

Article

A Sparsity-Aware Variable Kernel Width Proportionate Affine Projection Algorithm for Identifying Sparse Systems

Zhengxiong Jiang ¹, Yingsong Li ^{1,2,*} , Xinqi Huang ¹ and Zhan Jin ¹

¹ College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; jiangzhengxiong@hrbeu.edu.cn (Z.J.); huangxinqi@hrbeu.edu.cn (X.H.); jinzhan@hrueu.edu.cn (Z.J.)

² Key Laboratory of Microwave Remote Sensing, Chinese Academy of Sciences, Beijing 100190, China

* Correspondence: liyingsong@ieee.org

Received: 20 August 2019; Accepted: 25 September 2019; Published: 1 October 2019



Abstract: A sparsity-aware variable kernel width proportionate affine projection (AP) algorithm is devised for identifying sparse system in impulsive noise environments. For the devised algorithm, the symmetry maximum correntropy criterion (MCC) is employed to develop a new cost function for improving the PAP algorithm, then the variable kernel width and the l_p -norm-like constraint are incorporated into the cost-function, which is named as l_p -norm variable kernel width proportionate affine projection (LP-VPAP) algorithm. The devised LP-VPAP algorithm is investigated and verified under impulsive interference environments. Experimental results show that the LP-VPAP gets a faster convergence and provides a lower steady-state performance compared with AP, zero-attracting AP (ZA-AP), reweighted ZA-AP (RZA-AP), proportionate AP (PAP), MCC, variable kernel width MCC (VKW-MCC), and proportionate AP MCC (PAPMCC) algorithms.

Keywords: maximum correntropy criterion; l_p -norm; sparse system identification; impulsive interferences; proportionate affine projection algorithm

1. Introduction

Today, with the evolution of information technology, adaptive filtering (AF) algorithms are used in wireless communication, noise reduction, system identification (SI), and automatic control systems [1–6]. The famed least-mean-square (LMS) is one of the classic algorithms and is generally used [7,8]. Then, the normalized version of the LMS (NLMS) is proposed to improve the stability as well as increase the convergence performance [9]. However, the LMS-like algorithms encounter poor performance while the input is a strongly correlated signal. On the basis of the NLMS, the affine projection (AP) is developed by reusing input signal [10,11], which is regarded as a multi-order generalization of the NLMS algorithm, reusing the sample values of the current time and the previous time of the input signal. Compared with the LMS and NLMS, the behavior of the AP algorithms is more outstanding, especially for the colored inputs [12].

Recently, scholars have found that sparse systems are widespread in nature [13–18]. For example, in underwater communication systems, the underwater acoustic channel exhibits strong sparsity [19]. In network communications, the network echo channel also exhibits sparsity. [20]. However, the traditional LMS, NLMS and AP algorithms show great potential that can be further improved to use the sparse characteristic in the system's impulse response (IR). To fully use the sparse characteristic, the literature [21] has proposed the famed proportionate NLMS (PNLMS) algorithm by introducing a proportionate update technique to reasonably allocate the step sizes corresponding to each filter tap coefficient. Inspired by the PNLMS, the proportionate AP (PAP) algorithm was proposed by using the

idea in PNLMS to fully use the sparsity in the system via employing the data reusing principle [22]. Then, various proportionate-type AF algorithms were proposed and analyzed [23–25]. Moreover, a collection of zero-attraction (ZA) algorithms, such as the ZA-LMS and its reweighted form (RZA-LMS), ZA-AP, and RZA-AP algorithms, etc. [26–33], are developed based on the concept of compressed sensing (CS) theory [34] for sparse SI.

In many engineering fields, noise often exhibits strongly impulsive characteristics [35]. Traditional LMS-type and AP-type algorithms, which use the $E\{e^2(n)\}$ (where $e(n)$ is the error signal, and $E\{\cdot\}$ represents the expectation operator) to construct an expected cost-function, will encounter poor performance in impulsive noise environments. To find out the solution for handling these problems, the maximum correntropy criterion (MCC) was proposed to give resistance to impulse-noise [36,37]. Then, a series of AF algorithms based on MCC were proposed to resist impulse-noise [38–48]. The variable kernel width MCC (VKW-MCC) uses a variable kernel width technique to enhance the identification ability of the famed MCC [49]. The proportionate affine projection symmetry maximum correntropy criterion (PAPMCC) uses the MCC and the data reusing technique to enhance the robustness of the PAP [50].

The MCC criterion is used in this paper to build a new cost function to improve PAP algorithm, then integrate the variable kernel width technique and the l_p -norm-like constraint into the cost-function to develop the l_p -norm variable kernel width proportionate affine projection (LP-VPAP). The devised LP-VPAP algorithm is investigated and verified under impulsive noise environments. Experimental results show that the LP-VPAP converges the fastest and gets the lowest steady state misalignment compared with AP, ZA-AP, RZA-AP, PAP, MCC, VKW-MCC, and PAPMCC algorithms.

2. Previous Work of the MCC and AP Algorithms

2.1. The Basic MCC Algorithm

In the range of AF, the implementation schematic diagram for SI is presented in Figure 1.

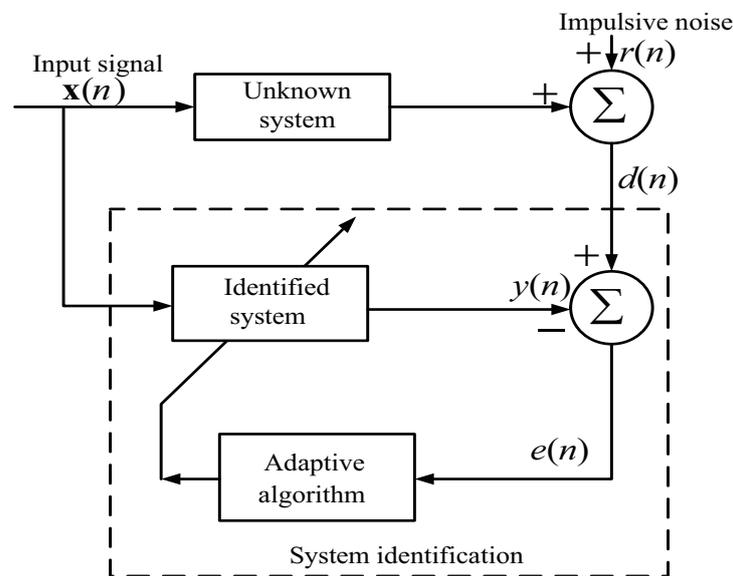


Figure 1. The schematic diagram of SI.

