} = 1$. The input signals are generated by a zero-mean Gaussian distribution with unit variance. The input signals are transmitted via the above mentioned unknown channel and contaminated by impulsive background noise generated by a 2-component Gaussian mixture distribution $MG(\sigma_1, \sigma_2, p_r)$. The simulation results are averaged over 100 independent Monte-Carlo experiments.

Firstly, we compare the MSD convergence curves of the proposed SGMCC algorithm with four parameters and investigate the effect of each parameter on the convergence performance of the SGMCC algorithm. Note that Figure 1 is obtained with $\sigma_1 = 0.2$, $\sigma_2 = 100$ and $p_r = 0.1$. At each simulation, we set different values for a specific parameter with the other three parameters fixed and observe the effect of the parameter on the corresponding MSD convergence curve.

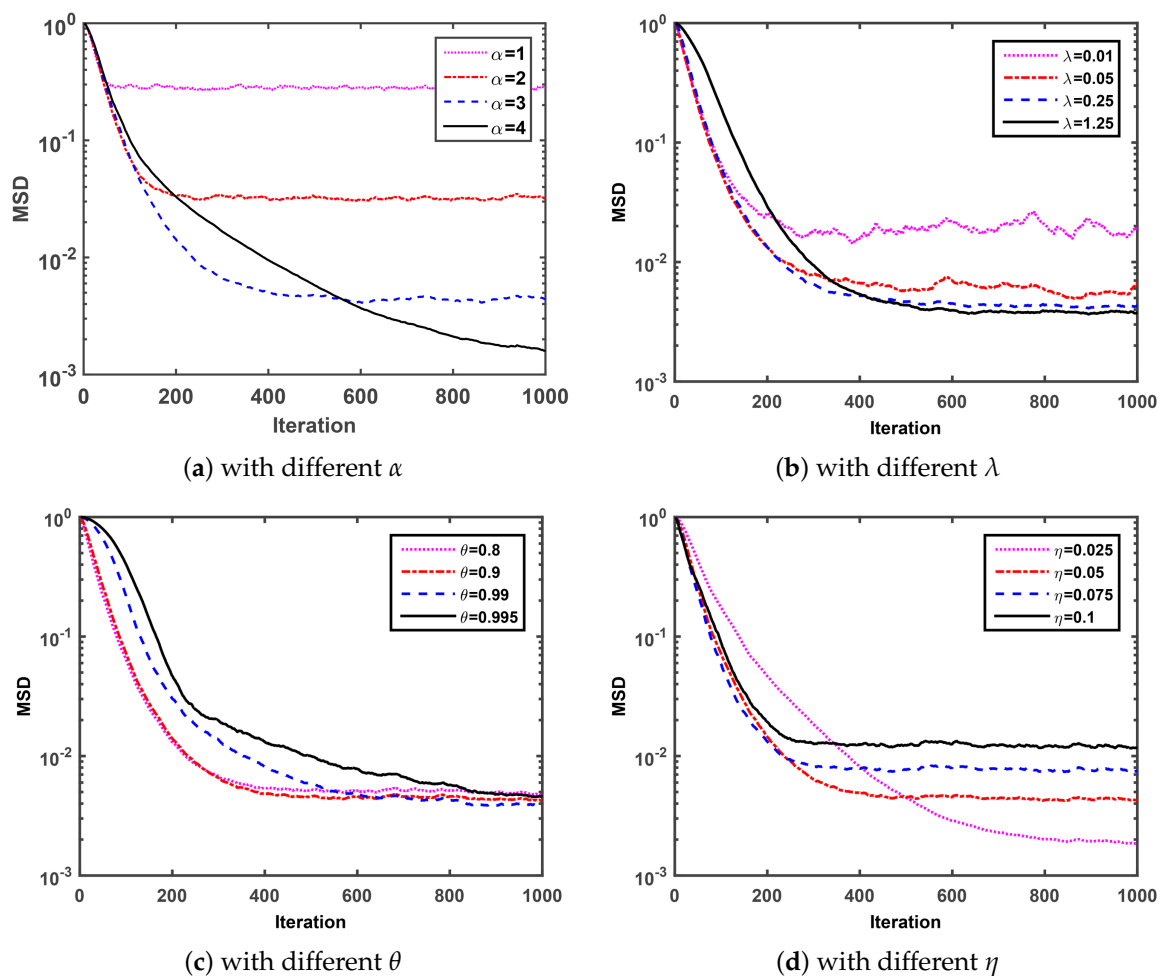


Figure 1. MSD convergence curves of SGMCC algorithm.

Under the condition of fixed parameter $\lambda = 0.25, \theta = 0.9$ and $\eta = 0.05$, Figure 1a shows that the SGMCC algorithm with larger shape parameter α converges slower to a lower steady-state MSD. As one can see in Figure 1b, under the condition of fixed parameter $\alpha = 3, \theta = 0.9$ and $\eta = 0.05$, the MSD convergence curve of SGMCC algorithm with smaller kernel parameter λ is not smooth and fluctuates around a larger range. However, a larger kernel parameter λ may lead to a slower