

Research Article

Kernel Least Logarithmic Absolute Difference Algorithm

Dongliang Fu,¹ Wei Gao ,² Wentao Shi,³ and Qunfei Zhang³

¹Shanghai Marine Equipment Research Institute, Shanghai 200031, China

²School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China

³School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Wei Gao; wei_gao@ujs.edu.cn

Received 2 July 2021; Revised 8 December 2021; Accepted 31 January 2022; Published 8 March 2022

Academic Editor: Akif Akgul

Copyright © 2022 Dongliang Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kernel adaptive filtering (KAF) algorithms derived from the second moment of error criterion perform very well in nonlinear system identification under assumption of the Gaussian observation noise; however, they inevitably suffer from severe performance degradation in the presence of non-Gaussian impulsive noise and interference. To resolve this dilemma, we propose a novel robust kernel least logarithmic absolute difference (KLLAD) algorithm based on logarithmic error cost function in reproducing kernel Hilbert spaces, taking into account of the non-Gaussian impulsive noise. The KLLAD algorithm shows considerable improvement over the existing KAF algorithms without restraining impulsive interference in terms of robustness and convergence speed. Moreover, the convergence condition of KLLAD algorithm with Gaussian kernel and fixed dictionary is presented in the mean sense. The superior performance of KLLAD algorithm is confirmed by the simulation results.

1. Introduction

Kernel adaptive filters as a tremendous breakthrough of the conventional linear adaptive filters have been widely used in many practical nonlinear applications including time series prediction [1], acoustic echo cancellation [2], channel equalization [3], abnormal event detection [4], etc. The scheme of kernel adaptive filtering (KAF) is to map the original input data into high or infinite dimensional feature space via kernel function and then apply the framework of typical linear adaptive filtering to the transformed data in the reproducing kernel Hilbert spaces (RKHS) leading to various KAF algorithms [5–8]. The kernel least-mean-square (KLMS) algorithm, as the benchmark among of KAF algorithms, is developed from the cost function of second-order statistic of the error between the desired signal and instantaneous estimate under the assumption of Gaussian noise for its mathematical simplicity and convenience [9]. Therefore, it can be ensured that the performances of KLMS-type algorithms only for Gaussian disturbance noise severely degrade, when the desired signals are corrupted by the impulsive interferences. In practical applications, e.g., underwater acoustic signal processing [10], wireless

communication environments [11], and radar cluster elimination [12], the impulsive noises with the statistical characteristics consisting of infrequency, short duration, and high amplitude are more rational than the ideal Gaussian noises.

For conventional linear robust adaptive filtering, the signed adaptive filters and their theoretical analyses have been extensively studied in [13–15]. The generalized maximum correntropy criterion (GMCC) algorithm with generalized Gaussian density function was proposed in [16], and its stability and steady-state mean square performance were also investigated. The authors of [17] proposed the least logarithmic absolute difference (LLAD) with the correntropy-induced metric (CIM) constraint in order to exploit system sparsity and suppress the interferences. The minimum kernel risk-sensitive loss (MKRSL) algorithm was developed to achieve fast convergence speed by applying risk-sensitive loss while yielding the robust performance to outliers [18]. When the shape parameter is equal to 2 and the risk-sensitive parameter gradually tends to trivial, both GMCC and MKRSL algorithms reduce to the ordinary maximum correntropy criterion (MCC) algorithm. The constrained least mean logarithmic square (CLMLS) based

on a relative logarithmic cost function and its variants were proposed in [19], and they were used in the application of sparse sensor array synthesis achieving the desired beam pattern with much less sensor elements. In [20], a robust least mean logarithmic square (RLMLS) algorithm and its variable step-size variant were presented to combat impulsive noises, and its theoretical mean square performance was also analyzed with the stationary white Gaussian inputs.

Meanwhile, the topic of nonlinear system identification corrupted by the impulsive noise also attracted large numbers of research interests [21]. It is thus important and necessary to study the robust KAF algorithm in order to resolve the divergence problem of performance caused by non-Gaussian impulsive noise. In [22], the kernel maximum correntropy criterion (KMCC) algorithm was developed by introducing the framework of MCC into RKHS in the presence of impulsive noises. By combining the kernel method and the affine projection sign algorithm, the kernel affine projection sign algorithm (KAPSA) was proposed in [23] to combat the non-Gaussian impulse interference. An improved variable forgetting factor recursive logarithmic least mean p th (IVFF-RLLMP) algorithm was developed in [24] for the Volterra system identification against the impulsive interference modeled by α -stable distribution. As a counterpart of GMCC algorithm, the generalized kernel maximum correntropy (GKMC) and the quantized GKMC (QGKMC) algorithm were developed in [25] for robust nonlinear adaptive filtering. In [26], the quantized minimum kernel risk-sensitive loss (QMKRSL) algorithm was proposed to achieve better and robust performance of nonlinear filtering for outliers. Motivated by the studies in [27, 28] on the Cauchy loss which has been successfully used in various robust learning applications, the multikernel minimum Cauchy kernel loss (MKMCKL) algorithm was reported in [29] showing the improved nonlinear filtering performance over counterpart single algorithm in the presence of extreme outliers. Recently, the kernel affine projection-like (KAPL) algorithm in RKHS was proposed and investigated for nonlinear channel equalization in scenarios of non-Gaussian noises [30]. The kernel least mean p -power (KLMP) algorithm was proposed to alleviate the adverse impact of impulsive noise in [31, 32], independently. More recently, Nyström kernel recursive generalized maximum correntropy (NKRGMCC) with probability density rank-based quantization sampling algorithm was proposed to improve the convergence performance for impulsive noises [33]. The tracking analysis of kernel signed error algorithm (KSEA) with Gaussian kernel for time-variant nonlinear system was analyzed thoroughly under the assumption of impulsive noise [34]. More recently, the logarithmic hyperbolic cosine-based adaptive filter (LHCAF) was proposed in [35] to address the issue of instability of its prototype algorithm, and the transient and steady-state analyses were also provided. Subsequently, the authors of [36] proposed the multiple random Fourier features Cauchy-loss conjugate gradient (MRFCCG) algorithm which has better performance than the classical KAF algorithms in terms of computational complexity and filtering accuracy.

Therefore, the cost functions adopting the frameworks of fractional order statistics of error or the distinct types of error measures are able to provide effective ways to reveal the robust performance against impulsive noises. Inspired by the family of linear adaptive filtering algorithms based on the logarithmic cost proposed in [37], our purpose is to extend this scheme into RKHS to obtain the robustness of KAF algorithm particularly within non-Gaussian impulsive noise environment. In this paper, the kernel least logarithmic absolute difference algorithm based on the logarithmic error cost framework is proposed to achieve the nonlinear system identification in the impulsive interference environments, which are more frequently encountered in practical applications. Simulation results illustrate the proposed KLLAD algorithm can consistently decrease the drastic perturbation of recursive weight coefficients caused by the large amplitude of instantaneous estimation error with low probability.