$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$ is used as system input signal, and the IR is modeled as $\mathbf{h}(n) = [h_0(n), \dots, h_{L-1}(n)]^T$ with L elements, and n represents the time index. The achieved signal $d(n)$ is

$$d(n) = \mathbf{x}^T(n)\mathbf{h}(n) + r(n), \quad (1)$$

in which $r(n)$ denotes the additive interferences. The output signal is defined as

$$y(n) = \mathbf{x}^T(n)\hat{\mathbf{h}}(n), \quad (2)$$

in which $\hat{\mathbf{h}}(n)$ stands for the estimation vector of IR. $e(n)$ is the estimated error written as

$$e(n) = d(n) - y(n). \quad (3)$$

The famed MCC aims to solve the following cost-function

$$J_{\text{MCC}}(n) = E \left[\exp \left(-\frac{e^2(n)}{2\sigma^2} \right) \right], \quad (4)$$

where σ represents Gaussian kernel. Removing expectation operator and computing the gradient of Equation (4), the iterative formula of the MCC is

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu_{\text{MCC}} \exp\left(-\frac{e^2(n)}{2\sigma^2}\right)e(n)\mathbf{x}(n), \quad (5)$$

in which μ_{MCC} is the step size.

2.2. AP Algorithm

The AP algorithm reuses the information from previous instants to accelerate the convergence, particularly for colored input. The input data of the AP is expressed as an $L \times P$ matrix

$$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \mathbf{x}(n-2), \dots, \mathbf{x}(n-P+1)]. \quad (6)$$

Herein, P denotes the order of the projection. The reference signal in the AP algorithm is $\mathbf{d}(n) = [d(n), d(n-1), d(n-2), \dots, d(n-P+1)]^T$. Then, the output is

$$\mathbf{y}(n) = \mathbf{X}^T(n)\hat{\mathbf{h}}(n) \quad (7)$$

with the error vector

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{y}(n). \quad (8)$$

The iterative formula of the famous AP is denoted as

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu_{\text{AP}}\mathbf{X}(n) \left[\mathbf{X}^T(n)\mathbf{X}(n) + \delta_{\text{AP}}\mathbf{I}_P \right]^{-1} \mathbf{e}(n), \quad (9)$$

in which μ_{AP} represents the total step size, $\delta_{\text{AP}} > 0$ has a small value, and \mathbf{I}_P is a $P \times P$ identity matrix.

3. The Developed LP-VPAP Algorithm

Herein, we detailedly analyze the developed LP-VPAP algorithm, which is implemented by the variable kernel width technique and the l_p -norm-like constraint to devise a novel cost-function. In addition, it can be summarized as the following problem:

$$\begin{aligned} & \frac{1}{2} \left\| \hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n) \right\|_{\mathbf{K}^{-1}(n)}^2 + \rho \left\| \mathbf{K}^{-1}(n)\hat{\mathbf{h}}(n+1) \right\|_p^p \\ & \text{subject to } \tilde{\mathbf{e}}(n) = \left[\mathbf{1}_P - \zeta \exp \left(-\frac{\mathbf{e}(n) \odot \mathbf{e}(n)}{2\sigma^2} \right) \right] \odot \mathbf{e}(n) \end{aligned} \quad (10)$$

where $\check{\mathbf{e}}(n) = \mathbf{d}(n) - \mathbf{X}^T(n)\hat{\mathbf{h}}(n+1)$, $\|\hat{\mathbf{h}}(n+1)\|_p^p = \sum_{i=1}^L |h_i|^p$, $\mathbf{1}_P$ is a $P \times 1$ vector with all the elements equal to 1, ζ denotes the step size, and $\mathbf{e}(n) \odot \mathbf{e}(n)$ is the Hadamard product. $\mathbf{K}(n)$ is a gain assignment matrix which allots different gain to each coefficient [21]:

$$\mathbf{K}(n) = \text{diag} \{k_0(n), k_1(n), \dots, k_{L-1}(n)\}. \quad (11)$$

The individual gain $k_l(n)$ is

$$k_l(n) = \frac{\varphi_l(n)}{\sum_{i=0}^{L-1} \varphi_i(n-1)}, \quad (12)$$

with

$$\begin{aligned} \varphi_l(n-1) &= \max \left\{ \alpha q, |\hat{h}_l| \right\}, \\ q &= \max \left\{ \beta, |\hat{h}_0|, |\hat{h}_1|, |\hat{h}_2|, \dots, |\hat{h}_{L-1}| \right\}. \end{aligned} \quad (13)$$

Parameters α and β can maintain coefficient updating from halting. Employing the Lagrange multiplier (LM) method, the cost-function of the developed LP-VPAP is

$$\begin{aligned} J(\hat{\mathbf{h}}(n+1)) &= \frac{1}{2} \|\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)\|_{\mathbf{K}^{-1}(n)}^2 + \rho \|\mathbf{K}^{-1}(n)\hat{\mathbf{h}}(n+1)\|_p^p \\ &+ \lambda \left\{ \check{\mathbf{e}}(n) - \left[\mathbf{1}_P - \zeta \exp \left(-\frac{\mathbf{e}(n) \odot \mathbf{e}(n)}{2\sigma^2} \right) \right] \odot \mathbf{e}(n) \right\} \end{aligned} \quad (14)$$

in which $\lambda = [\lambda_0, \lambda_1, \dots, \lambda_{P-1}]$. Then, the gradients of $J(\hat{\mathbf{h}}(n+1))$ with respect to $\hat{\mathbf{h}}(n+1)$ and λ are calculated, respectively.

$$\frac{\partial J(\hat{\mathbf{h}}(n+1))}{\partial \hat{\mathbf{h}}(n+1)} = \mathbf{K}^{-1}(n) (\hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n)) + \rho \frac{p \mathbf{K}^{-1}(n) \text{sgn}(\hat{\mathbf{h}}(n+1))}{\varepsilon + |\hat{\mathbf{h}}(n+1)|^{1-p}} - \mathbf{X}(n) \lambda^T \quad (15)$$

and

$$\frac{\partial J(\hat{\mathbf{h}}(n+1))}{\partial \lambda} = \check{\mathbf{e}}(n) - \left[\mathbf{1}_P - \zeta \exp \left(-\frac{\mathbf{e}(n) \odot \mathbf{e}(n)}{2\sigma^2} \right) \right] \odot \mathbf{e}(n). \quad (16)$$

