

Article

Newtonian-Type Adaptive Filtering Based on the Maximum Correntropy Criterion

Pengcheng Yue ^{1,*},[†] , Hua Qu ^{1,2,†}, Jihong Zhao ^{1,3} and Meng Wang ²

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

² School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China; avictor.wally@stu.xjtu.edu.cn

³ School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710061, China

* Correspondence: heaven.1452@stu.xjtu.edu.cn

† These authors contributed equally to this work.

Received: 19 July 2020; Accepted: 21 August 2020; Published: 22 August 2020



Abstract: This paper provides a novel Newtonian-type optimization method for robust adaptive filtering inspired by information theory learning. With the traditional minimum mean square error (MMSE) criterion replaced by criteria like the maximum correntropy criterion (MCC) or generalized maximum correntropy criterion (GMCC), adaptive filters assign less emphasis on the outlier data, thus become more robust against impulsive noises. The optimization methods adopted in current MCC-based LMS-type and RLS-type adaptive filters are gradient descent method and fixed point iteration, respectively. However, in this paper, a Newtonian-type method is introduced as a novel method for enhancing the existing body of knowledge of MCC-based adaptive filtering and providing a fast convergence rate. Theoretical analysis of the steady-state performance of the algorithm is carried out and verified by simulations. The experimental results show that, compared to the conventional MCC adaptive filter, the MCC-based Newtonian-type method converges faster and still maintains a good steady-state performance under impulsive noise. The practicability of the algorithm is also verified in the experiment of acoustic echo cancellation.

Keywords: Newtonian method; robust adaptive filter; maximum correntropy criterion; steady-state performance analysis; acoustic echo cancellation

1. Introduction

Adaptive filtering is widely used in many areas including system identification, channel equalization, interference cancelling, acoustic echo cancellation (AEC), etc. [1–5]. Traditional adaptive filtering methods based on minimum mean square error (MMSE) criterion perform well in the presence of Gaussian noise, and the optimization methods adopted are mostly least mean square (LMS)-type or recursive least square (RLS)-type [6]. LMS-type adaptive filtering uses gradient descent characterized with a low convergence speed and a very low arithmetic complexity, while RLS-type adaptive filtering, free from the selection problem of the optimal step size, converges much faster at the cost of higher complexity, and is afflicted with stability problems caused by error propagation and unregulated matrix inversion [7]. Kalman filter is also an important optimization method in state estimation [6].

MMSE criterion-based adaptive filtering suffers from impulsive noise for its sensitivity to large outliers. To deal with this problem, robust adaptive filtering has been researched extensively. A popular robust solution is to change the MMSE criterion to other criteria that suppress impulsive noise, such as

Huber loss [8], least p-norm criterion [9], etc. In recent years, information theory learning (ITL) was found suitable to deal with non-Gaussian noises [10–14]. Inspired by ITL, maximum correntropy criterion (MCC) or generalized maximum correntropy criterion (GMCC)-based adaptive filtering was studied [15–20]. Most of the aforementioned robust algorithms were LMS-like [21,22] or RLS like [23–25], which is to say that the optimization methods used were limited to gradient descent and fixed point iteration [26]. MCC-based Kalman filter as an optimization method of state estimation was studied in [27]. Ref. [28] implied that a Newtonian algorithm could be utilized in the MCC state estimation, as correntropy is a differentiable function. However, different methods such as Newton's method and all its derivative algorithms that converge faster than gradient descent are seldom considered in MCC-based adaptive filtering. As an inspiring attempt to enrich the optimization methods, ref. [29,30] proposed a correntropy-based Levenberg–Marquardt algorithm that converges faster than maximum correntropy-based gradient descent algorithm and performs well dealing with heavy tailed non-Gaussian noise. This work revealed the potential adoption of more optimization methods except gradient descent and fixed point iteration. Especially, Newtonian-type optimization methods of MCC-based robust adaptive filtering are far from complete and still need to be improved.

Adaptive filtering based on Newtonian or quasi-Newtonian methods was proved to be serviceable for its fast convergence rate [31–36]. Adaptive filtering based on Newtonian methods, known as LMS-Newton, models the input sequence as an autoregressive (AR) process, and usually focuses on the acceleration of the estimate of the input autocorrelation matrix [32,33]. Adaptive filtering based on quasi-Newtonian methods, known as quasi-Newtonian adaptive filtering, usually updates the approximation of the Newtonian direction (typically the inverse Hessian matrix) by formulas similar to that of BFGS [34,37]. Adaptive filtering based on the Gauss–Newton method or Levenberg–Marquardt (LM) method also put forward approximations of the Hessian matrix that are easy to compute using the Jacobian matrix [35]. Moreover, there are many methods adopted to enhance the robustness of Newtonian-type adaptive filtering [31,34,36]. Reference [36] proposed a robust algorithm and revealed that the weighting function related to the cost function in the algorithm is the key that ensures the robustness. Inspired by that, we adopt LMS-Newton to solve the optimization of the cost function based on MCC, and we call this the MCC-Newton adaptive filtering method, which enhances the existing body of knowledge in MCC-based adaptive filtering. The proposal is in the category of the most commonly used linear transversal filter.

The main contributions of this paper are as follows: (1) the Newtonian-type optimization method is introduced in MCC-based adaptive filter and the recursive updating equation of the impulse response is derived. (2) The steady-state performance analysis is discussed theoretically and compared with that in experiments. According to the theoretical analysis, the guideline of parameter selection of the algorithm is provided. (3) The algorithm is applied in system identification and acoustic echo cancellation in experiments to verify the practicability.

The paper is divided into five parts. Section 2 presents the conventional Newtonian-type adaptive filter based on the MMSE criterion and introduces the MCC. Section 3 proposes the Newtonian-type adaptive filter based on MCC and gives the recursive solution of the impulsive response. The complexity of the algorithm is also compared with that of the other algorithms. Section 4 analyzes the steady-state performance of the algorithm theoretically. The experiments verifying the steady-state performance discussion is displayed in Section 5, and there are also experiments showing that the proposed algorithm is robust in the presence of impulsive noise and converges faster than the gradient descent-based adaptive filter algorithms. Besides, the experiment of acoustic echo cancellation verifies the practicability of the algorithm. Section 6 gives the conclusion of the paper.