convergence speed. As observed in Figure 1c, under the condition of fixed parameter $\alpha = 3, \lambda = 0.25$ and $\eta = 0.05$, the SGMCC algorithm with different smooth factor θ values converges to the almost same steady-state MSD after 1000 iterations. In addition, the SGMCC algorithm with smaller θ converges faster but the improvement of the convergence speed is not obvious when θ smaller than 0.9. At last in Figure 1d, under the condition of fixed parameter $\alpha = 3, \lambda = 0.25$ and $\theta = 0.9$, the SGMCC algorithm with smaller step-size parameter η can achieve lower steady-state MSD at the cost of a slower convergence speed and vice versa.

Secondly, we compare the convergence MSD performance of our proposed algorithm with several recently published algorithms to demonstrate its convergence performance. The parameters of the GMCC algorithm is set the same as that of the SGMCC algorithm. The parameters of other algorithms are carefully adopted according to the corresponding reference. The compared algorithms and their corresponding parameters were summarized in Table 2.

Table 2. Parameters setting.

Algorithm	Parameters
CMCC [4]	$\lambda = 0.5, \mu_a = 4.5, \beta = 0.8, \gamma = 1.5, \sigma = 2$ $\mu_1 = 0.05, \mu_2 = 0.01, \varepsilon_1 = \varepsilon_2 = 0.00001$
VSSA [16]	$\alpha = 0.95, \beta = 0.97, \gamma = 0.0005$
GMCC [21]	$\alpha = 3, \lambda = 0.25, \eta = 0.05$
SGMCC	$\alpha = 3, \lambda = 0.25, \eta = 0.05, \theta = 0.9$