Assume that $\hat{\mathbf{h}}(n+1) \approx \hat{\mathbf{h}}(n)$ and let

$$\begin{aligned} \frac{\partial J(\hat{\mathbf{h}}(n+1))}{\partial \hat{\mathbf{h}}(n+1)} &= \mathbf{0} \\ \frac{\partial J(\hat{\mathbf{h}}(n+1))}{\partial \lambda} &= \mathbf{0}. \end{aligned} \quad (17)$$

Then, one can obtain

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mathbf{K}(n) \mathbf{X}(n) \lambda^T - \rho \frac{p \text{sgn}(\hat{\mathbf{h}}(n))}{\varepsilon + |\hat{\mathbf{h}}(n)|^{1-p}}, \quad (18)$$

where ε is a small positive constant which aims to prevent division by 0 and

$$\mathbf{d}(n) = \mathbf{X}^T(n) \hat{\mathbf{h}}(n+1) + \left[\mathbf{1}_P - \zeta \exp \left(-\frac{\mathbf{e}(n) \odot \mathbf{e}(n)}{2\sigma^2} \right) \right] \odot \mathbf{e}(n). \quad (19)$$

From Equations (18) and (19), we have

$$\lambda^T = \zeta [\mathbf{X}^T(n) \mathbf{K}(n) \mathbf{X}(n)]^{-1} \exp \left(-\frac{\mathbf{e}(n) \odot \mathbf{e}(n)}{2\sigma^2} \right) \odot \mathbf{e}(n) + \rho [\mathbf{X}^T(n) \mathbf{K}(n) \mathbf{X}(n)]^{-1} \mathbf{X}^T(n) \frac{p \text{sgn}(\hat{\mathbf{h}}(n))}{\varepsilon + |\hat{\mathbf{h}}(n)|^{1-p}} \quad (20)$$

Then, the iterative formula is denoted as

$$\begin{aligned} \hat{\mathbf{h}}(n+1) = & \hat{\mathbf{h}}(n) + \zeta \mathbf{K}(n) \mathbf{X}(n) \left[\mathbf{X}^T(n) \mathbf{K}(n) \mathbf{X}(n) \right]^{-1} \exp \left(-\frac{\mathbf{e}(n) \odot \mathbf{e}(n)}{2\sigma^2} \right) \odot \mathbf{e}(n) \\ & + \rho \mathbf{K}(n) \mathbf{X}(n) \left[\mathbf{X}^T(n) \mathbf{K}(n) \mathbf{X}(n) \right]^{-1} \mathbf{X}^T(n) \frac{p \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\varepsilon + |\hat{\mathbf{h}}(n)|^{1-p}} - \rho \frac{p \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\varepsilon + |\hat{\mathbf{h}}(n)|^{1-p}} \end{aligned} \quad (21)$$

In practice, the third term in Equation (21) plays a minor role and hence it can be ignored here [51,52]. To enhance the behavior of correntropy-based algorithm, a variable kernel width technique is used to automatically adjust the σ in Equation (4). Suppose that there is no impulse-interference. After obtaining $e(n)$, the kernel-width which yields the error with the maximum damping along the direction of gradient-ascent is solved by

$$\max_{\sigma(n)} J_{\text{MCC}}(n) = \exp \left(-\frac{e^2(n)}{2\sigma^2(n)} \right). \quad (22)$$

For Equation (22), calculate its derivative in terms of $e(n)$, then let the derivative be zero, we obtain

$$\sigma(n) = e(n) \sigma'(n). \quad (23)$$

Solving homogeneous differential Equation (23), we can get

$$\sigma(n) = \theta |e(n)|. \quad (24)$$

$\sigma(n)$ represents the optimal kernel-width, and θ denotes a positive constant. In practical terms, the received signal could contain impulse-interferences. To get the robustness, a observed sliding window of $e(n)$ is used to optimize $\sigma(n)$.

$$\sigma(n) = \min(\sigma_0, \theta \bar{e}(n)), \quad (25)$$

in which σ_0 represents a kernel bound, and $\bar{e}(n)$ is denoted as

$$\bar{e}(n) = \chi \bar{e}(n-1) + (1-\chi) \min [e(n), e(n-1), \dots, e(n-N_0+1)], \quad (26)$$

in which $0 \ll \chi < 1$ denotes a forgetting factor, and N_0 represents the observation length. Then, Equation (21) is changed to be

$$\begin{aligned} \hat{\mathbf{h}}(n+1) = & \hat{\mathbf{h}}(n) + \zeta \mathbf{K}(n) \mathbf{X}(n) \left[\mathbf{X}^T(n) \mathbf{K}(n) \mathbf{X}(n) + \delta_{\text{PAP}} \mathbf{I}_P \right]^{-1} \exp \left(-\frac{\mathbf{e}(n) \odot \mathbf{e}(n)}{2\sigma^2(n)} \right) \odot \mathbf{e}(n) \\ & - \rho \frac{p \operatorname{sgn}(\hat{\mathbf{h}}(n))}{\varepsilon + |\hat{\mathbf{h}}(n)|^{1-p}} \end{aligned} \quad (27)$$

Equation (27) is the final iterative formula for the developed LP-VPAP algorithm.

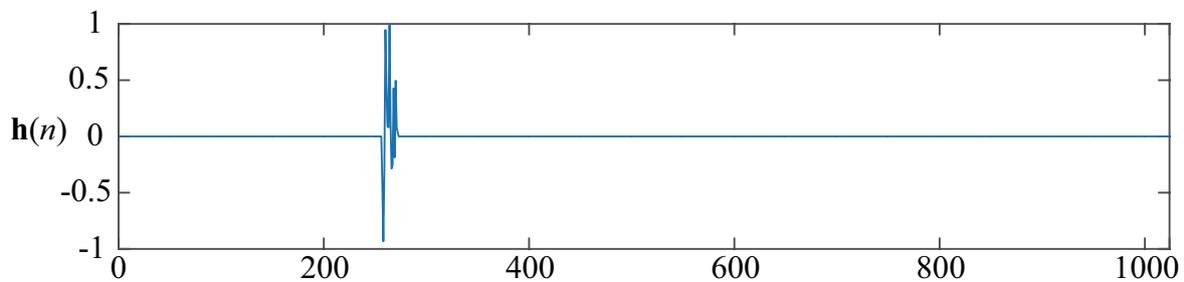
The computation complexity of the LP-VPAP is compared with MCC, VKW-MCC, AP, ZA-AP, RZA-AP, PAP, and PAPMCC algorithms with respect to the total number of addition, multiplication, and division in each iteration. The comparison is presented in Table 1. It is clear to see that the LP-VPAP algorithm has a modest increase in computational complexity compared with that of the PAPMCC algorithm.