2. Preliminaries

2.1. Conventional Newtonian-Type Adaptive Filtering

The adaptive filter update equation of the conventional LMS-Newton algorithm is implemented as in [33,36]:

$$e(n) = d(n) - y(n) \quad (1)$$

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu e(n) \hat{\mathbf{R}}^{-1}(n) \mathbf{X}(n), \quad (2)$$

where $e(n)$ is the estimation error at time point n , $d(n)$ is the observed system output, and the linear transversal filter estimated output $y(n) = \mathbf{X}^T(n) \mathbf{W}(n)$ is the filter parameter $\mathbf{W}(n)$ multiplying the input \mathbf{X} . μ is the step size, and $\hat{\mathbf{R}}$ is the estimated autocorrelation matrix of $\mathbf{X}(n)$ which is assumed to be known in the ideal LMS-Newton algorithm. Note that when $\hat{\mathbf{R}}$ equals to the identity matrix I , (2) becomes the equation of the conventional LMS algorithm. The ideal algorithm is easy to analyze theoretically, but it is considered impractical for the computational complexity.

In practical articles considering the application of the LMS-Newton method in acoustic echo cancellation, the input $\mathbf{X}(n)$ is modeled as an autoregressive (AR) process in which the order is much smaller than the length of the filter [33,38,39]. According to the characteristics of AR modeling, there are many efficient ways to simplify the computation and updating of $\hat{\mathbf{R}}^{-1}(n) \mathbf{X}(n)$. Some of the articles estimate $\hat{\mathbf{R}}^{-1}(n)$ first and then perform the matrix multiplication [38,39], while other articles directly compute the vector $\hat{\mathbf{R}}^{-1}(n) \mathbf{X}(n)$ without the estimation of $\mathbf{R}(n)$ [33]. Owing to the efficient updating of $\hat{\mathbf{R}}^{-1}(n) \mathbf{X}(n)$, the modified practical LMS-Newton algorithms could keep very small computational complexity, which is approximately equal to that of the conventional LMS algorithm, adding a negligible updating operation of $\hat{\mathbf{R}}^{-1}(n) \mathbf{X}(n)$. Practical Newtonian-type adaptive filtering has the potential to converge as fast as RLS-type adaptive filtering while maintaining a low computational complexity.

2.2. Maximum Correntropy Criterion

Conventional Newtonian-type adaptive filtering derives the equation of the adaptive filter parameter according to the MMSE criterion, which performs well under Gaussian noise but suffers from non-Gaussian noise. In this paper, MCC is introduced to enhance the robustness of the Newtonian-type adaptive filtering. Correntropy in ITL is used to measure the similarity of two random variables, and is defined as [10]:

$$V(X, Y) = E[\kappa(X, Y)] = \int \kappa(x, y) dF_{XY}(x, y), \quad (3)$$

where X, Y denote two random variables; $E[\cdot]$ is the expectation operator; $F_{XY}(x, y)$ represents the joint distribution function of (X, Y) ; and $\kappa(\cdot)$ stands for a Mercer kernel which is in general the Gaussian kernel defined as:

$$\kappa_{\beta}(x, y) = G_{\beta}(x - y) = \frac{1}{\sqrt{2\pi\beta}} \exp\left(-\frac{|x - y|^2}{2\beta^2}\right), \quad (4)$$

where β is the Gaussian kernel width, and $1/\sqrt{2\pi\beta}$ is the normalization parameter.

In practical adaptive filtering, the available data $\{x_n, y_n\}_{n=1}^N$ of X, Y are discrete and the joint distribution $F_{XY}(x, y)$ can be estimated by the Parzen kernel estimator as:

$$\hat{V}_N^{\beta}(X, Y) = \frac{1}{N} \sum_{n=1}^N \kappa_{\beta}(x_n - y_n). \quad (5)$$

2.3. Comparison of Different Criteria

This section presents the characteristic of robustness of MCC to outliers and gives a formal motivation for the use of MCC. The loss functions of different criteria are the metrics of the estimation

error e . We compare the correntropy induced loss function defined in (4) with the MSE loss function and its general version, Lp norm loss function.

The MSE loss, in other words, L2 loss is defined as:

$$L_{MSE} = |e|^2. \tag{6}$$

The Lp norm loss can be defined as

$$L_{Lp} = |e|^p, \tag{7}$$

where p usually satisfies $1 < p < 2$. When $p = 2$, Lp norm loss becomes L2 loss.

Note that the correntropy induced loss function (C-loss) defined in (4) is a concave function. So in MCC, we try to find the maximum of the cost function to reduce the estimation error. However, we usually wish the loss function to be convex, so that the function value will get larger as the absolute value of the error increases. So we used a modified version of C-loss [40] to compare with the other loss functions. The modified C-loss is convex and can be defined as:

$$L_C = \frac{1}{\sqrt{2\pi}\beta} (1 - \exp(-\frac{|e|^2}{2\beta^2})). \tag{8}$$

Figure 1 shows the loss function of different criteria. For each convex loss function, the function value increases with the growth of the absolute value of the estimation error. When an outlier appears, the value of L2-loss and Lp-loss becomes very large, but the value of C-loss remains relatively stable. Hence one can conclude that compared with the conventional loss functions, C-loss is robust against outliers.

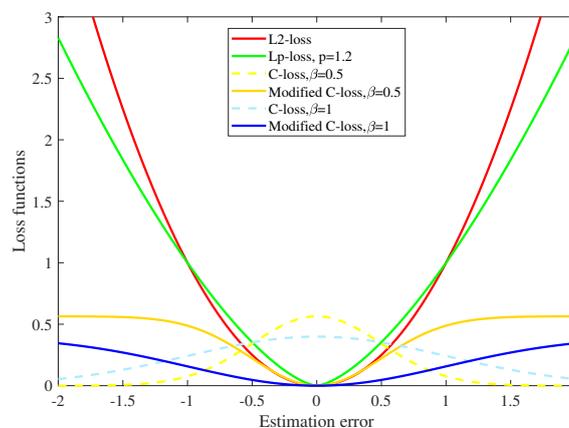


Figure 1. Comparison of different loss functions.