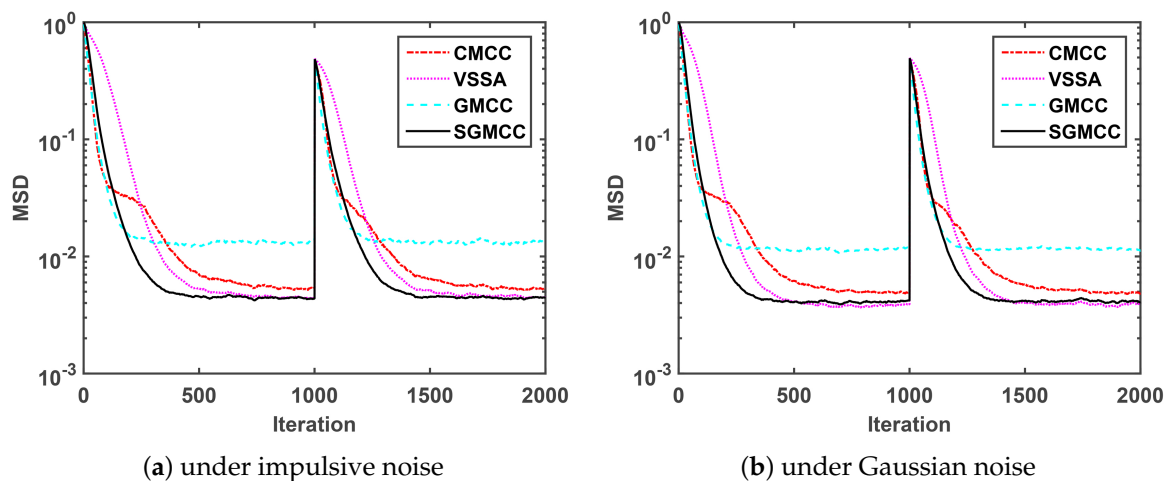


Figure 2. Comparison of the mean square deviation (MSD) curves versus iteration.

Figure 2a shows a comparison of the MSD convergence curves of different algorithms under impulsive noise generated by $MG(0.2, 100, 0.1)$. As observed in Figure 2a, the proposed SGMCC algorithm achieves a much lower steady-state MSD than the GMCC algorithm with the same parameters. The SGMCC algorithm and the VSSA algorithm converge to the same and the lowest steady-state MSD. On the other hand, the proposed SGMCC algorithm converges faster than the VSSA algorithm and the CMCC algorithm. Different from Figure 2a simulated under impulsive noise, Figure 2b is conducted under Gaussian noise. We can see that Figure 2b shows the similar results with Figure 2a. So we can conclude that, compared with the other three algorithms, the proposed SGMCC

algorithm achieves the lowest steady-state MSD with a relatively fast convergence speed. Also, the proposed SGMCC algorithm is robust against both impulsive noise and Gaussian noise.