Notation: We use normal font small letters x for scalar variables, boldface small letters \mathbf{x} for column vectors, and boldface capital letters \mathbf{X} for matrices. The superscript $(\cdot)^T$ represents the transpose of a vector or a matrix. The expectation is denoted by $\mathbb{E}\{\cdot\}$, and matrix trace is denoted by $\text{tr}\{\cdot\}$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. The notation $\|\cdot\|_2$ is the ℓ_2 -norm of its matrix or vector argument; $\|\mathbf{x}\|_A^2$ is the weighted square value $\mathbf{x}^T \mathbf{A} \mathbf{x}$. Notation $\text{sgn}\{\cdot\}$ is the signum function. The operator $\text{eig}_{\max}\{\mathbf{X}\}$ denotes the maximum eigenvalue of matrix \mathbf{X} . Identity matrix of size $N \times N$ is denoted by \mathbf{I}_N .

2. Preliminaries of KAF Algorithms

Let \mathcal{H} denote a Hilbert space of real-valued function ψ from a compact subspace $\mathcal{U} \subset \mathbb{R}^L$. Function $\kappa: \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}$ is a reproducing kernel, and $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ is the induced RKHS with its inner product. The following unknown nonlinear system is considered:

$$y_n = f^*(\mathbf{x}_n) + z_n. \quad (1)$$

The scalar y_n and the vector $\mathbf{x}_n \in \mathbb{R}^L$ denote the desired output and the input signal, respectively. Here, f^* represents the optimum functional to be identified, and z_n stands for the non-Gaussian impulsive noise modeled as the contaminated-Gaussian (CG) noise as follows [13, 38]:

$$z_n = v_n + b_n \epsilon_n, \quad (2)$$

where both v_n and ϵ_n are statistically independent zero-mean white Gaussian noises with the variances σ_v^2 and $\sigma_\epsilon^2 = K\sigma_v^2$, with parameter $1 \ll K$. Moreover, the random sequence b_n is from a Bernoulli random process with the probability of $\Pr(b_n = 1) = p_r$ and $\Pr(b_n = 0) = 1 - p_r$. Notice that v_n and $b_n \epsilon_n$ represent the common Gaussian noise and particular impulsive interference component, respectively. Subsequently, the probability density function (pdf) of the impulsive CG noise z_n is given by

$$p_z = (1 - p_r) \mathcal{N}(0, \sigma_v^2) + p_r \mathcal{N}(0, (K + 1)\sigma_v^2). \quad (3)$$

When $p_r = 0$, the impulsive CG noise z_n deteriorates to a common white Gaussian noise with zero-mean and variance

σ_v^2 , namely, impulsive interference free. Correspondingly, the variance of impulsive noise z_n is given by

$$\sigma_z^2 = \mathbb{E}\{z_n^2\} = (1 - p_r)\sigma_v^2 + p_r(K + 1)\sigma_v^2. \quad (4)$$

Hence, z_n is statistically independent of any other signals. The reason why we adopt the CG model is that it makes the analysis of mean stability of KLLAD algorithm mathematically tractable.

Given a sample set of pairs of input vectors and desired output scalars, i.e., $\{\mathbf{x}_n, y_n\}_{n=1}^N$, we aim at estimating a nonlinear regression function ψ that relates input data \mathbf{x}_n and output data y_n corrupted by impulsive noise z_n . Let \mathcal{H} be a RKHS with kernel κ . By virtue of the representer theorem [39], the function ψ in \mathcal{H} that minimizes the regularized least-squares problem:

$$\min_{\psi \in \mathcal{H}} \sum_{n=1}^N |y_n - \psi(\mathbf{x}_n)|^2 + \varepsilon \|\psi\|_{\mathcal{H}}^2, \quad (5)$$

with ε a nonnegative regularization constant, can be written as the kernel expansion in terms of available training data in form of

$$\psi = \sum_{n=1}^N w_n \kappa(\cdot, \mathbf{x}_n). \quad (6)$$

When $\varepsilon = 0$, it will not affect the derivation of algorithm. It can be seen from (5) and (6) that the functional representation is formulated as the parametric vector form. However, solution (6) is infeasible to be performed in an online manner, because the algorithm can not cope with the linear increase of the size N of the model as the latest input data available. A commonly used strategy is to adopt a finite-order model of the form [5]

$$\bar{\psi} = \sum_{m=1}^{M_n} w_m \kappa(\cdot, \mathbf{x}_{\omega, m}). \quad (7)$$

The set of selected input data $\mathcal{D}_n = \{\mathbf{x}_{\omega, m}\}_{m=1}^{M_n}$ is so-called online dictionary with the length M_n and is generated by the informative criteria from input vectors \mathbf{x}_n in an online manner, e.g., coherence criterion [1], surprise criterion [5], and approximate linear dependency [40]. Note that (6) seems to be the same as (7) in form, whereas the length M_n determined by the online sparsification criterion is analogue to the order of transversal filter. Without loss of generality, we chose the coherence criterion which allows the candidate \mathbf{x}_n to be inserted into the dictionary if its maximum coherence remains below the given threshold δ_κ , namely,

$$\max_{m=1,2,\dots,M_n} |\kappa(\mathbf{x}_n, \mathbf{x}_{\omega, m})| \leq \delta_\kappa, \quad (8)$$

where $0 < \delta_\kappa \leq 1$ determines both the level of sparsity and the coherence of the dictionary. Based on the stochastic gradient of cost functions $J(e_n)$ with respect to \mathbf{w} , using the parametric finite-order model (7), we can then obtain the corresponding recursive update equation of the KAF algorithms:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta \frac{\partial J(e_n)}{\partial \mathbf{w}} \mathbf{\kappa}_{\omega, n}. \quad (9)$$

With the positive step-size η and the weight coefficients vector $\mathbf{w}_n = [w_n(1), w_n(2), \dots, w_n(M_n)]^\top$, the instantaneous estimation error e_n is given by

$$e_n = y_n - \mathbf{w}_n^\top \mathbf{\kappa}_{\omega, n}. \quad (10)$$

Moreover, the kernelized input vector $\mathbf{\kappa}_{\omega, n}$ is defined by

$$\mathbf{\kappa}_{\omega, n} = [\kappa(\mathbf{x}_n, \mathbf{x}_{\omega, 1}), \kappa(\mathbf{x}_n, \mathbf{x}_{\omega, 2}), \dots, \kappa(\mathbf{x}_n, \mathbf{x}_{\omega, M_n})]^\top. \quad (11)$$

Adopting the mean-squared error (MSE) criterion for (9) leads to the scheme of KLMS algorithm described as follows:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta e_n \mathbf{\kappa}_{\omega, n}, \quad (12)$$

where the complete KLMS algorithm with online sparsification criterion is not present for clarity.

It can be easily observed from (12) that the KLMS-type algorithms suffer from severe performance degradation and even divergence caused by the instantaneous estimation error given in (10), which is contaminated by the non-Gaussian impulsive noise defined in (2). It is thus critical for the design of robust KAF algorithm to effectively suppress the adverse effects of impulsive interference noise.

3. KLLAD Algorithm

In this section, we shall derive the KLLAD algorithm based on the logarithmic cost function.