Table 1. Computational Complexity in Each Iteration.

Algorithm	Addition	Multiplication	Division
MCC	$2L$	$2L + 5$	1
VKW-MCC	$2L + 2$	$2L + 8$	1
AP	$(2P^2 + P)L$	$(2P^2 + 3P)L + P^2$	0
ZA-AP	$(2P^2 + P + 1)L$	$(2P^2 + 3P + 1)L + P^2$	0
RZA-AP	$(2P^2 + P + 2)L$	$(2P^2 + 3P + 2)L + P^2$	L
PAP	$2PL^2 + (2P^2 - P + 1)L - 1$	$2PL^2 + (2P^2 + 3P + 1)L + P^2$	L
PAPMCC	$2PL^2 + (2P^2 - P + 1)L - 1$	$2PL^2 + (2P^2 + 3P + 1)L + P^2 + 2P$	$L + P$
LP-VPAP	$2PL^2 + (2P^2 - P + 3)L + 1$	$2PL^2 + (2P^2 + 3P + 2)L + P^2 + 2P + 4$	$2L + P$

4. Performance Analysis

Herein, several examples are presented to investigate the performance of the devised LP-VPAP with the framework of SI. The background noise $r(n)$ consists of the impulsive noise $i(n)$ and white-Gaussian-noise (WGN) $v(n)$, where $i(n)$ is modeled by the Bernoulli distribution for $i(n) = b(n)g(n)$, $g(n)$ is another WGN and $b(n)$ is Bernoulli-process for $P(b(n) = 1) = 0.1$. The signal-noise-ratio (SNR) and signal-interface-ratio (SIR), which are defined by $\text{SNR} = 10 \log_{10} \frac{\delta_x^2}{\delta_v^2}$ and $\text{SIR} = 10 \log_{10} \frac{\delta_x^2}{\delta_i^2}$ (where δ_x^2 , δ_v^2 , and δ_i^2 represent the variances of $x(n)$, $v(n)$, and $i(n)$, respectively), are set as 30 dB and 10 dB, respectively. In all experiments, $L = 1024$ and $P = 4$ are selected. Regularization parameters are set to be $\delta_{\text{AP}} = \delta_{\text{ZA-AP}} = \delta_{\text{RZA-AP}}$ and $\delta_{\text{LP-VPAP}} = \delta_{\text{PAP}} = \frac{1}{L} \delta_{\text{AP}}$ [53]. The kernel width of the MCC is set as 1, and the observation length of VKW-MCC and LP-VPAP algorithms is set to be 25. The network echo channel used for the simulation, which is a classical sparse channel presented in Figure 2, whose active coefficients distributed in [257,272], is considered to evaluate the proposed LP-VPAP. The behavior of the LP-VPAP is evaluated by normalized-misalignment (NM) that has a definition of $10 \log_{10} (\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2 / \|\mathbf{h}\|_2^2)$.

**Figure 2.** The IR used in the examples.

4.1. Performance of the LP-VPAP with Different p and ρ

Firstly, the effects of p and ρ on the behavior for devised LP-VPAP algorithm is investigated. The colored noise (CN), which is obtained from WGN filtering through an autoregressive with a pole at 0.8, is used as the input. Herein, $\xi = 0.225$. The results given in Figure 3 illustrate that when p is selected to 0.5 or 0.3, the LP-VPAP algorithm achieves the fastest convergence.

Secondly, the effects of ρ on the convergence for the LP-VPAP is analyzed and discussed. Herein, $p = 0.5$ is selected. The results given in Figure 4 illustrate that the NM of the LP-VPAP algorithm increases with the increment of ρ , but the convergence rate becomes faster. Therefore, a balance between convergence and NM should be taken into consideration.

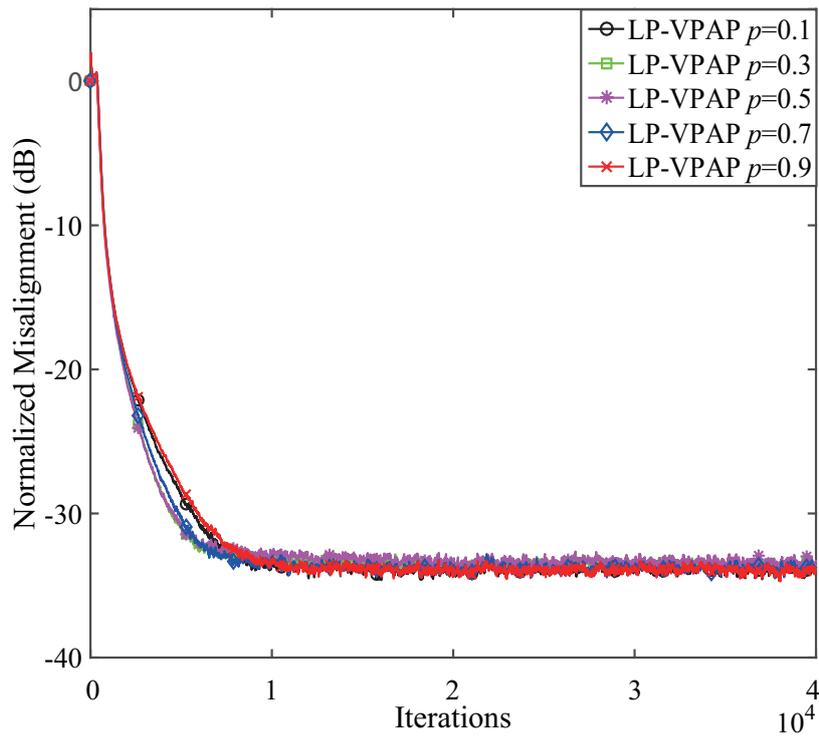


Figure 3. Effects of p on the devised LP-VPAP.