3. A Newtonian-Type Adaptive Filtering Based on MCC

In our robust Newtonian-type adaptive filtering, MCC is used to construct the cost function $J = E[p(e(n))]$, in which $p(e(n))$ is the loss function of the error $e(n)$ at time point n , and can be written as $G_\beta(e(n))$ in MCC instead of $e^2(n)$, the L2 loss in MMSE criterion. To calculate the optimal estimation of the adaptive filter parameter $\mathbf{W}(n)$, the gradient of the cost function with respect to $\mathbf{W}(n)$ is set to zero and we can derive that

$$E [G_\beta(d - y)(d - y)\mathbf{X}] = 0 \tag{9}$$

$$E [G_\beta(e) (d - \mathbf{X}^T \mathbf{W}) \mathbf{X}] = 0 \tag{10}$$

and

$$E [G_\beta(e)\mathbf{X}\mathbf{X}^T] \cdot \mathbf{W} = E [G_\beta(e)d \cdot \mathbf{X}]. \tag{11}$$

In adaptive filtering, the data are discrete and expectation operators can be presented using their sample estimator:

$$\mathbf{R}_{MCC} \cdot \mathbf{W} = \mathbf{P}_{MCC}, \tag{12}$$

where $\mathbf{R}_{MCC} = \sum_{n=1}^N q(e(n)) \mathbf{X}(n)\mathbf{X}(n)^T$ and $\mathbf{P}_{MCC} = \sum_{n=1}^N q(e(n)) d(n) \cdot \mathbf{X}(n)$. $q(e(n))$ is the weighting function of RLS-type MCC-based adaptive filtering that values how much the n -th sample data influences the filtering. In Newtonian-type and RLS-type adaptive filtering, the weighting function can be calculated by [19,36]:

$$q(e(i)) = |p(e(i))'/e(i)|, \tag{13}$$

where $p(e(n))$ is the loss function or the error measurement. One can calculate that the error measurement and the weighting function are coincidentally $G_\beta(e(n))$ in terms of MCC. In MMSE criterion, $p(e(n))$ is set to $e^2(n)$, then $q(e(i))$ equals to constant 2 and (12) will become the conventional RLS solution. The weighting functions of different criteria are displayed in Figure 2, which shows that MCC assigns very little weight on sample data that cause large estimation error compared with MMSE criterion and least p-norm criterion. Therefore, MCC is robust against outliers and able to diminish the impact of impulsive noises.

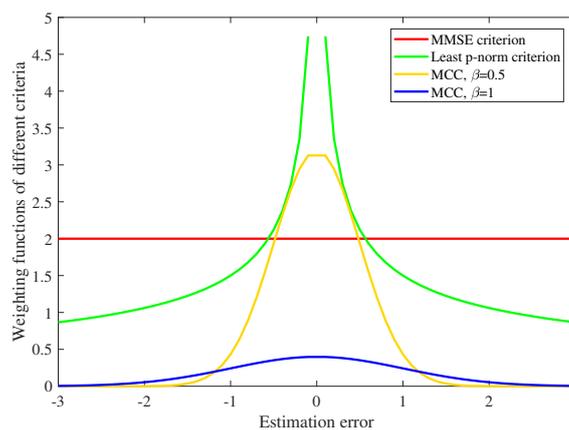


Figure 2. Weighting functions of different criteria.

Similar to the derivation process of the conventional LMS-Newton, the gradient of \mathbf{W} can be calculated as

$$\nabla_{MCC} = \frac{\partial J_{MCC}(n)}{\partial w} = 2\mathbf{R}_{MCC} \cdot \mathbf{W} - 2\mathbf{P}_{MCC}. \tag{14}$$

Adding a multiplier $1/2 * \mathbf{R}_{MCC}^{-1}$ to both sides of the formula, one can get

$$\mathbf{W}_{MCC}^* = \mathbf{W} - \frac{1}{2} \mathbf{R}_{MCC}^{-1} \nabla_{MCC}. \tag{15}$$

Meanwhile, from (10) we can derive the gradient ∇_{MCC} as

$$\nabla_{MCC} = \frac{1}{N} \sum_{n=1}^N G_\beta(e(n)) (e(n)) \mathbf{X}(n) \tag{16}$$

where we retain the information of all moments but not take the instantaneous gradient as a substitute like other algorithms.

Combining (15) and (16) one can derive the expression of the MCC-based Newtonian-type adaptive filtering as:

$$\mathbf{W}_{MCC}(n+1) = \mathbf{W}_{MCC}(n) + \mu \cdot \mathbf{R}_{MCC}^{-1}(n) \mathbf{F}_{MCC}(n) \mathbf{X}_{MCC}(n), \quad (17)$$

where $\mathbf{F}_{MCC}(n)$ is a vector in which each element is $G_\beta(e(n))$ ($e(n)$) of each time point, and $\mathbf{X}_{MCC}(n)$ is a matrix composed by vectors from $\mathbf{X}(1)$ to $\mathbf{X}(n)$. A step size μ is also added in the recursive equation to enhance the flexibility of the algorithm. Note that the MCC-Newton algorithm of this paper is different from that of [36], because the error vector and the input matrix in (17) contain the information of all the previous time points. In practical implementation of calculating the gradient according to (16), a forgetting factor and a sliding window could be also adopted to enhance the tracking ability and reduce the computational complexity, just like in the sliding-exponential window RLS adaptive filtering or sliding window LMS algorithm [41,42].

4. Steady-State Performance Analysis

The steady-state performance analysis [21,22,43,44] of adaptive filtering is a theoretical foundation and gives the guideline for the parameter selection.

Firstly, the assumptions used throughout the analysis are given as follows.

A1: The additive noise sequence $v(n)$ with variance σ_v^2 is independent and identically distributed (i.i.d.), and is independent of the input sequence $\mathbf{X}(n)$.

A2: The filter is enough long so that the a priori error $e_a(n)$ is zero-mean Gaussian and independent of the background noise $v(n)$.

We define the steady-state MSE as:

$$MSE = \lim_{n \rightarrow \infty} E \left[e^2(n) \right]. \quad (18)$$

The steady-state excess mean square error (EMSE) can be presented as:

$$EMSE = \lim_{n \rightarrow \infty} E \left[e_a^2(n) \right]. \quad (19)$$

So

$$MSE = EMSE + \sigma_v^2. \quad (20)$$

In our theoretical analysis of the steady-state behavior of the algorithm, the desired output of the unknown system $d(n)$ can be presented as:

$$d(n) = \mathbf{W}_0^T \mathbf{X}(n) + v(n), \quad (21)$$

where \mathbf{W}_0 is the optimal impulsive response vector of the adaptive filter which could not be measured directly, and $v(n)$ is the additive noise at time point n .