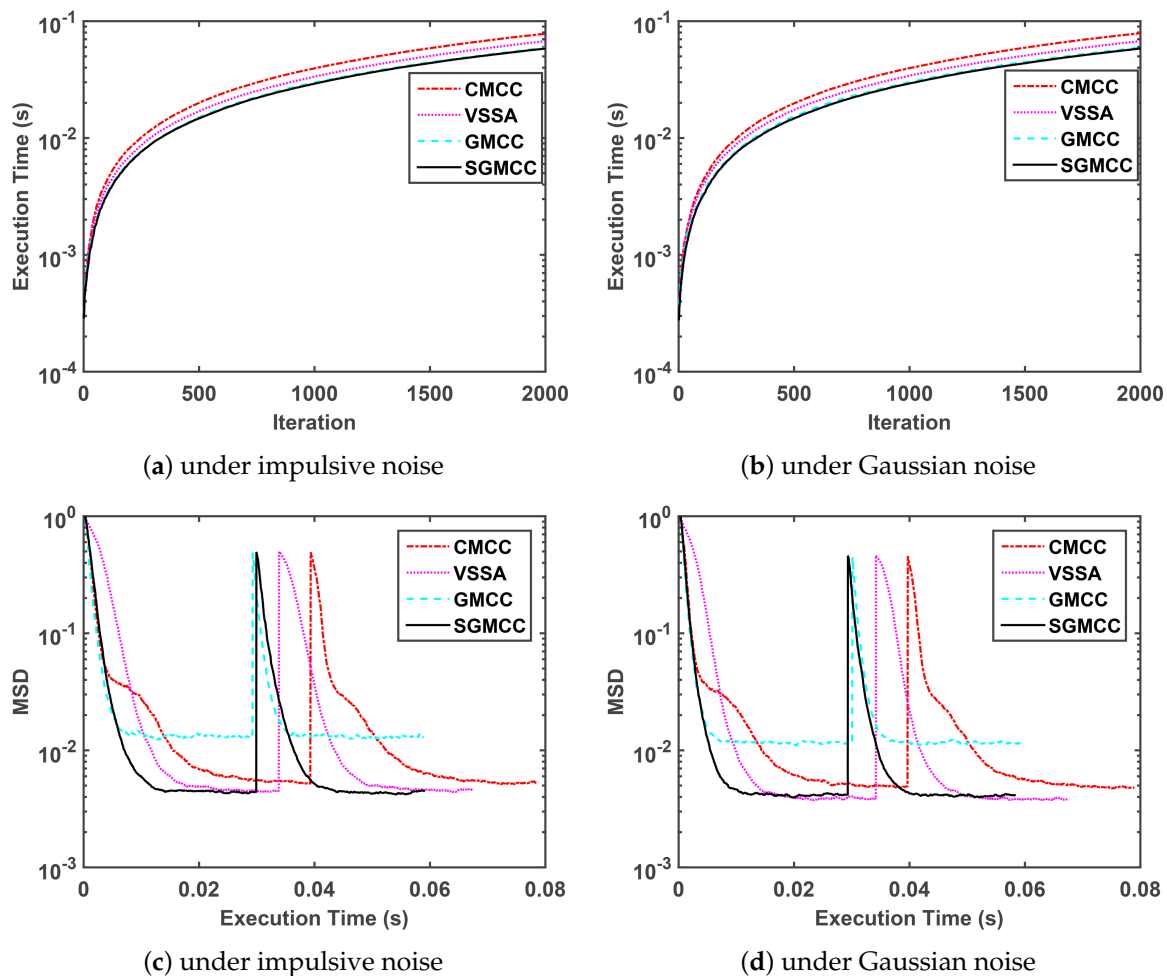


Figure 3. Comparison of the execution time.

Thirdly, we compare the execution time of our proposed algorithm with several recently published algorithms in the next Figure 3. The parameter setting of four algorithms in Figure 3 is same with that in Figure 2, which is shown in Table 2. Note that the measured execution time of four algorithms is related to the running platform and specific code-programming efficiency.

Figure 3a,b shows a comparison of the execution time curves of different algorithms versus iteration under impulsive noise and Gaussian noise, respectively, and these two subfigures show the same performance trend. Figure 3c,d shows a comparison of the steady-state MSD curves of different algorithms versus execution time under impulsive noise and Gaussian noise respectively and these two subfigures show the same performance trend. Concretely speaking, Figure 3a,b demonstrate that the proposed SGMCC algorithm owns the shortest run time and is close to the GMCC algorithm. Figure 3c,d demonstrate that the SGMCC algorithm owns the shortest run time by observing the length of the curve tail. Although the run time of the proposed SGMCC algorithm is also is close to that of the GMCC algorithm, the SGMCC algorithm achieves a much lower steady-state MSD. Moreover, the superiority of the SGMCC algorithm in run time can be ensured by computational complexity shown in Table 1.