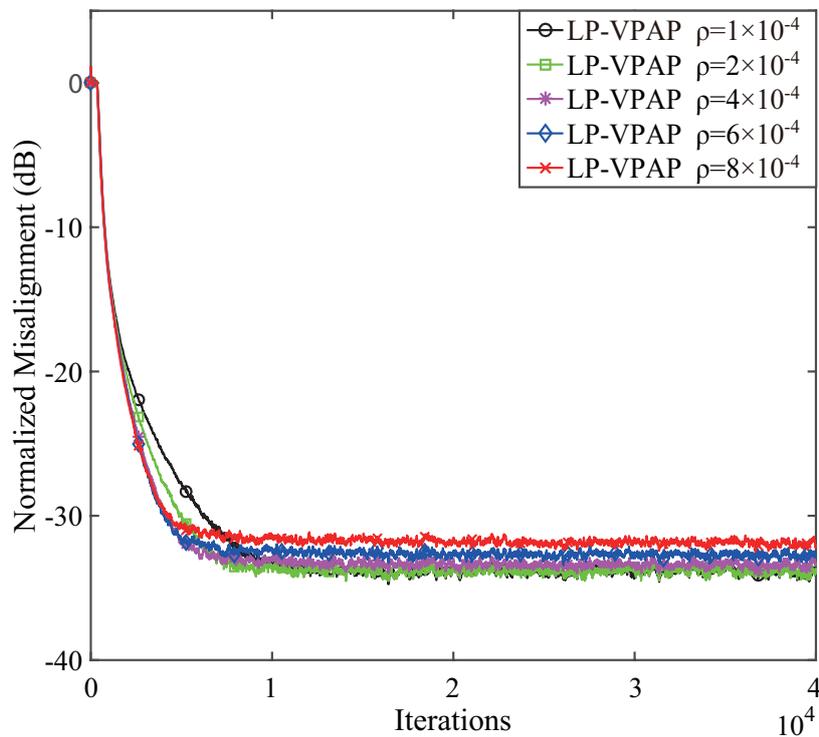


Figure 4. Effects of ρ on the devised LP-VPAP.

4.2. Performance Comparisons of the LP-VPAP Algorithm under Different Input Signals

According to the simulation above, we found that the devised LP-VPAP obtains better performance when $p = 0.5$ and $\rho = 4 \times 10^{-7}$ are selected. Then, the identification behavior of the LP-VPAP is compared with AP, ZA-AP, RZA-AP, PAP, MCC, VKW-MCC, and PAPMCC

algorithms with WGN and CN input. The step-sizes for other algorithms are presented as follows: $\mu_{AP} = \mu_{ZA-AP} = \mu_{RZA-AP} = 0.025$, $\mu_{PAP} = 0.005$, $\mu_{MCC} = 0.0006$, $\mu_{VKW-MCC} = 0.00175$, and $\mu_{PAPMCC} = 0.015$. The performance comparisons of the LP-VPAP with various inputs are presented in Figures 5 and 6, respectively. It is clear to see that the LP-VPAP algorithm achieves the lowest NM and fast convergence rate. When the input is CN, the convergence speed of the MCC and VKW-MCC becomes much slower, while the proposed LP-VPAP is still better than all the related algorithms by considering the convergence and identification error.

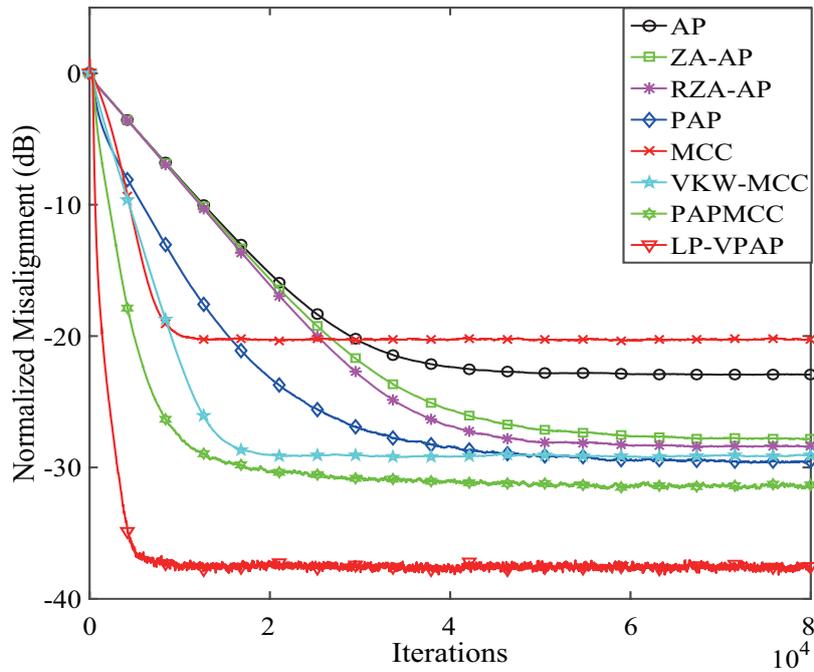


Figure 5. Performance comparisons of the devised LP-VPAP . Input: WGN.

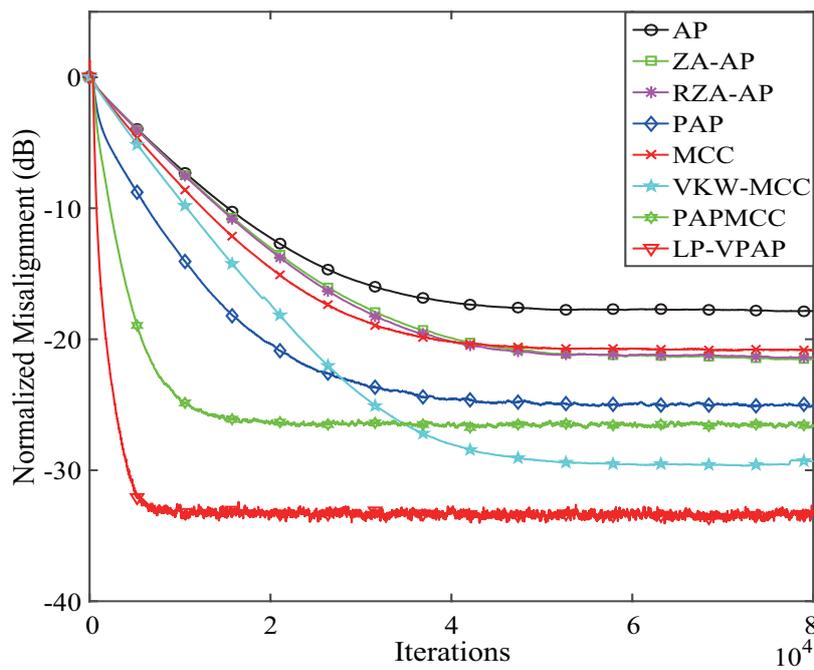


Figure 6. Performance comparisons of the devised LP-VPAP . Input: CN.