The weight error vector is presented as:

$$\tilde{\mathbf{W}}(n) = \mathbf{W}_0 - \mathbf{W}(n). \quad (22)$$

To simplify the structure of the discussed expressions, we assume that the width of the sliding window is 1, then (17) becomes

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu \cdot G_\beta(e(n)) e(n) \mathbf{R}^{-1}(n) \mathbf{X}(n). \quad (23)$$

Combining (23) and (22), one can obtain:

$$\tilde{\mathbf{W}}(n+1) = \tilde{\mathbf{W}}(n) - \mu \cdot G_\beta(e(n)) e(n) \mathbf{R}^{-1}(n) \mathbf{X}(n). \quad (24)$$

We define prior and posterior errors as

$$\begin{aligned} e_a(n) &= \tilde{\mathbf{W}}(n)^T \mathbf{X}(n) \\ e_p(n) &= \tilde{\mathbf{W}}(n+1)^T \mathbf{X}(n) \end{aligned} \tag{25}$$

Premultiplying $\mathbf{X}^T(n)$ on both sides of (24), one can get

$$e_p(n) = e_a(n) - \mu \cdot G_\beta(e(n))e(n)\mathbf{X}(n)^T \mathbf{R}^{-1}(n)\mathbf{X}(n). \tag{26}$$

As a stable algorithm, the MCC-Newton converges to the steady-state when time point n comes to a very large number. So the weight error vector satisfies [44]:

$$\lim_{n \rightarrow \infty} E \left[\|\tilde{\mathbf{W}}(n)\|^2 \right] = \lim_{n \rightarrow \infty} E \left[\|\tilde{\mathbf{W}}(n+1)\|^2 \right] \tag{27}$$

Combining (25) and (27), one can derive:

$$\lim_{n \rightarrow \infty} E \left[e_p(n)^2 \right] = \lim_{n \rightarrow \infty} E \left[e_a(n)^2 \right]. \tag{28}$$

At the steady-state, the time index n can be omitted for brevity since the distributions of the input and error signals are independent with n . We also omit the limit operators of the formulas and rewrite (28) as:

$$E \left(e_a^2 \right) = E \left(e_p^2 \right). \tag{29}$$

Substituting (26) into (29), one can derive that

$$E \left(e_a^2 \right) = E \left(e_a^2 \right) - 2E \left(e_a \mu G_\beta(e) e \cdot \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right) + E \left(\left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right)^2 \cdot \left(\mu G_\beta(e) e \right)^2 \right) \tag{30}$$

and

$$2E \left(e_a \mu G_\beta(e) e \right) = E \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \cdot \left(\mu G_\beta(e) e \right)^2 \right). \tag{31}$$

$\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X}$ is independent from $\mu G_\beta(e) e$, so that $E \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \cdot \left(\mu G_\beta(e) e \right)^2 \right) = E \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right) \cdot E \left(\left(\mu G_\beta(e) e \right)^2 \right)$.

Assume that $f(e) = \left(\mu G_\beta(e) e \right)$, then (31) becomes

$$2E \left(e_a \cdot f(e) \right) = E \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right) E \left(f(e)^2 \right). \tag{32}$$

It is known that $e(n) = e_a(n) + v(n)$. The simplification of (32) needs the Taylor expansion of $f(e)$ [21].

$$f(e) = f(v + e_a) = f(v) + f'(v)e_a + \frac{1}{2}f''(v)e_a^2 + o(e_a). \tag{33}$$

Then one can derive the following expressions:

$$E \left[e_a f(e) \right] = E \left[\left(e_a f(v) + f'(v)e_a^2 + o(e_a) \right) \right] \approx E \left[f'(v) \right] EMSE \tag{34}$$

and

$$E \left[f^2(e) \right] \approx E \left[f^2(v) \right] + E \left[f(v)f''(v) + |f'(v)|^2 \right] EMSE. \tag{35}$$

Substituting (34) and (35) into (32), one can obtain

$$EMSE = \frac{E \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right) \cdot E \left(f^2(v) \right)}{2E \left(f'(v) \right) - E \left(\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \right) \cdot E \left(f(v)f''(v) + f^2(v) \right)} \tag{36}$$

where $f(v) = \mu v \exp(-v^2/(2\beta^2))$, $f'(v) = \exp(-v^2/(2 * \beta^2)) - (v^2 \exp(-v^2/(2\beta^2)))/\beta^2$ and $f''(v) = (v^3 \exp(-v^2/(2\beta^2)))/\beta^4 - (3v \exp(-v^2/(2\beta^2)))/\beta^2$. Then, one can get the closed form solution of EMSE. The three parameters affecting the steady-state EMSE are step size μ , the noise deviation v^2 and β , the kernel width of MCC.

5. Experiments and Results

In this section, several experiments were carried out. In experiment 1, we simulated the algorithms in a simple system identification scenario to verify their effectiveness. Experiment 2 discussed the influence of the kernel width parameter selection of MCC-Newton under Gaussian and non-Gaussian noises. Experiment 3 compared Newtonian-type and LMS-type algorithms on the Correntropy performance surface. Experiment 4 compared the theoretical and the experimental results of EMSE of MCC-Newton. The effectiveness of MCC-Newton in practical acoustic echo cancellation was examined in experiment 5. Experiment 6 compared the echo return loss enhancement performance of different algorithms in practical acoustic echo cancellation.

Experiment 1:

The system impulsive response \mathbf{W}_o was set as a vector with 20 entries, which was consistent with the order of the adaptive filter. The 10th element of \mathbf{W}_o was set to 1, and the other elements were 0. The iteration number was set to 2000. The input signal followed Gaussian distribution with zero mean and unit variance. The desired signal was generated by the convolution of \mathbf{W}_o and the input signal, added by the system noise. We executed 100 Monte Carlo runs so that the average simulation result was available for the discussion. We simulated in both Gaussian and α stable noises, and compared two criteria (MMSE criterion and MCC)-based algorithms, each optimized with three different methods. The additive Gaussian noise had zero mean and the noise variance was 0.09. The additive α stable noise [45] was set with the same variance with the Gaussian noise, and the characteristic exponent α was set to 1.2, which is a measure of the thickness of the tails of the α stable distribution. When α is equal to 1 or 2, α stable distribution becomes the Cauchy distribution and the Gaussian distribution as special cases. To compare the convergence rate of the algorithms, step sizes were adjusted to ensure that the steady-state MSD of different algorithms were close to each other. The results are shown in Figures 3 and 4.

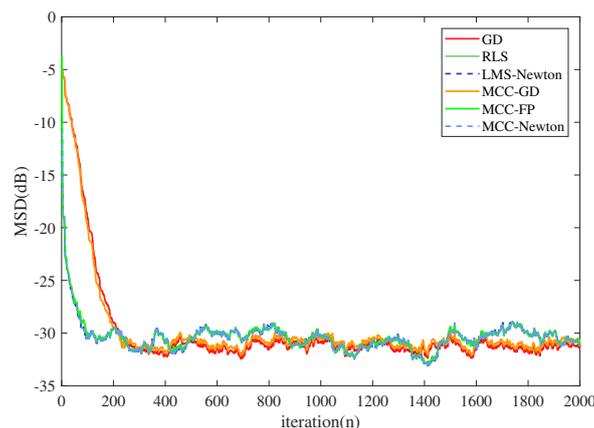


Figure 3. The performance of different algorithms under Gaussian noise.