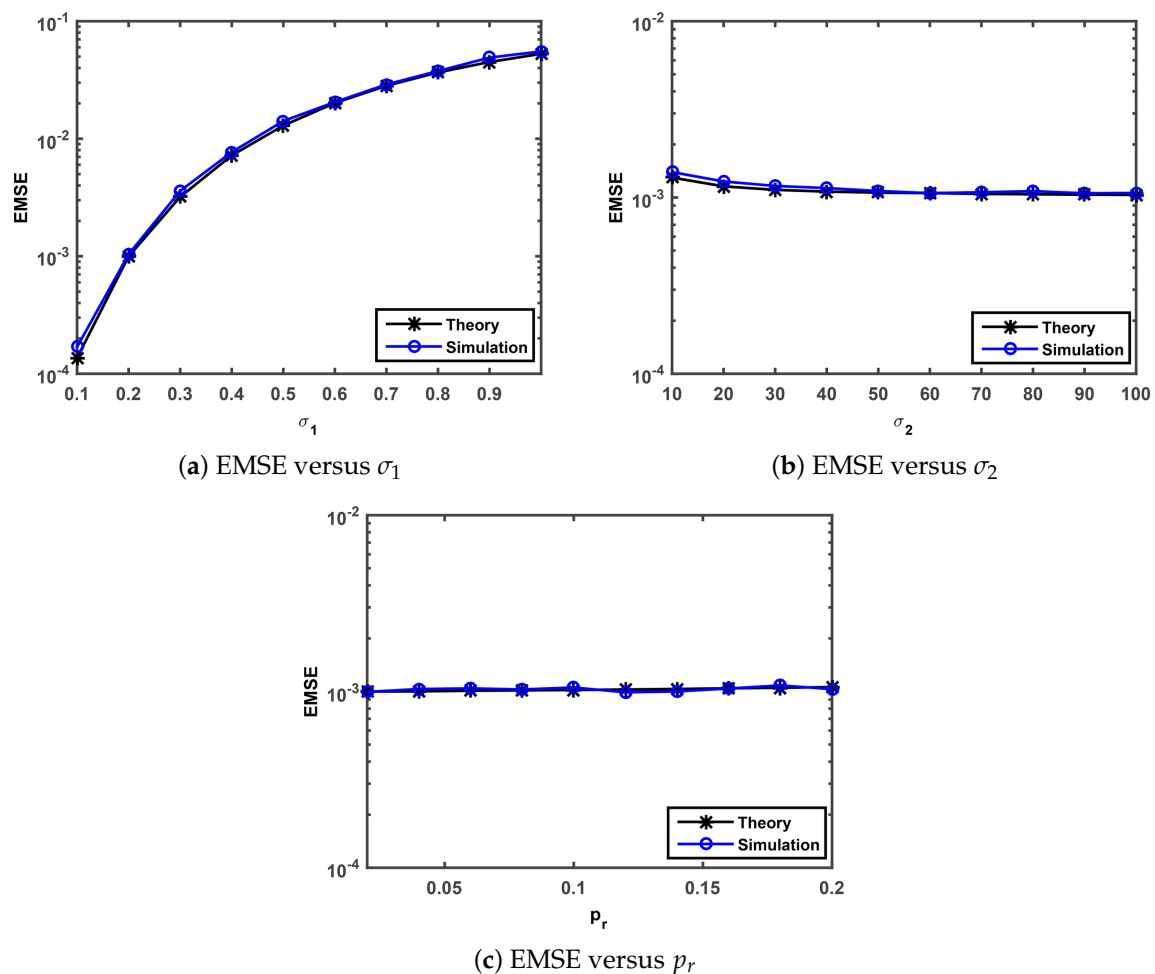


Figure 4. Comparison of the theoretical result with the simulation result.

Finally, we perform simulations to confirm the steady-state behavior analysis presented in Section 3.3. The theoretical steady-state EMSEs are evaluated by Equation (47). The simulated steady state EMSEs were calculated as a average over the last 500 iterations of EMSE that is averaged over 50 independent Monte-Carlo simulation with 5000 iterations. We investigate the theoretical and simulated EMSEs under $MG(\sigma_1, \sigma_2, p_r)$ versus Gaussian noise standard deviation σ_1 , impulsive noise standard deviation σ_2 , impulsive noise occurrence probability p_r . Note that the parameters of the SGMCC algorithm in Figure 4 are set as $\alpha = 3$, $\lambda = 0.25$, $\theta = 0.9$ and $\eta = 0.01$.

The comparisons of the theoretical and simulated value on EMSE versus σ_1 , σ_2 and p_r are shown in Figure 4a–c, respectively. As observed in Figure 4a, the simulated EMSEs match well the theoretical EMSEs computed by Equation (47) and the simulation value grows with the increase of the parameter σ_1 . As one can see from Figure 4b,c, the simulated EMSEs also match well the theoretical EMSEs. In Figure 4b, the EMSE values show a very slow downward trend and no significant changes with the increase of the parameter σ_2 from 10 to 60 and from 60 to 100, respectively. In Figure 4c, the EMSE values keep almost the same value with the range of the parameter p_r value from 0.02 to 0.2. This implies that the steady-state behavior analysis presented in Section 3.3 is valid and the proposed SGMCC algorithm is robust against impulsive noise.