4.3. Tracking Behavior of the LP-VPAP

Herein, the tracking behavior of the LP-VPAP is investigated by two different systems. The first system, whose dominant taps distributed in [257,272], is presented in Figure 7a, and the second system, whose dominant taps distributed in [769,784] and [257,272], is presented in Figure 7b. Other parameters are consistent with those in the previous subsection. Simulation results under different input signal are presented in Figures 8 and 9. It is clear to see that the LP-VPAP tracks the two different sparse systems, and converges the fastest and achieves the lowest NM. In the comparisons, only the sparsity-aware algorithms like ZA-AP, RZA-AP, PAMPCC and the proposed LP-VPAP algorithms will lead to estimation error increase. In Figures 8 and 9, we find this result since our proposed LP-VPAP algorithm is also a sparsity-aware algorithm. Thus, when the system is getting less sparse, our proposed LP-VPAP algorithm will result in a high estimation error like the ZA-AP, RZA-AP and PAMPCC algorithms. However, the LP-VPAP algorithm still provides the lowest estimation error.

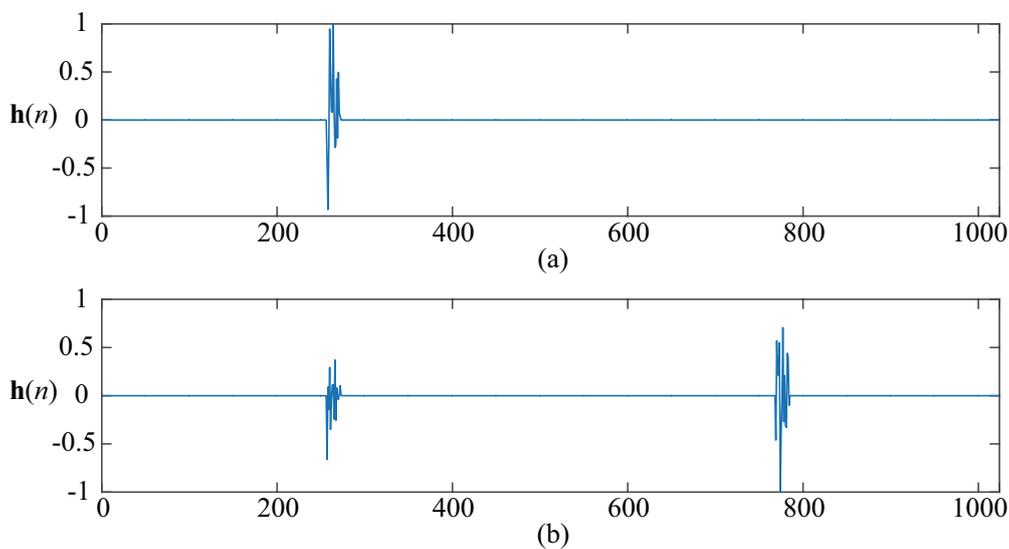


Figure 7. The impulse response used in simulation below. (a) One-cluster system; (b) Two-cluster system.

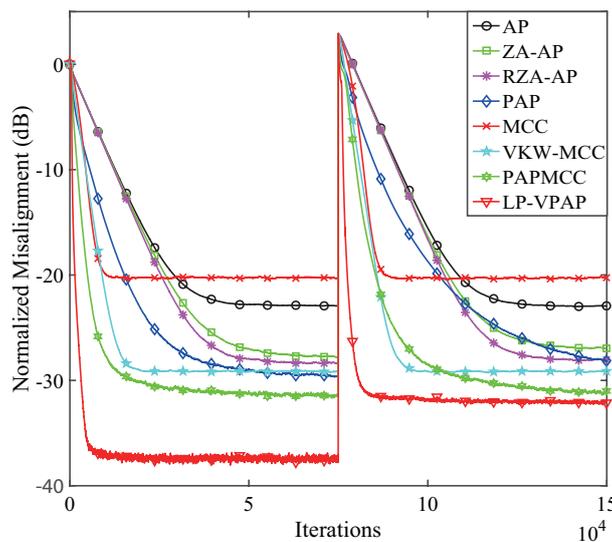


Figure 8. Tracking performance of the devised LP-VPAP. Input: WGN.

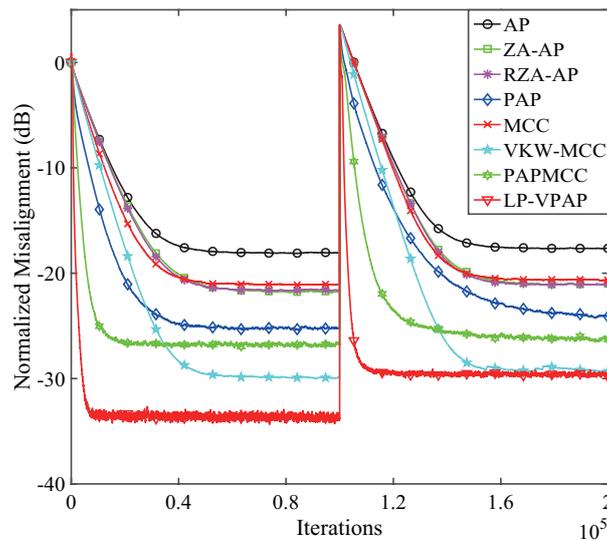


Figure 9. Tracking performance of the devised LP-VPAP. Input: CN.

4.4. Performance Comparisons of the LP-VPAP Algorithm with a Less Sparse Echo Path

In this subsection, the performance of the LP-VPAP is investigated using CN input with a less sparse echo path whose dominant taps are distributed in [257,384]. Herein, $\rho = 5 \times 10^{-8}$ is selected. Other parameters are the same as those in the previous simulations. Simulation results are presented in Figure 10. It is clear to see that although the steady-state error of the devised LP-VPAP increases, it is still better than the related algorithms for providing the fastest convergence speed.

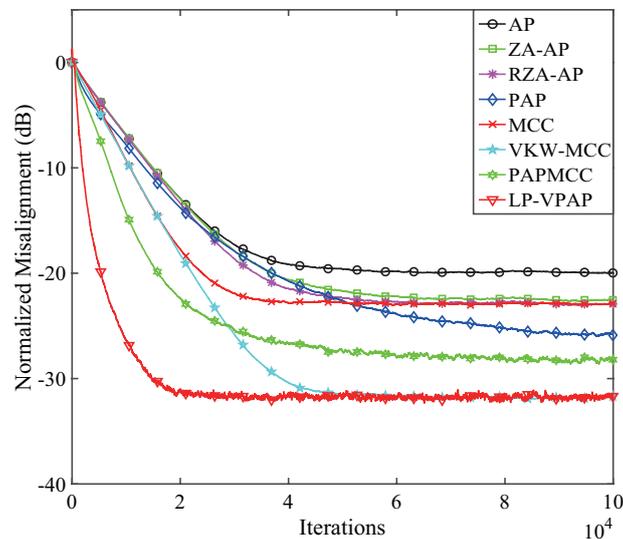


Figure 10. Performance comparisons of the LP-VPAP algorithm with a less sparse echo path.