Logarithmic cost as one type of relative cost measures is capable of providing relatively legitimate amplification for the ordinary error fluctuation and significant attenuation for very large error value induced by impulsive interferences. It has been illustrated in [37] that the logarithmic cost function can proportionally adjust the weight coefficients for small and large error values depending on the combination weights varying with time. Therefore, we introduce the differentiable combined logarithmic error cost function as follows:

$$J(e_n) = F(e_n) - \frac{1}{\lambda} \ln(1 + \lambda F(e_n)), \quad (13)$$

where λ is a design parameter, and $F(e_n)$ is the conventional cost function of instantaneous estimation error. Note that the universal adaptive filtering algorithm obtained from (13) is able to update the coefficients by small error and suppress the drastic perturbation of error simultaneously.

Taking the derivative of (13) with respect to weight vector \mathbf{w} , we obtain

$$\nabla_{\mathbf{w}} J(e_n) = \nabla_{\mathbf{w}} F(e_n) \frac{\lambda F(e_n)}{1 + \lambda F(e_n)}. \quad (14)$$

Since the impulsive CG noise exists, the kernel adaptive filters based on quadratic cost function have to face the severe performance degradation. Nevertheless, the signed adaptive filtering algorithms based on the ℓ_1 -norm of

estimation error are robust to the impulsive noise because of its recursive update equation only utilizing the sign of instantaneous estimation error. In order to mitigate the adverse impact of impulsive noise on KAF, let the conventional cost function of error $F(e_n)$ be $\mathbb{E}\{|e_n|\}$ using (10), namely,

$$F(e_n) = \mathbb{E}\{|y_n - \mathbf{w}_n^\top \boldsymbol{\kappa}_{\omega,n}|\}. \quad (15)$$

It should be pointed out that cost function (13) using (15) can perform less attenuation for the small estimation errors and more attenuation for relatively large estimation errors simultaneously. Substituting (15) into (14) leads to

$$\frac{\partial J(e_n)}{\partial \mathbf{w}} = \frac{\partial F(e_n)}{\partial \mathbf{w}} \frac{\lambda F(e_n)}{1 + \lambda F(e_n)}. \quad (16)$$

Applying the steepest-descent method to minimize the logarithmic cost function (13), thus the weights vector of KLLAD algorithm can be solved iteratively by

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta \left[-\frac{\partial J(e_n)}{\partial \mathbf{w}} \right]. \quad (17)$$

By (16), the recursive update equation of KLLAD algorithm can be reformulated as

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \eta \frac{\partial F(e_n)}{\partial \mathbf{w}} \frac{\lambda F(e_n)}{1 + \lambda F(e_n)}, \quad (18)$$

where the subgradient in the second term on the right hand side of (18) is calculated as

$$\frac{\partial F(e_n)}{\partial \mathbf{w}} \approx \text{sgn}(e_n). \quad (19)$$

Note that the above approximation notation implies that the subgradient of conventional cost function $F(e_n)$ is replaced by its instantaneous estimate. Substituting (15) and (19) into (18) and then removing the expectations, the stochastic update equation of KLLAD algorithm can be rewritten as

$$\begin{aligned} \mathbf{w}_{n+1} &= \mathbf{w}_n + \eta \frac{\lambda |e_n|}{1 + \lambda |e_n|} \text{sgn}(e_n) \boldsymbol{\kappa}_{\omega,n} \\ &= \mathbf{w}_n + \eta \frac{\lambda e_n}{1 + \lambda |e_n|} \boldsymbol{\kappa}_{\omega,n}, \end{aligned} \quad (20)$$

where $\boldsymbol{\kappa}_{\omega,n}$ denotes the $M_n \times 1$ dimensional kernelized input vector as in (11). For universal nonlinear filtering performance, we only focus on using the Gaussian kernel function $\kappa(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|_2^2 / 2\xi^2)$, with $\xi > 0$ the kernel bandwidth in this paper. Moreover, the Gaussian kernel function has been successfully used in the theoretical analysis of KAF algorithms due to its mathematical convenience and tractability in the derivation [34, 41]. Based on (20), the recursive update equations of KLLAD algorithm based on the finite-order model using the coherence criterion are presented in the following.

At each time instant n , the input \mathbf{x}_n will be decided into the case of rejection or reception according to the coherence criterion (8) for online dictionary \mathcal{D}_n as follows:

(i) Rejection: if $\max_{m=1,2,\dots,M_n} |\kappa(\mathbf{x}_n, \mathbf{x}_{\omega,m})| > \delta_\kappa$

$$\begin{aligned} \mathcal{D}_{n+1} &= \mathcal{D}_n, M_{n+1} = M_n, \\ \mathbf{w}_{n+1} &= \mathbf{w}_n + \eta \frac{\lambda e_n}{1 + \lambda |e_n|} \boldsymbol{\kappa}_{\omega,n}. \end{aligned} \quad (21)$$

Note that the dimensions of kernelized input vector $\boldsymbol{\kappa}_{\omega,n}$ and weight vector \mathbf{w}_n remain still in this case.

(ii) Reception: if $\max_{m=1,2,\dots,M_n} |\kappa(\mathbf{x}_n, \mathbf{x}_{\omega,m})| \leq \delta_\kappa$

$$\begin{aligned} \mathcal{D}_{n+1} &= \mathcal{D}_n \cup \{\mathbf{x}_n\}, M_{n+1} = M_n + 1, \\ \mathbf{w}_{n+1} &= \begin{bmatrix} \mathbf{w}_n \\ 0 \end{bmatrix} + \eta \frac{\lambda e_n}{1 + \lambda |e_n|} \boldsymbol{\kappa}_{\omega,n}. \end{aligned} \quad (22)$$

Note that the dimensions of kernelized input vector $\boldsymbol{\kappa}_{\omega,n}$ and weight vector \mathbf{w}_n are augmented to make the update equation of weight vector valid.

It should be pointed that the threshold δ_κ is selected by grid search over the interval $[0, 1)$ to determine the sparsity of online dictionary \mathcal{D}_n as in [1, 5–7].

The KLLAD algorithm described by (21) and (22) combines the advantages of KLMS and KSEA algorithms leading to the improved convergence performance taking into account Gaussian noise and non-Gaussian impulsive noise existing simultaneously. When the design parameter λ is set to large value, the KLLAD algorithm can achieve the robust convergence performance against the impulsive noise. In addition, the KLLAD algorithm has faster convergence rate than the KSEA algorithm especially for highly correlated input signals. The scheme of KLLAD algorithm is summarized as Algorithm 1.

By (9), the recursive update equation of KAF algorithms can be expressed in a general form of

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta f(e_n) e_n \boldsymbol{\kappa}_{\omega,n}, \quad (23)$$

where $f(e_n)$ is defined as the nonlinear error function which performs like the generalized variable step-size against non-Gaussian impulsive noise. According to (23), various nonlinear KAF algorithms can be readily obtained by solving distinct cost functions in the RKHS. Hence, the error functions of some KAF algorithms mentioned previously are listed in Table 1.