5. Conclusions

In this paper, we propose a smoothed GMCC algorithm called SGMCC to improve the performance of steady-state MSD in the presence of impulsive noise. To avoid the high steady-state MSD caused by the stochastic gradient, instead of taking the exponential weighted average of gradient

vector to approximate the expectation of the gradient vector, we take the exponential weighted average of the variable step-size so that the SGMCC algorithm can be viewed as a sign GMCC algorithm with smoothed variable step-size. Convergence performance analyses are derived to demonstrate the robustness of the proposed algorithm and it is verified by the previous simulation part. Simulation results also demonstrate that the proposed SGMCC algorithm is robust against impulsive noise and converges fast with lower computational complexity.

Author Contributions: H.Q. and Y.S. modeled and analyzed the problem. J.Z. proposed and implemented the result. Y.S. wrote the paper.

Funding: This research was supported by the National Natural Science Foundation of China (No.61531013) and the National Science and Technology Major Project of China (No.2018ZX03001016).

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

References

- Principe, J.C. *Information Theoretic Learning*; Springer: New York, NY, USA, 2010.
- Santamaria, I.; Pokharel, P.P.; Principe, J.C. Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Trans. Signal Process.* **2006**, *54*, 2187–2197. [[CrossRef](#)]
- 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](#)]
- Shi, L.; Lin, Y. Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference. *IEEE Signal Process. Lett.* **2014**, *21*, 1385–1388. [[CrossRef](#)]
- Ma, W.; Qu, H.; Zhao, J.; Chen, B. Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-gaussian environments. *J. Frankl. Inst.* **2015**, *352*, 2708–2727. [[CrossRef](#)]
- Wang, W.; Zhao, J.; Qu, H.; Chen, B.; Principe, J.C. A switch kernel width method of correntropy for channel estimation. In Proceedings of the International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015; pp. 1–7.
- Wang, W.; Zhao, J.; Qu, H.; Chen, B.; Principe, J.C. An adaptive kernel width update method of correntropy for channel estimation. In Proceedings of the IEEE International Conference on Digital Signal Processing, Singapore, 21–24 July 2015; pp. 916–920.
- Ogunfunmi, T.; Paul, T. The quaternion maximum correntropy algorithm. *IEEE Trans. Circuits Syst. II Express Briefs* **2015**, *62*, 598–602. [[CrossRef](#)]
- Mandanas, F.; Kotropoulos, C. Robust multidimensional scaling using a maximum correntropy criterion. *IEEE Trans. Signal Process.* **2016**, *65*, 919–932. [[CrossRef](#)]
- Kulikova, M. Square-root algorithms for maximum correntropy estimation of linear discrete-time systems in presence of non-gaussian noise. *Syst. Control Lett.* **2017**, *108*, 8–15. [[CrossRef](#)]
- Singh, A.; Principe, J.C. Using correntropy as a cost function in linear adaptive filters. In Proceedings of the International Joint Conference on Neural Networks, Atlanta, Georgia, 14–19 June 2009; pp. 2950–2955.
- 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.
- Kwong, R.; Johnston, E. A variable step size lms algorithm. *IEEE Trans. Signal Process.* **1992**, *40*, 1633–1642. [[CrossRef](#)]
- Ang, W.; Farhang-Boroujeny, B. A new class of gradient adaptive stepsize lms algorithms. *IEEE Trans. Signal Process.* **2001**, *49*, 805–810.
- Shin, H.; Sayed, A.; Song, W. Variable step-size nlms and affine projection algorithms. *IEEE Signal Process. Lett.* **2004**, *11*, 132–135. [[CrossRef](#)]
- Li, Y.; Lee, T.; Wu, B. A variable step-size sign algorithm for channel estimation. *Signal Process.* **2014**, *102*, 304–312. [[CrossRef](#)]
- Wang, R.; Chen, B.; Zheng, N.; Principe, J.C. A variable step-size adaptive algorithm under maximum correntropy criterion. In Proceedings of the International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015; pp. 1–5.