5. Conclusions

In this paper, a sparsity-aware variable kernel width PAP is devised for sparse SI under impulsive noise environments. The developed LP-VPAP is realized by developing the variable kernel width technique and the l_p -norm-like constraint. The key parameters are analyzed to discuss the performance of the developed LP-VPAP. Simulation results show the LP-VPAP can speed up the convergence and improve the estimation accuracy compared with AP, ZA-AP, RZA-AP, PAP, MCC, VKW-MCC, and PAMCC algorithms.

Author Contributions: Z.J. (Zhengxiong Jiang) wrote the draft, programmed the code and finished the simulations. Y.L. put forward the LP-VPAP algorithm and checked the codes. X.H. and Z.J. (Zhan Jin) provided some useful analysis for LP-VPAP algorithm. All authors approved the contents of this paper.

Funding: This paper is supported by the National Key Research and Development Program of China (2016YFE0111100), Key Research and Development Program of Heilongjiang (GX17A016), the Science and Technology innovative Talents Foundation of Harbin (2016RAXXJ044), the Natural Science Foundation of Beijing (4182077), Natural Science Foundation of Heilongjiang Province, China (F2017004), China Postdoctoral Science Foundation (2017M620918 and 2019T120134), the Fundamental Research Funds for the Central Universities (HEUCFG201829 and 2072019CFG0801), and Opening Fund of Acoustics Science and Technology Laboratory (SSKF2016001).

Conflicts of Interest: There is no conflict of interest.

References

1. Ljung, L. *System Identification: Theory for the User*; Prentice-Hall: Upper Saddle River, NJ, USA, 1987.
2. Benesty, J.; Gänslér, T.; Morgan, D.R.; Sondhi, M.M.; Gay, S.L. *Advances in Network and Acoustic Echo Cancellation*; Springer: Berlin, Germany, 2001.
3. Li, Y.; Wang, Y.; Jiang, T. Sparse-aware set-membership NLMS algorithms and their application for sparse channel estimation and echo cancelation. *AEU-Int. J. Electron. Commun.* **2006**, *70*, 895–902. [[CrossRef](#)]
4. Li, Y.; Wang, Y.; Yang, R.; Ablu, F. A soft parameter function penalized normalized maximum correntropy criterion algorithm for sparse system identification. *Entropy* **2017**, *19*, 1–16. [[CrossRef](#)]
5. Li, Y.; Wang, Y.; Jiang, T. Norm-adaption penalized least mean square/fourth algorithm for sparse channel estimation. *Signal Process.* **2016**, *128*, 243–251. [[CrossRef](#)]
6. Li, Y.; Jiang, Z.; Jin, Z.; Han, X.; Yin, J. Cluster-sparse proportionate NLMS algorithm with the hybrid norm constraint. *IEEE Access* **2018**, *6*, 47794–47803. [[CrossRef](#)]
7. Haykin, S. *Adaptive Filter Theory*, 2nd ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 1991.
8. Widrow, B.; Stearns, S.D. *Adaptive Signal Processing*; Prentice-Hall: Upper Saddle River, NJ, USA, 1985.
9. Douglas, S.C.; Meng, T.Y. Normalized data nonlinearities for LMS adaptation. *IEEE Trans. Signal Process.* **1994**, *42*, 1352–1365. [[CrossRef](#)]
10. Ozeki, K.; Umeda, T. An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. *Electr. Commun. Jpn.* **1984**, *67*, 19–27. [[CrossRef](#)]
11. Shin, H.; Sayed, A.H.; Song, W. Variable step-size NLMS and affine projection algorithms. *Electr. Commun. Jpn.* **2004**, *11*, 132–135. [[CrossRef](#)]
12. Li, Y.; Jiang, Z.; Osman, O.M.O.; Han, X.; Yin, J. Mixed norm constrained sparse APA algorithm for satellite and network echo channel estimation. *IEEE Access* **2018**, *6*, 65901–65908. [[CrossRef](#)]
13. Etter, D. Identification of sparse impulse response systems using an adaptive delay filter. In Proceedings of the 1985 IEEE International Conference on Acoustics, Speech and Signal Processing, Tampa, FL, USA, 26–29 April 1985; pp. 1169–1172.
14. Li, W.; Preisig, J.C. Estimation of rapidly time-varying sparse channels. *IEEE J. Ocean. Eng.* **2007**, *32*, 927–939. [[CrossRef](#)]
15. Li, Y.; Wang, Y.; Jiang, T. Sparse least mean mixed-norm adaptive filtering algorithms for sparse channel estimation applications. *Int. J. Commun. Syst.* **2017**, *30*, 1–14. [[CrossRef](#)]
16. Li, Y.; Zhang, C.; Wang, S. Low-complexity non-uniform penalized affine projection algorithm for sparse system identification. *Circuits Syst. Signal Process.* **2016**, *35*, 1–14. [[CrossRef](#)]
17. Shi, W.; Li, Y.; Zhao, L.; Liu, X. Controllable sparse antenna array for adaptive beamforming. *IEEE Access* **2019**, *7*, 6412–6423. [[CrossRef](#)]
18. Zhang, X.; Jiang, T.; Li, Y.; Zakharov, Y. A novel block sparse reconstruction method for DOA estimation with unknown mutual coupling. *IEEE Commun. Lett.* **2019**. [[CrossRef](#)]
19. Berger, C.R.; Zhou, S.; Preisig, J.C.; Willett, P. Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing. *IEEE Trans. Signal Process.* **2010**, *58*, 1708–1721. [[CrossRef](#)]
20. Gänslér, T.; Gay, S.L.; Sondhi, M.M.; Benesty, J. Double-talk robust fast converging algorithms for network echo cancellation. *IEEE Trans. Speech Audio Process.* **2000**, *8*, 656–663. [[CrossRef](#)]