It should be pointed out that the error function of MKRSL algorithm reduces to that of KMCC algorithm as the risk-sensitive parameter $\lambda_f \rightarrow 0_+$ or that of KLMS algorithm as the kernel bandwidth $\xi_f \rightarrow \infty$ [18]. This means that the KMCC algorithm is a particular case of MKRSL algorithm in fact. Thus, we are only concerned with the KMCC algorithm with the characteristic of robustness for impulsive noises in the following. Correspondingly, the curves of error functions listed in Table 1 with different parameters are plotted in Figure 1. First of all, it can be observed from Figure 1(a) that the value of error function of KLMS algorithm is constant one without any ability of suppressing the impulsive noise due to its quadratic cost

- (1) **Input:** $\{\mathbf{x}_n, y_n\}$, $n = 1, 2, \dots$
- (2) **Initialization:** select the step-size $\eta > 0$, the Gaussian kernel bandwidth $\xi > 0$, the threshold $\delta_\kappa > 0$, the parameter $\lambda > 0$, the dictionary $\mathcal{D}_0 = \{\mathbf{x}_1\}$, and weight coefficient $w_0(1) = 0$.
- (3) **for** $n = 1, 2, \dots$, **do**
- (4) **if** $\max_{m=1,2,\dots,M_n} |\kappa(\mathbf{x}_n, \mathbf{x}_{\omega,m})| > \delta_\kappa$ Update \mathbf{w}_{n+1} via (21);
- (5) **elseif** $\max_{m=1,2,\dots,M_n} |\kappa(\mathbf{x}_n, \mathbf{x}_{\omega,m})| \leq \delta_\kappa$ Update \mathbf{w}_{n+1} via (22);
- (6) **end if** Obtain solution $\bar{\psi}(\mathbf{x}_n) = \sum_{m=1}^{M_n} w_n(m) \kappa(\mathbf{x}_n, \mathbf{x}_{\omega,m})$.
- (7) **end for**

ALGORITHM 1: KLLAD algorithm.

TABLE 1: Error functions of several KAF algorithms.

Algorithm	Error function $f(e_n)$
KLMS	$f(e_n) = 1$
KMCC	$f(e_n) = \kappa(y_n, \phi(\mathbf{x}_n)) = \exp(-\ e_n\ _2^2 / 2\xi_f^2)$
KLMP	$f(e_n) = e_n ^{p-2}$
KLLAD	$f(e_n) = \lambda/1 + \lambda e_n $

function for Gaussian noise. We can then see from Figures 1(b)–1(d) that the output values of nonlinear error functions of KMCC, KLMP, and KLLAD algorithms are effectively attenuated even for the very large instantaneous estimation errors caused by impulsive interferences. Specifically, the attenuation rate of error function of KMCC with small kernel bandwidth ξ_f is very fast, whereas its maximum output values corresponding to the relatively small error inputs particularly near zero are always less than 1. Consequently, the KMCC algorithm has to take much larger step-size η than the KLMP and KLLAD algorithms. Although both error functions have similar shape of output values for KLMP algorithm with small p and KLLAD algorithm with large λ , the error function of KLLAD consistently gives much smaller outputs for larger error inputs than that of KLMP.

4. Convergence Condition on Mean Stability of the KLLAD Algorithm

In this section, we shall investigate the sufficient convergence condition of KLLAD algorithm via Gaussian kernel function with the fixed dictionary in the mean sense.

For mathematical tractability of theoretical analysis, the CG modeled impulsive noise z_n given in (1), instead of the standard symmetric α -stable distribution, has been successfully used in the theoretical analysis of robust adaptive filters [13, 14, 34].

One the one hand, the coherence criterion guarantees the length M_n is infinite [1]. On the other hand, it is true that the length M_n of online dictionary \mathcal{D}_n gradually tends to be invariant in the steady-state phase. Therefore, we only need to consider the recursive update equation (21) of KLLAD algorithm with the steady fixed dictionary \mathcal{D} with constant length M in the context of the derivation of convergence condition on mean stability.

We start with introducing the weight error vector defined by

$$\mathbf{v}_n = \mathbf{w}_n - \mathbf{w}^*, \quad (24)$$

where $\mathbf{w}^* = [w_1^*, w_2^*, \dots, w_M^*]^\top$ represents the optimal weight vector of vector-valued form of nonlinear system. It should be emphasized that \mathbf{w}^* corresponds to the functional f^* based on the specific dictionary elements. From (1) and (24), the instantaneous estimation error (10) can be rewritten as

$$e_n = z_n - \mathbf{v}_n^\top \bar{\mathbf{k}}_{\omega,n}. \quad (25)$$

With the kernelized input vector with fixed dictionary \mathcal{D}

$$\bar{\mathbf{k}}_{\omega,n} = [\kappa(\mathbf{x}_n, \mathbf{x}_{\omega,1}), \kappa(\mathbf{x}_n, \mathbf{x}_{\omega,2}), \dots, \kappa(\mathbf{x}_n, \mathbf{x}_{\omega,M})]^\top. \quad (26)$$

Subtracting \mathbf{w}^* from both sides of (21), and using (25), we obtain

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \eta\lambda \frac{z_n - \mathbf{v}_n^\top \bar{\mathbf{k}}_{\omega,n}}{1 + \lambda|e_n|} \bar{\mathbf{k}}_{\omega,n}. \quad (27)$$

Taking the expected values of both sides of (27), it follows that

$$\begin{aligned} \mathbb{E}\{\mathbf{v}_{n+1}\} &= \mathbb{E}\{\mathbf{v}_n\} + \eta\lambda \mathbb{E}\left\{ \frac{z_n \bar{\mathbf{k}}_{\omega,n}}{1 + \lambda|e_n|} \right\} \\ &\quad - \eta\lambda \mathbb{E}\left\{ \frac{\bar{\mathbf{k}}_{\omega,n} \bar{\mathbf{k}}_{\omega,n}^\top}{1 + \lambda|e_n|} \mathbf{v}_n \right\}. \end{aligned} \quad (28)$$

Since the impulsive CG noise z_n with zero-mean is assumed to be statistically independent of any other signals as mentioned above, and $\bar{\mathbf{k}}_{\omega,n} \bar{\mathbf{k}}_{\omega,n}^\top$ is assumed to be statistically independent of weight error vector \mathbf{v}_n , i.e., the modified independence assumption (MIA) widely used in [34, 41, 42], then (28) can be reformulated as

$$\mathbb{E}\{\mathbf{v}_{n+1}\} = \left(\mathbf{I}_M - \eta\lambda \mathbb{E}\left\{ \frac{\bar{\mathbf{k}}_{\omega,n} \bar{\mathbf{k}}_{\omega,n}^\top}{1 + \lambda|e_n|} \right\} \right) \mathbb{E}\{\mathbf{v}_n\}. \quad (29)$$

For theoretical analysis tractability, the second term of (29) can be approximated as