18. Roenko, A.; Lukin, V.; Djurovi, I.; Simeunovi, M. Estimation of parameters for generalized gaussian distribution. In Proceedings of the 6th International Symposium on Communications, Control and Signal Processing, Athens, Greece, 21–23 May 2014; pp. 376–379.
19. Ahmed, Q.; Park, K.; Alouini, M. Ultrawide bandwidth receiver based on a multivariate generalized gaussian distribution. *IEEE Trans. Wirel. Commun.* **2014**, *14*, 1800–1810. [[CrossRef](#)]
20. Ghofrani, P.; Wang, T.; Schmeink, A. Robust Kalman Filter Algorithm Based on Generalized Correntropy for Ultra-Wideband Ranging in Industrial Environment. *IEEE Access* **2019**, *7*, 27490–27500.
21. Chen, B.; Xing, L.; Zhao, H.; Zheng, N. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [[CrossRef](#)]
22. Wu, Z.; Peng, S.; Chen, B.; Zhao, H. Robust hammerstein adaptive filtering under maximum correntropy criterion. *Entropy* **2014**, *52*, 131–139. [[CrossRef](#)]
23. Al-Naffouri, T.Y.; Sayed, A.H. Adaptive filters with error nonlinearities: mean-square analysis and optimum design. *EURASIP J. Appl. Signal Process.* **2001**, *2001*, 192–205. [[CrossRef](#)]
24. Yu, Y.; Zhao, H.; Chen, B. Steady-state mean-square-deviation analysis of the sign subband adaptive filter algorithm. *Signal Process.* **2016**, *120*, 36–42. [[CrossRef](#)]
25. Li, K.; Principe, J.C. Transfer learning in adaptive filters: The nearest instance centroid-estimation kernel least-mean-square algorithm. *IEEE Trans. Signal Process.* **2017**, *65*, 6520–6535. [[CrossRef](#)]
26. Price, R. A useful theorem for nonlinear devices having gaussian inputs. *IEEE Trans. Inf. Theory* **1958**, *4*, 69–72. [[CrossRef](#)]
27. Bang, S.; Ann, S.; Song, I. Performance analysis of the dual sign algorithm for additive contaminated-gaussian noise. *IEEE Signal Process. Lett.* **1994**, *1*, 196–198. [[CrossRef](#)]
28. Gini, F.; Reggiannini, R.; Mengali, U. The modified cramer-rao bound in vector parameter estimation. *IEEE Trans. Commun.* **1998**, *46*, 52–60. [[CrossRef](#)]
29. Pelekanakis, K.; Chitre, M. Adaptive sparse channel estimation under symmetric alpha-stable noise. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 3183–3195. [[CrossRef](#)]
30. Jellali, Z.; Atallah, L. Fast fading channel estimation by kalman filtering and cir support tracking. *IEEE Trans. Broadcast.* **2017**, *63*, 635–643. [[CrossRef](#)]
31. Kari, D.; Marivani, I.; Khan, F.; Sayin, M.; Kozat, S. Robust adaptive algorithms for underwater acoustic channel estimation and their performance analysis. *Digit. Singal Process.* **2017**, *68*, 57–68. [[CrossRef](#)]
32. Ghofrani, P.; Wang, T.; Schmeink, A. A fast converging channel estimation algorithm for wireless sensor networks. *IEEE Trans. Signal Process.* **2018**, *66*, 3169–3184. [[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/>).