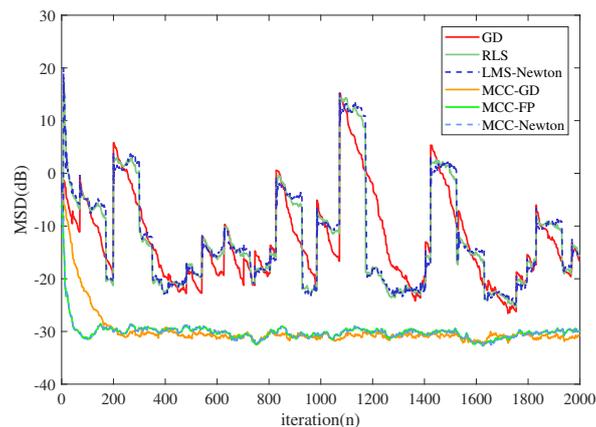


Figure 4. The performance of different algorithms under α stable noise.

From Figures 3 and 4 one can observe that: (1) all the algorithms converge to a steady-state in the presence of Gaussian noise. However, in α stable noise environment, MMSE criterion-based algorithms could not converge to a steady-state but MCC-based algorithms still converge well. (2) Newtonian-type and RLS-type algorithms converge in similar rate, and are much faster than LMS-type algorithms.

Experiment 2:

We simulated the performance of MCC-Newton algorithm with different kernel widths set and the results are presented in Figures 5 and 6. The parameter settings are the same as the previous experiment. From Figure 5 one can observe that under a Gaussian noise environment, the MCC-Newton algorithm performs as well as LMS-Newton algorithm, and the kernel width selection of MCC-Newton does not make much difference in terms of the steady-state MSD. However, as Figure 6 shows, LMS-Newton does not converge under the α stable noise, and MCC-Newton converges to a steady-state only when the kernel width is set small. One can also observe that MCC-Newton with a large kernel width performs similar to LMS-Newton which agrees well with the relation of the MMSE criterion and the MCC.

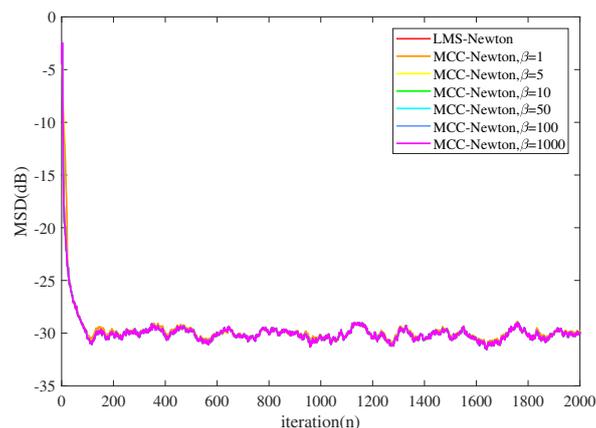


Figure 5. The performance of the MCC-Newton with different kernel width β under Gaussian noise.

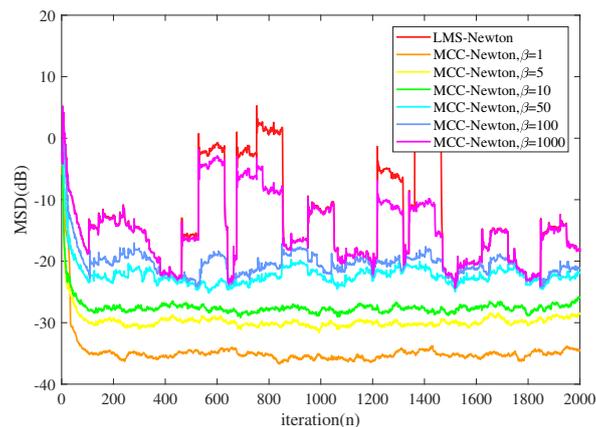


Figure 6. The performance of the maximum correntropy criterion (MCC)-Newton with different kernel width β under α stable noise.

Experiment 3:

To image the difference of the convergence process of MCC-based Newtonian-type and LMS-type algorithms, we simulated a system identification scenario of two tap weights, and presented the weight tracks of the algorithms on the correntropy performance surface [23].

The system impulsive response was set as [1,2], and the initial one was [1,−1]. Figure 7 shows the weight tracks (i.e., how the impulsive response \mathbf{W} changes to approach the optimal one \mathbf{W}_o) of different algorithms. One can observe that there are many twists and turns in the curve of MCC_GD than MCC_Newton, so that MCC_GD needs much more steps to reach an estimation close to \mathbf{W}_o , which is consistent with one of the results of experiment 1; that Newtonian-type algorithms converge faster than LMS-type ones. Figure 7 also shows that the ideal MCC_Newton algorithm possesses the best direction of optimization and the weight track of MCC_GD is approximately consistent with the gradient descent. However, it seems hard for the practical MCC_Newton algorithm to find the right direction at the beginning of the iteration. This can be explained by the temporary imprecision of the correlation matrix estimation in the practical MCC_Newton. The difference between the ideal and the practical MCC_Newton algorithms is the calculation of the correlation matrix of the input signal. The ideal MCC_Newton algorithm can use the perfectly calculated correlation matrix through the whole iteration, but the practical one can only build the estimated correlation matrix from the received input signals, which causes the imprecision of the estimation of the matrix at the beginning of the iteration. When the number of iterations increases, the estimation becomes more accurate, so that the practical MCC-Newton algorithm will achieve a similar performance with the ideal one after a certain number of iterations.

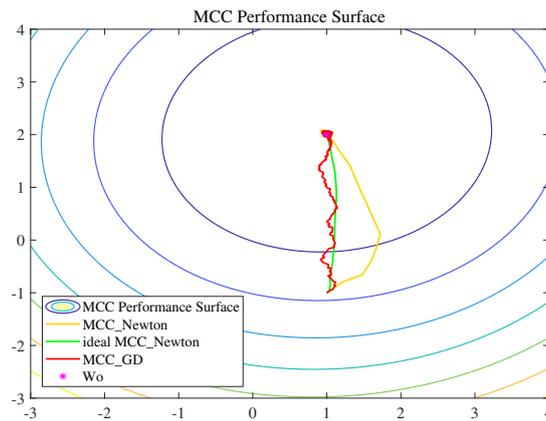


Figure 7. The weight tracks for the gradient descent-based algorithm (MCC_GD) and the Newtonian-type algorithm (ideal MCC_Newton and MCC_Newton) on the Correntropy performance surface.