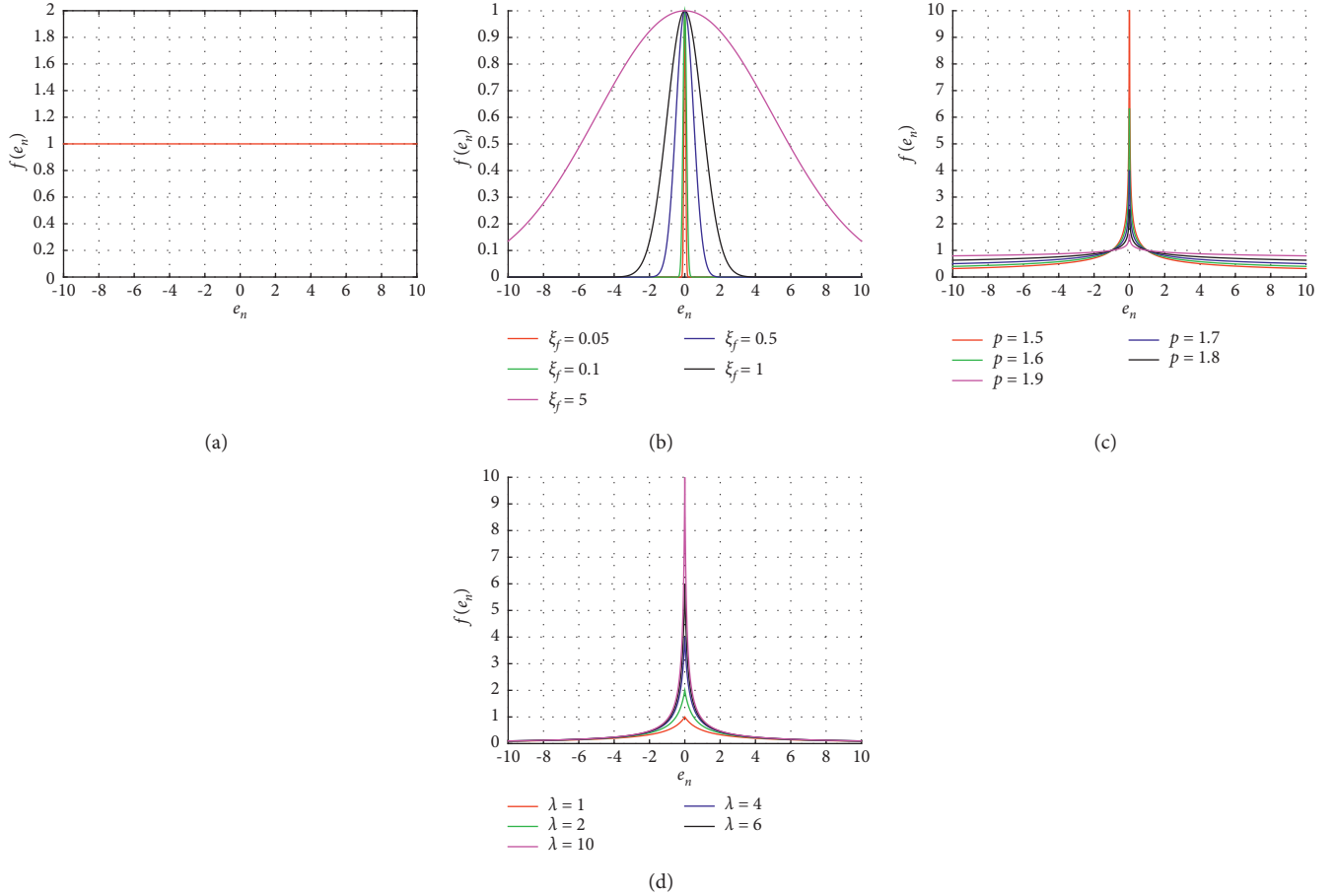


FIGURE 1: Curves of error functions $f(e_n)$ with $e_n \in [-10, 10]$. (a) KLMs algorithm. (b) KMCC algorithm. (c) KLMP algorithm. (d) KLLAD algorithm.

$$\mathbb{E}\left\{\frac{\tilde{\mathbf{k}}_{\omega,n}\tilde{\mathbf{k}}_{\omega,n}^T}{1+\lambda|e_n|}\right\} \approx \frac{\mathbb{E}\{\tilde{\mathbf{k}}_{\omega,n}\tilde{\mathbf{k}}_{\omega,n}^T\}}{1+\lambda\mathbb{E}\{|e_n|\}}. \quad (30)$$

On the one hand, the numerator of (30) is the autocorrelation matrix of the kernelized input vector defined by

$$\mathbf{R}_{\kappa\kappa} = \mathbb{E}\{\tilde{\mathbf{k}}_{\omega,n}\tilde{\mathbf{k}}_{\omega,n}^T\} \in \mathbb{R}^{M \times M}. \quad (31)$$

It has been determined in the theoretical analysis of KAF algorithms [34, 41]. The (i, j) -th element of matrix $\mathbf{R}_{\kappa\kappa}$ can be computed by

$$\begin{aligned} [\mathbf{R}_{\kappa\kappa}]_{ij} &= \left| \mathbf{I}_L + \frac{2}{\xi^2} \mathbf{R}_{xx} \right|^{-1/2} \\ &\times \exp\left(-\frac{1}{4\xi^2} \left[2\left(\|\mathbf{x}_{\omega,i}\|_2^2 + \|\mathbf{x}_{\omega,j}\|_2^2\right) \right. \right. \\ &\left. \left. - \|\mathbf{x}_{\omega,i} + \mathbf{x}_{\omega,j}\|_2^2 \left(\mathbf{I}_L + \xi^2 \mathbf{R}_{xx}^{-1/2} \right)^{-1} \right] \right), \end{aligned} \quad (32)$$

with the input covariance matrix $\mathbf{R}_{xx} = \mathbb{E}\{\mathbf{x}_n \mathbf{x}_n^T\} \in \mathbb{R}^{L \times L}$. On the other hand, using (25) and the MIA assumption, the expression of MSE for KLLAD algorithm is given by

$$\mathbb{E}\{e_n^2\} \approx \sigma_z^2 + \text{tr}\{\mathbf{R}_{\kappa\kappa} \mathbf{V}_n\}, \quad (33)$$

with the autocorrelation matrix of weight error vector $\mathbf{V}_n = \mathbb{E}\{\mathbf{v}_n \mathbf{v}_n^T\} \in \mathbb{R}^{M \times M}$. Although the recursion of \mathbf{V}_n is not explicitly provided, it has trivial influence on the convergence condition on the mean stability of KLLAD algorithm as shown below. By the result of (4), (33) can be expressed as

$$\mathbb{E}\{e_n^2\} = (1-p_r)\sigma_v^2 + p_r(K+1)\sigma_v^2 + \text{tr}\{\mathbf{R}_{\kappa\kappa} \mathbf{V}_n\}. \quad (34)$$

The expectation in denominator of (30) can be roughly approximated as

$$\mathbb{E}\{|e_n|\} \approx \sqrt{(1-p_r)\sigma_v^2 + p_r(K+1)\sigma_v^2 + \text{tr}\{\mathbf{R}_{\kappa\kappa} \mathbf{V}_n\}}. \quad (35)$$

$$\mathbb{E}\{\mathbf{v}_{n+1}\} = \left(\mathbf{I}_M - \eta\lambda \frac{\mathbf{R}_{\kappa\kappa}}{1 + \lambda\sqrt{(1-p_r)\sigma_v^2 + p_r(K+1)\sigma_v^2 + \text{tr}\{\mathbf{R}_{\kappa\kappa}\mathbf{V}_n\}}} \right) \mathbb{E}\{\mathbf{v}_n\}. \quad (36)$$