21. Duttweiler, D.L. Proportionate normalized least-mean-squares adaptation in echo cancelers. *IEEE Trans. Speech Audio Process.* **2000**, *8*, 508–518. [[CrossRef](#)]
22. Gansler, T.; Benesty, J.; Gay, S.L.; Sondhi, M.M. A robust proportionate affine projection algorithm for network echo cancellation. In Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, 5–9 June 2000; pp. 793–796.
23. Benesty, J.; Gay, S.L. An improved PNLMS algorithm. In Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, 13–17 May 2002; pp. 1881–1884.
24. Jin, Z.; Li, Y.; Wang, Y. An enhanced set-membership PNLMS algorithm with a correntropy induced metric constraint for acoustic channel estimation. *Entropy* **2017**, *19*, 1–14. [[CrossRef](#)]
25. Deng, H.; Doroslovački, M. Improving convergence of the PNLMS algorithm for sparse impulse response identification. *IEEE Signal Process. Lett.* **2005**, *12*, 181–184. [[CrossRef](#)]
26. Chen, Y.; Gu, Y.; Hero, A.O. Sparse LMS for system identification. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 3125–3128.
27. Gu, Y.; Jin, J.; Mei, S. L_0 norm constraint LMS for sparse system identification. *IEEE Signal Process. Lett.* **2009**, *16*, 774–777.
28. Su, G.; Jin, J.; Gu, Y.; Wang, J. Performance analysis of l_0 norm constraint least mean square algorithm. *IEEE Trans. Signal Process.* **2012**, *60*, 2223–2235. [[CrossRef](#)]
29. Jiang, S.; Gu, Y. Block-sparsity-induced adaptive filter for multi-clustering system identification. *IEEE Trans. Signal Process.* **2015**, *63*, 5183–5330. [[CrossRef](#)]
30. Jin, J.; Qu, Q.; Gu, Y. Robust zero-point attraction least mean square algorithm on near sparse system identification. *IET Signal Process.* **2013**, *7*, 210–218. [[CrossRef](#)]
31. Meng, R.; de Lamare, R.C.; Nascimento, V.H. Nascimento. Sparsity-aware affine projection adaptive algorithms for system identification. In Proceedings of the Sensor Signal Processing for Defence (SSPD 2011), London, UK, 27–29 September 2011; pp. 793–796.
32. Yu, Y.; Zhao, H.; Chen, B. Sparseness-controlled proportionate affine projection sign algorithms for acoustic echo cancellation. *Circuits Syst. Signal Process.* **2015**, *34*, 3933–3948. [[CrossRef](#)]
33. 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.* **2016**, *10*, 503–510. [[CrossRef](#)]
34. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
35. Vega, L.R.; Rey, H.; Benesty, J.; Tressens, S. A new robust variable step-size NLMS algorithm. *IEEE Trans. Signal Process.* **2008**, *56*, 1878–1893. [[CrossRef](#)]
36. Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Príncipe, J.C. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [[CrossRef](#)]
37. Chen, B.; Wang, J.; Zhao, H.; Zheng, N.; Príncipe, J.C. Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Process. Lett.* **2015**, *22*, 1723–1727. [[CrossRef](#)]
38. 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.-Eng. Appl. Math.* **2015**, *352*, 2708–2727. [[CrossRef](#)]
39. Li, Y.; Jiang, Z.; Shi, W.; Han, X.; Chen, B. Blocked maximum correntropy criterion algorithm for cluster-sparse system identifications. *IEEE Trans. Circuits Syst. II-Express Briefs* **2019**. [[CrossRef](#)]
40. Shi, W.; Li, Y.; Wang, Y. Noise-free maximum correntropy criterion algorithm in non-Gaussian environment. *IEEE Trans. Circuits Syst. II-Express Briefs* **2019**. [[CrossRef](#)]
41. Chen, B.; Xing, L.; Liang, J.; Zheng, N.; Príncipe, J.C. Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion. *IEEE Signal Process. Lett.* **2014**, *21*, 880–884.
42. Chen, B.; Príncipe, J.C. Maximum correntropy estimation is a smoothed MAP estimation. *IEEE Signal Process. Lett.* **2012**, *19*, 491–494. [[CrossRef](#)]
43. Wu, Z.; Shi, J.; Zhang, X.; Ma, W.; Chen, B. Kernel recursive maximum correntropy. *Signal Process.* **2015**, *117*, 11–16. [[CrossRef](#)]
44. Chen, B.; Wang, X.; Lu, N.; Wang, S.; Cao, J.; Qin, J. Mixture correntropy for robust learning. *Pattern Recognit.* **2018**, *79*, 318–327. [[CrossRef](#)]
45. Wang, F.; He, Y.; Wang, S.; Chen, B. Maximum total correntropy adaptive filtering against heavy-tailed noises. *Signal Process.* **2017**, *141*, 84–95. [[CrossRef](#)]

46. Wang, S.; Dang, L.; Chen, B.; Duan, S.; Wang, L.; Chi, K T. Random fourier filters under maximum correntropy criterion. *IEEE Trans. Circuits Syst. I-Regul. Pap.* **2018**, *65*, 3390–3403. [[CrossRef](#)]
47. Qian, G.; Wang, S. Generalized complex correntropy: Application to adaptive filtering of complex data. *IEEE Access* **2018**, *6*, 19113–19120. [[CrossRef](#)]
48. Wang, S.; Dang, L.; Wang, W.; Qian, G.; Chi, K T. Kernel adaptive filters with feedback based on maximum correntropy. *IEEE Access* **2018**, *6*, 10540–10552. [[CrossRef](#)]
49. Huang, F.; Zhang, J.; Zhang, S. Adaptive filtering under a variable kernel width maximum correntropy criterion. *IEEE Trans. Circuits Syst. II-Express Briefs* **2017**, *64*, 1247–1251. [[CrossRef](#)]
50. Jiang, Z.; Li, Y.; Hunag, X. A correntropy-based proportionate affine projection algorithm for estimating sparse channels with impulsive noise. *Entropy* **2019**, *21*, 1–13. [[CrossRef](#)]
51. Das, R.L.; Chakraborty, M. Improving the performance of the PNLMS algorithm using l_1 norm regularization. *IEEE Trans. Audio Speech Lang. Process.* **2016**, *24*, 1280–1290. [[CrossRef](#)]
52. Li, Y.; Li, W.; Yu, W.; Wan, J.; Li, Z. Sparse adaptive channel estimation based on l_p -norm-penalized affine projection algorithm. *Int. J. Antennas Propag.* **2014**, *2014*, 1–8.
53. Benesty, J.; Paleologu, C.; Ciochina, S. On regularization in adaptive filtering. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *19*, 1734–1742. [[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/>).