Experiment 4:

In this experiment, we compared the theoretical derived EMSE in Section 4 with the simulated one to confirm the theoretical result. The input signal was generated from a Gaussian distribution with zero mean and unit variance, and the additive noise was Gaussian distributed with specified variance. We executed 100 Monte Carlo runs to get the average simulation result of EMSE. From (36), we can learn that the steady-state EMSE are determined by step size μ , the noise deviation v^2 and β , the kernel width of MCC. So we separately simulated the steady-state performance versus different parameters. In each simulation, one parameter varied with the other two parameters stayed unchanged, so we could display the influence of the specified parameter to the steady-state EMSE in Figures 8–10.

Figure 8 represents the theoretical and simulated EMSEs under different step sizes. The kernel width β was set to 5, and the noise variance σ_v^2 was set to 0.09. The step size μ scaled from 0.01 to 0.15. Figure 9 shows the theoretical and simulated EMSEs versus different kernel widths which scaled from 1 to 25. The step size was 0.07 and the noise variance was set as 0.09. Figure 10 displays the theoretical and simulated EMSEs under noise variances scaling from 0.01 to 0.25. The step size was 0.07 and the kernel width was set to 5. One could observe that the simulation results agrees well with the theoretical analysis. The steady-state performance of the algorithm degrades with the increase of the step size, the kernel width and the noise variance. In other words, we can choose smaller step size and kernel width to help MCC-Newton algorithm achieve better steady-state performance.

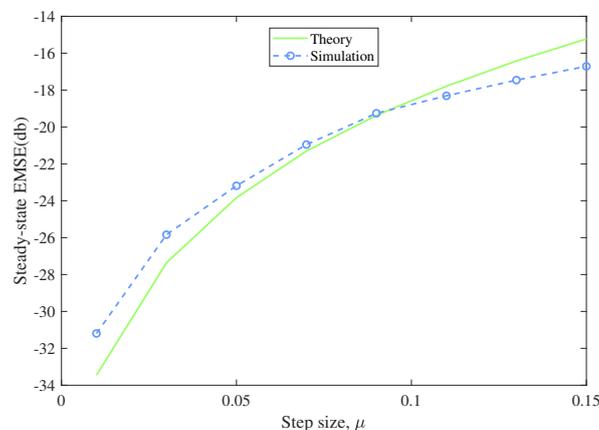


Figure 8. Theoretical and simulated EMSEs versus the step size μ ($\beta = 5, \sigma_v^2 = 0.09$).

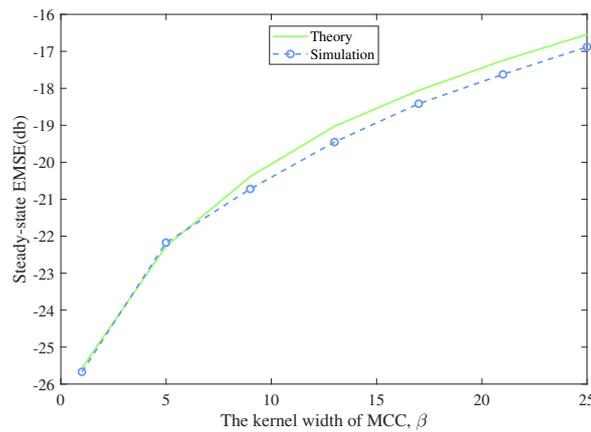


Figure 9. Theoretical and simulated excess mean square errors (EMSEs) versus the kernel width β ($\mu = 0.07, \sigma_v^2 = 0.09$).

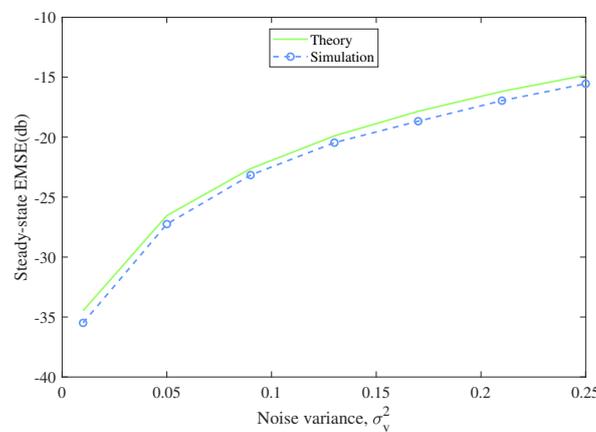


Figure 10. Theoretical and simulated EMSEs versus the noise variance σ_v^2 ($\mu = 0.07, \beta = 5$).

Experiment 5:

We examined the algorithm in practical acoustic echo cancellation and the parameters in this experiment were set as follows. We recorded two different simple voices that both last for 4.0 s as the near end speech and the far end speech. In ideal acoustic echo cancellation, the echo caused by the far end speech should be well cancelled from the mixed speech picked up by the near end microphone thus the output of the canceller could be close to the near end speech. The size of the impulsive response \mathbf{W}_o was set to 2000, which was large enough so that the generated echo can be heard clearly.

To evaluate the performance in AEC, we introduce the echo return loss enhancement (ERLE) which is defined as:

$$ERLE = \frac{E \left\{ (d(n) - \epsilon(n))^2 \right\}}{E \left\{ (e(n) - \epsilon(n))^2 \right\}}, \tag{37}$$

where $d(n)$ is the echo signal of the far end speech picked up by the microphone, which contains the additive noise $\epsilon(n)$ and the original echo signal. $e(n)$ is the residual error of AEC which is transmitted back to the far end together with the near end speech. So, ERLE tells us how large the estimated echo is compared with the residual error of AEC, without the influence of the additive noise.

We used the MCC-Newton algorithm in practical acoustic echo cancellation where the additive noise picked up by the near end microphone follows non-Gaussian distribution. To focus on the problem of AEC, it was assumed that we already had an ideal detection scheme to tell when near end speech was present. Figure 11 shows the practical effect and the ERLE performance of MCC-Newton.

One can observe that: (1) the microphone signal contains the near end speech, the far end speech and its echo signal added with the system noise. (2) The AEC output is free from the far-end speech echo and is very close to the original near-end speech. However, it could not get rid of the influence of noise. (3) ERLE becomes larger when there exists echo signal of the far end speech. The algorithm could achieve up to 50 dB ERLE.

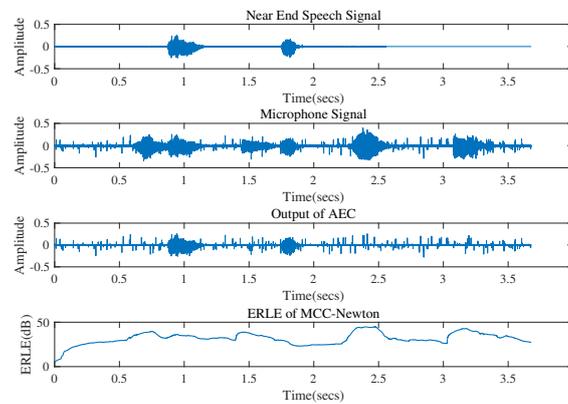


Figure 11. Effect of the MCC-Newton algorithm and the echo return loss enhancement performance.