Then, (29) can be further determined by (36), as shown at the top of next page. By (36), thus the sufficient convergence condition on mean stability is given by

$$0 < \eta < \frac{2 + 2\lambda\sqrt{(1-p_r)\sigma_v^2 + p_r(K+1)\sigma_v^2 + \text{tr}\{\mathbf{R}_{\kappa\kappa}\mathbf{V}_n\}}}{\text{eig}_{\max}\{\mathbf{R}_{\kappa\kappa}\}}. \quad (37)$$

Thus, (37) can be rigorously reformulated as

$$0 < \eta < \frac{2 + 2\lambda\sigma_v\sqrt{1 + p_rK}}{\text{eig}_{\max}\{\mathbf{R}_{\kappa\kappa}\}}, \quad (38)$$

where the calculation of autocorrelation matrix of weight error vector \mathbf{V}_n is not used as explained before.

5. Simulation Results

In this section, we evaluated the performance of the proposed KLLAD algorithm in the context of impulsive noise by the simulation results. All the curves are obtained by averaging over 200 independent Monte Carlo trails.

5.1. Example I. In general, the optimal weight vectors of KAF algorithms, which correspond to the elements of online dictionary built by an online manner, are unavailable a priori. As a consequence, it is difficult to exhibit the convergence performance, particularly the mean-square-deviation (MSD). In order to obviously demonstrate the excellent mean and mean square convergence performance of KLLAD algorithm, the desired output of an ideal nonlinear synthesis system consisting of the optimum weight vector and the kernelized inputs is given by

$$y_n = \sum_{m=1}^M w_m^* \exp\left(-\frac{\|\mathbf{x}_{\omega,m} - \mathbf{x}_n\|_2^2}{2\xi^2}\right) + z_n, \quad (39)$$

where the kernel bandwidth is set to $\xi = 0.65$, and z_n is the non-Gaussian impulsive CG noise with parameters $K = 1 \times 10^4$, $\sigma_v^2 = 1$, and $p_r = 0.05$.

Furthermore, the CG noise model defined by (2) can be generalized into the mixture Gaussian noise model by changing the distribution of the random variable ϵ_n into the uniform, binary, Rayleigh, Laplacian distributions, etc. Then, we are able to thoroughly investigate the variations of nonlinear filtering performance of KLLAD algorithm in the presence of distinct impulsive noises with large outliers.

Meanwhile, the preselected dictionary with 5 elements utilized for the kernelized nonlinear mapping is given by

$$\begin{aligned} \mathcal{D} &= \{\mathbf{x}_{\omega,1}, \mathbf{x}_{\omega,2}, \mathbf{x}_{\omega,3}, \mathbf{x}_{\omega,4}, \mathbf{x}_{\omega,5}\} \\ &= \left\{ \begin{bmatrix} 0.72 \\ 1.44 \end{bmatrix}, \begin{bmatrix} 3.31 \\ 1.28 \end{bmatrix}, \begin{bmatrix} -3.03 \\ -2.75 \end{bmatrix}, \begin{bmatrix} 1.48 \\ -1.66 \end{bmatrix}, \begin{bmatrix} -1.28 \\ -0.32 \end{bmatrix} \right\}, \end{aligned} \quad (40)$$

which is generated by the coherence criterion from the input signals a priori. Correspondingly, the learning curve of the MSD is defined as

$$\text{MSD}_n = \frac{1}{T} \sum_{t=1}^T \|\mathbf{w}_{t,n} - \mathbf{w}^*\|_2^2, \quad (41)$$

where T is the total number of Monte Carlo runs. The input signal was assumed to be a sequence of statistically independent vectors $\mathbf{x}_n = [x_{1,n}, x_{2,n}]^T$ with correlated samples satisfying $x_{1,n} = 0.5x_{2,n} + \tau_{x,n}$, where $x_{2,n}$ is a white Gaussian noise sequence with variance $\sigma_{x_2}^2 = 1$ and $\tau_{x,n}$ is a white Gaussian so that $x_{1,n}$ has variance $\sigma_{x_1}^2 = 1$. The optimum weight vector \mathbf{w}^* was supposed to be abruptly changed from $[0.3, 0.1, -0.15, -0.3, -0.1]^T$ to $[0.05, 0.25, -0.05, -0.2, -0.35]^T$ at time instant $n = 2 \times 10^4$.

The set of parameters of all used algorithms for Example I is listed in Table 2. It should be pointed that the step-size of KLLAD algorithm is selected by satisfying the convergence condition presented by (38).

The convergence curves of weight coefficients obtained by implementing KLMS, KMCC, KLMP, and KLLAD algorithms are illustrated in Figures 2–5, respectively. As illustrated in Figure 2, the averaged curves of weight coefficients of KLMS algorithm are not able to tend to the optimal weight coefficients during the two stages due to the interference effect of impulsive CG noise. In contrast, the averaged curves of weight coefficients of KMCC, KLMP, and KLLAD algorithms are smooth and all converge to two optimal weight coefficients within the two stages, as shown in Figures 3–5. In addition, the effectiveness of convergence condition on the mean stability (38) is validated by Figure 5. More importantly, Figures 6(a)–6(f) show that the filtering performance of KLLAD algorithm consistently outperforms those of KLMS, KMCC, and KLMP algorithms in terms of robustness, convergence rate, and accuracy of the learning curves of MSD during the two stages using the mixture Gaussian noise model based on the normal, uniform, binary, Rayleigh, Laplacian, and symmetric α -stable distributions, respectively. Therefore, the robust performance of KLLAD algorithm is validated by the simulation results of nonstationary nonlinear system identification in the presence of non-Gaussian impulsive noise.

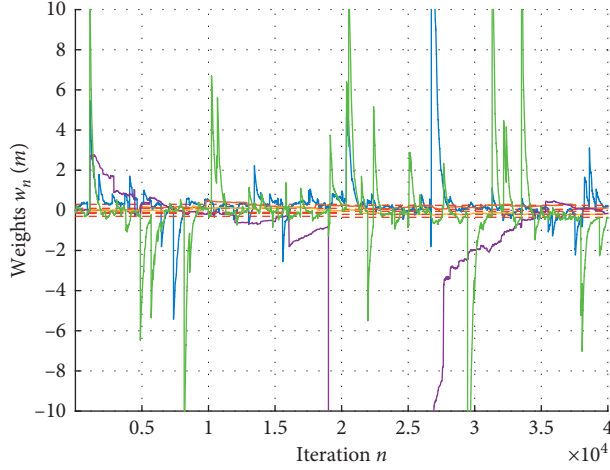
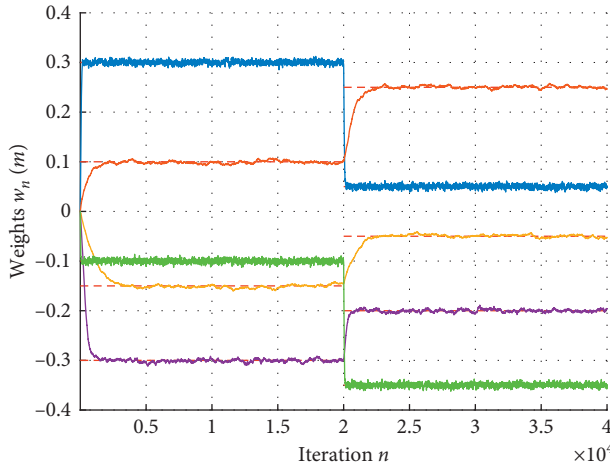
5.2. Example II. As the second example, we consider the input random sequence generated from the following relation:

$$u_n = \rho u_{n-1} + \sigma_u \sqrt{1 - \rho^2} \zeta_n, \quad (42)$$

where ζ_n is a random noise following the i.i.d. standard normal distribution. Here, the correlation factor ρ and the standard deviation σ_u of random sequence u_n were all

TABLE 2: Parameter settings of used algorithms for Example I.