Experiment 6:

We compared the ERLEs of MCC-Newton algorithms with that of LMS-Newton and the conventional LMS-type algorithms to see their performance under practical AEC scenario. The parameter settings are similar with that of the previous experiment. Figure 12 shows the result, from which one can observe that: (1) the ERLEs of all the algorithms tends to increase at the early stage of the experiment, and the ERLEs of Newtonian-type algorithms increase faster than that of LMS-type algorithms. The possible explanation is that the algorithms need to converge to a steady-state in the early stage iterations and Newtonian-type algorithms converge faster than LMS-type ones. (2) Compared with the other algorithms, MCC-Newton achieves higher ERLE performance. This confirms the practical effectiveness of MCC-Newton algorithm in the presence of heavy tailed mixed Gaussian noise.

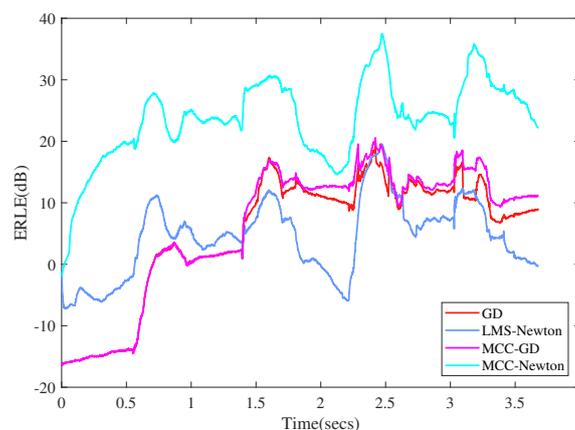


Figure 12. Echo return loss enhancement (ERLE) of different algorithms.

6. Conclusions

MCC is recently adopted to deal with the heavy tailed impulsive noise problem in robust adaptive filtering. In this paper, a Newtonian-type method is innovatively used to solve the MCC-based adaptive

filtering, of which the existing optimization methods are usually LMS-type or RLS-type. Experiments demonstrate that the Newtonian-type MCC-based adaptive filtering converges as fast as RLS-type ones, and is much faster than LMS-type. The steady-state performance of the Newtonian-type MCC adaptive filtering is theoretically analyzed and is verified to be consistent with the experiment results. It is also revealed in experiments that a smaller kernel width helps the Newtonian-type MCC-based adaptive filtering perform better under impulsive noises, which could be a guide for the selection of the parameter. Experiments also show that the MCC-Newton algorithm is practical in acoustic echo cancellation under heavy-tailed system noise. In future work, better approximation methods of the Hessian matrix could be involved to decrease the computational complexity and further improve the performance of the algorithm.

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

Funding: This work was supported by the National Key Research and Development Project under Grant 2018YFB18003600 and the National Natural Science Foundation of China under Grant 61531013.

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

References

1. Widrow, B.; Stearns, S.D. *Adaptive Signal Processing*; Prentice-Hall: Upper Saddle River, NJ, USA, 2008.
2. Schafhuber, D.; Matz, G.; Hlawatsch, F. Adaptive Wiener filters for time-varying channel estimation in wireless OFDM systems. In Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), Hong Kong, China, 6–10 April 2003; Volume 4, p. IV-688, doi:10.1109/ICASSP.2003.1202736. [[CrossRef](#)]
3. Doukopoulos, X.G.; Moustakides, G.V. Blind adaptive channel estimation in ofdm systems. *IEEE Trans. Wirel. Commun.* **2006**, *5*, 1716–1725, doi:10.1109/TWC.2006.1673083. [[CrossRef](#)]
4. Benesty, J.; Amand, F.; Gilloire, A.; Grenier, Y. Adaptive filtering algorithms for stereophonic acoustic echo cancellation. In Proceedings of the 1995 International Conference on Acoustics, Speech, and Signal Processing, Detroit, MI, USA, 9–12 May 1995; Volume 5, pp. 3099–3102.
5. Aung, M.T.S.; Shi, Z.; Kikuuwe, R. A new noise-reduction filter with sliding mode and low-pass filtering. In Proceedings of the 2014 IEEE Conference on Control Applications (CCA), Juan Les Antibes, France, 8–10 October 2014; pp. 1029–1034.
6. Haykin, S. *Adaptive Filter Theory*; Prentice Hall: Upper Saddle River, NJ, USA, 2002; Volume 4.
7. Madisetti, V. *Digital Signal Processing Fundamentals*, 2nd ed.; CRC Press, Inc.: Boca Raton, FL, USA, 2009.
8. Petrus, P. Robust Huber adaptive filter. *IEEE Trans. Signal Process.* **1999**, *47*, 1129–1133. [[CrossRef](#)]
9. Ma, W.; Chen, B.; Qu, H.; Zhao, J. Sparse least mean p-power algorithms for channel estimation in the presence of impulsive noise. *Signal Image Video Process.* **2015**, *10*, doi:10.1007/s11760-015-0757-5. [[CrossRef](#)]
10. Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer Publishing Company, Incorporated: New York, NY, USA, 2010; pp. 1385–1392.
11. Chen, B.; Zhu, Y.; Hu, J.; Principe, J.C. *System Parameter Identification: Information Criteria and Algorithms*; Tsinghua University Press: Beijing, China, 2014.
12. Liu, X.; Qu, H.; Zhao, J.; Yue, P.; Wang, M. Maximum Correntropy Unscented Kalman Filter for Spacecraft Relative State Estimation. *Sensors* **2016**, *16*, 1530. [[CrossRef](#)] [[PubMed](#)]
13. Liu, X.; Qu, H.; Zhao, J.; Yue, P. Maximum correntropy square-root cubature Kalman filter with application to SINS/GPS integrated systems. *ISA Trans.* **2018**, *80*, 195–202, doi:10.1016/j.isatra.2018.05.001. [[CrossRef](#)]
14. Yue, P.; Qu, H.; Zhao, J.; Wang, M.; Liu, X. A robust blind adaptive multiuser detection based on maximum correntropy criterion in satellite CDMA systems. *Trans. Emerg. Telecommun. Technol.* **2019**, *30*, e3605, doi:10.1002/ett.3605. [[CrossRef](#)]
15. 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, Atlanta, GA, USA, 14–19 June 2009; pp. 2950–2955.