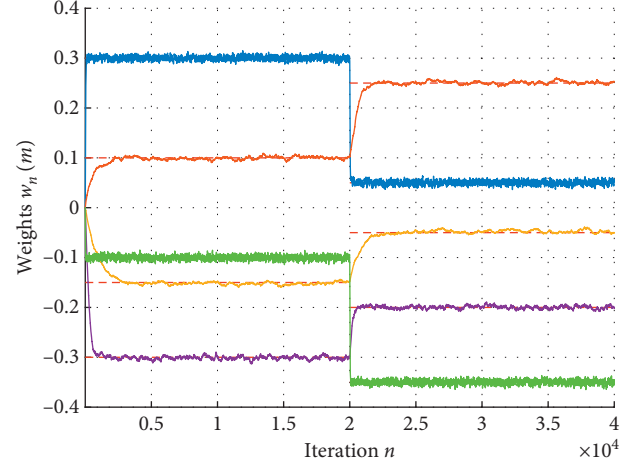
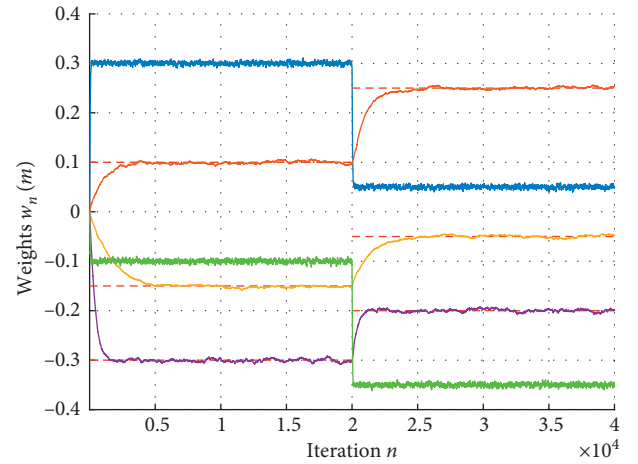
Algorithm	Parameters
KLMS	$\xi = 0.65; \eta = 0.05;$
KMCC	$\xi = 0.65; \eta = 3; \xi_f = 0.05;$
KLMP	$\xi = 0.65; \eta = 0.05; p = 1.1$
KLLAD	$\xi = 0.65; \eta = 0.08; \lambda = 15$

FIGURE 2: Convergence of the coefficients $w_n(m)$ for KLMS algorithm (the dotted lines are for the optimal weight coefficients, and solid lines are for weight coefficients of KLMS algorithm).FIGURE 3: Convergence of the coefficients $w_n(m)$ for KMCC algorithm (the dotted lines are for the optimal weight coefficients, and solid lines are for weight coefficients of KMCC algorithm).

chosen as 0.5. The desired output of nonlinear system is generated as follows:

$$\begin{cases} t_n = 0.5u_n - 0.3u_{n-1}, \\ y_n = t_n - 1.25t_n^2 + 0.25t_n^3 + z_n, \end{cases} \quad (43)$$

where z_n is the non-Gaussian impulsive CG noise with parameters $K = 1 \times 10^4$, $\sigma_v^2 = 1$, and $p_r = 0.1$. At each time instant n , the input vectors $\mathbf{x}_n = [u_n, u_{n-1}]^T$ are used to estimate the nonlinear desired response signal y_n contaminated by the impulsive noise. In order to clearly investigate the characteristic of convergence of weight coefficients, the 6

FIGURE 4: Convergence of the coefficients $w_n(m)$ for KLMP algorithm (the dotted lines are for the optimal weight coefficients, and solid lines are for weight coefficients of KLMP algorithm).FIGURE 5: Convergence of the coefficients $w_n(m)$ for KLLAD algorithm (the dotted lines are for the optimal weight coefficients, and solid lines are for weight coefficients of KLLAD algorithm).

elements of dictionary were chosen by the coherence criterion from the input vectors a priori as follows:

$$\begin{aligned} \mathcal{D} &= \{\mathbf{x}_{\omega,1}, \mathbf{x}_{\omega,2}, \mathbf{x}_{\omega,3}, \mathbf{x}_{\omega,4}, \mathbf{x}_{\omega,5}, \mathbf{x}_{\omega,6}\} \\ &= \left\{ \begin{bmatrix} 0.17 \\ -1.92 \end{bmatrix}, \begin{bmatrix} -1.62 \\ -0.18 \end{bmatrix}, \begin{bmatrix} 0.52 \\ 1.55 \end{bmatrix}, \begin{bmatrix} 2.90 \\ 1.92 \end{bmatrix}, \begin{bmatrix} -2.0 \\ -2.47 \end{bmatrix}, \begin{bmatrix} 2.65 \\ -0.82 \end{bmatrix} \right\}. \end{aligned} \quad (44)$$

The set of parameters of used algorithms for Example II is listed in Table 3. Likewise, the step-size of KLLAD algorithm in second example is also determined from the range of convergence condition according to (38).

Although the fixed dictionary with 6 elements is used to compare the differences of convergent performance among used algorithms, the corresponding optimal weight