16. Ma, W.; Qu, H.; Gui, G.; Xu, L.; 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](#)]
17. 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](#)]
18. He, Y.; Wang, F.; Yang, J.; Rong, H.; Chen, B. Kernel adaptive filtering under generalized Maximum Correntropy Criterion. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 1738–1745.
19. Yue, P.; Qu, H.; Zhao, J.; Wang, M. An Adaptive Channel Estimation Based on Fixed-Point Generalized Maximum Correntropy Criterion. *IEEE Access* **2020**, *8*, 66281–66290. [[CrossRef](#)]
20. Liu, X.; Qu, H.; Zhao, J.; Chen, B. State space maximum correntropy filter. *Signal Process.* **2017**, *130*, 152–158. [[CrossRef](#)]
21. 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.
22. Wang, W.; Zhao, J.; Qu, H.; Chen, B.; Principe, J.C. Convergence performance analysis of an adaptive kernel width MCC algorithm. *AEU Int. J. Electron. Commun.* **2017**, *76*, 71–76. [[CrossRef](#)]
23. Singh, A.; Principe, J.C. A closed form recursive solution for Maximum Correntropy training. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14–19 March 2010; pp. 2070–2073.
24. Chen, B.; Wang, J.; Zhao, H.; Zheng, N.; Principe, J.C. Convergence of a Fixed-Point Algorithm under Maximum Correntropy Criterion. *IEEE Signal Process. Lett.* **2015**, *22*, 1723–1727. [[CrossRef](#)]
25. Zhao, J.; Zhang, H.; Wang, G. Fixed-point generalized maximum correntropy: Convergence analysis and convex combination algorithms. *Signal Process.* **2019**, *154*, 64–73. [[CrossRef](#)]
26. Agarwal, R.P.; Meehan, M.; O'Regan, D. *Fixed Point Theory and Applications*; Nova Science Publishers Inc.: Huntington, NY, USA, 2001; Volume 101, pp. 1–16.
27. Chen, B.; Liu, X.; Zhao, H.; Principe, J.C. Maximum correntropy kalman filter. *Automatica* **2017** *76*, 70–77. [[CrossRef](#)]
28. Miranda, V.; Santos, A.; Pereira, J. State Estimation Based on Correntropy: A Proof of Concept. *IEEE Trans. Power Syst.* **2009**, *24*, 1888–1889. [[CrossRef](#)]
29. Heravi, A.R.; Hodtani, G.A. A new robust correntropy based Levenberg-Marquardt algorithm. In Proceedings of the 2016 Iran Workshop on Communication and Information Theory (IWCIT), Tehran, Iran, 3–4 May 2016; pp. 1–6.
30. Heravi, A.R.; Hodtani, G.A. Comparison of the Convergence Rates of the New Correntropy-Based Levenberg–Marquardt (CLM) Method and the Fixed-Point Maximum Correntropy (FP-MCC) Algorithm. *Circuits Syst. Signal Process.* **2018**, *37*, 2884–2910. [[CrossRef](#)]
31. Bhotto, M.Z.A.; Antoniou, A. Robust Quasi-Newton Adaptive Filtering Algorithms. *IEEE Trans. Circuits Syst. II Express Briefs* **2011**, *58*, 537–541. [[CrossRef](#)]
32. Rao, H.I.K.; Farhang-Boroujeny, B. Fast LMS/Newton Algorithms for Stereophonic Acoustic Echo Cancellation. *IEEE Trans. Signal Process.* **2009**, *57*, 2919–2930. [[CrossRef](#)]
33. Farhang-Boroujeny, B. Fast LMS/Newton algorithms based on autoregressive modeling and their application to acoustic echo cancellation. *IEEE Trans. Signal Process.* **1997**, *45*, 1987–2000. [[CrossRef](#)]
34. de Campos, M.L.R.; Antoniou, A. A robust quasi-Newton adaptive filtering algorithm. In Proceedings of the IEEE International Symposium on Circuits and Systems—ISCAS '94, London, UK, 30 May–2 June 1994; Volume 2, pp. 229–232.
35. Dash, P.; Krishnanand, K.; Padhee, M. Fast recursive Gauss–Newton adaptive filter for the estimation of power system frequency and harmonics in a noisy environment. *IET Gener. Transm. Distrib.* **2011**, *5*, 1277–1289. [[CrossRef](#)]
36. Zhou, Y.; Chan, S.C.; Ho, K.L. A new LMS/Newton algorithm for robust adaptive filtering in impulsive noise. In Proceedings of the 2004 12th European Signal Processing Conference, Vienna, Austria, 6–10 September 2004; pp. 705–708.
37. De Campos, M.L.R.; Antoniou, A. A new quasi-Newton adaptive filtering algorithm. *IEEE Trans. Circuits Syst. II Analog. Digit. Signal Process.* **1997**, *44*, 924–934. [[CrossRef](#)]

38. Marshal, D.F.; Jenkins, W.K. A fast quasi-Newton adaptive filtering algorithm. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP-88), New York, NY, USA, 11–14 April 1988; Volume 3, pp. 1377–1380.
39. Berberidis, K.; Rantos, S.; Palicot, J. A step-by-step quasi-Newton algorithm in the frequency domain and its application to adaptive channel equalization. *IEEE Trans. Signal Process.* **2004**, *52*, 3335–3344. [[CrossRef](#)]
40. Chen, L.; Qu, H.; Zhao, J.; Chen, B.; Principe, J.C. Efficient and robust deep learning with Correntropy-induced loss function. *Neural Comput. Appl.* **2016**, *27*, 1019–1031. [[CrossRef](#)]
41. Song, H.-K. A channel estimation using sliding window approach and tuning algorithm for MLSE. *IEEE Commun. Lett.* **1999**, *3*, 211–213. [[CrossRef](#)]
42. Liu, H.; He, Z. A sliding-exponential window RLS adaptive filtering algorithm: Properties and applications. *Signal Process.* **1995**, *45*, 357–368. [[CrossRef](#)]
43. Qu, H.; Shi, Y.; Zhao, J. A Smoothed Algorithm with Convergence Analysis under Generalized Maximum Correntropy Criteria in Impulsive Interference. *Entropy* **2019**, *21*, 1099. [[CrossRef](#)]
44. Lu, L.; Zhao, H. Steady-state performance analysis of the recursive maximum correntropy algorithm and its application in adaptive beamforming with alpha-stable noise. *arXiv* **2017**, arXiv:1701.08407.
45. Kuruoglu, E.E.; Fitzgerald, W.J.; Rayner, P.J.W. Near optimal detection of signals in impulsive noise modeled with a symmetric/spl alpha/-stable distribution. *IEEE Commun. Lett.* **1998**, *2*, 282–284. [[CrossRef](#)]



© 2020 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/>).