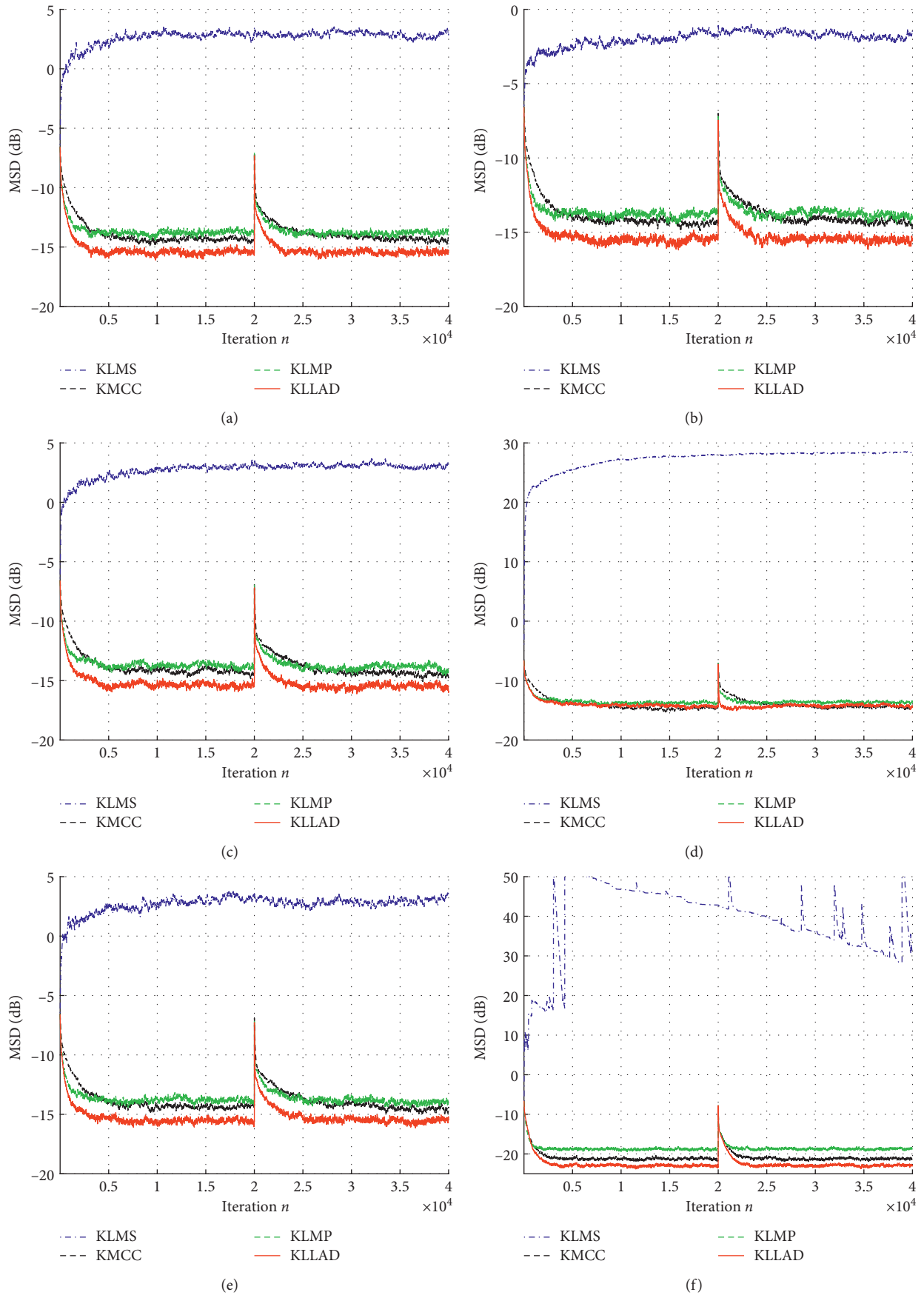
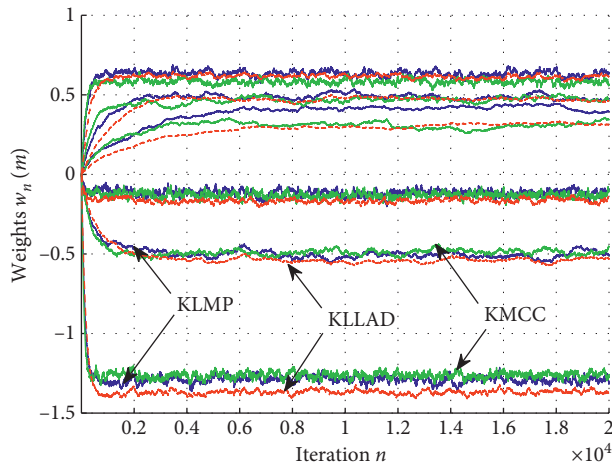


FIGURE 6: Comparisons of learning curves of MSD for KLMS, KMCC, KLMP, and KLLAD algorithms with different non-Gaussian distribution noises. (a) Normal distribution noise. (b) Uniform distribution noise with $(-1, 1)$. (c) Binary distribution noise with $(-1, 1)$. (d) Rayleigh distribution noise with $\sigma^2 = 4$. (e) Laplacian distribution noise with $\mu = 0$ and $\sigma^2 = 1$. (f) Symmetric α -stable distribution noise with $\alpha = 0.9$, $\gamma = 0.5$, and $\delta = 0$.

TABLE 3: Parameter settings of used algorithms for Example II.

Algorithm	Parameters
KMCC	$\xi = 0.75; \eta = 1; \xi_f = 0.5;$
KLMP	$\xi = 0.75; \eta = 0.1; p = 0.7$
KLLAD	$\xi = 0.75; \eta = 0.1; \lambda = 15$

FIGURE 7: Convergence of the coefficients $w_n(m)$ for KMCC, KLMP, and KLLAD algorithms.

coefficients are still unknown in fact. The KLMS algorithm is ignored due to its severe instability of convergence. Figure 7 shows that the mean value curves of weight coefficients of KLLAD algorithm converge much smoother and faster compared with the KLMP and KMCC algorithms, which verified the superiorities of KLLAD algorithm.

6. Conclusion

In this paper, we presented a novel KLLAD algorithm based on the logarithmic error cost criterion under the assumption of non-Gaussian impulsive CG noise. The KLLAD algorithm can effectively mitigate the instability of convergence learning curves caused by impulsive noise. The simulation results demonstrated that the proposed KLLAD algorithm has an excellent performance compared to the KLMS, KMCC, and KLMP algorithms in the presence of impulsive noise.

Data Availability

The ideal mathematical model data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant nos. 61701200 and 61701529.

References

- [1] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.
- [2] J. M. Gil-Cacho, M. Signoretto, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Nonlinear acoustic echo cancellation based on a sliding-window leaky kernel affine projection algorithm," *IEEE Transactions on Audio Speech and Language Processing*, vol. 21, no. 9, pp. 1867–1878, 2013.
- [3] Y. Nakajima and M. Yukawa, "Nonlinear channel equalization by multi-kernel adaptive filter," in *Proceedings of the IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications*, pp. 384–388, SPAWC, Cesme, Turkey, June, 2012.
- [4] T. Wang, J. Chen, P. Honeine, and H. Snoussi, "Abnormal event detection via multikernel learning for distributed camera networks," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 989450, 213 pages, 2015.
- [5] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*, John Wiley & Sons, New-York, NY, USA, 2010.
- [6] M. Yukawa, "Multikernel adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4672–4682, 2012.
- [7] B. Chen, S. Zhao, P. Zhu, and J. C. Principe, "Quantized kernel recursive least squares algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1484–1491, 2013.
- [8] M. Zhang, X. Wang, X. Chen, and A. Zhang, "The kernel conjugate gradient algorithms," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4377–4387, 2018.
- [9] W. Liu, P. P. Pokharel, and J. C. Principe, "The kernel least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 543–554, 2008.
- [10] M. A. Chitre, J. R. Potter, and S.-H. Ong, "Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 2, pp. 497–503, 2006.
- [11] A. Mahmood, M. Chitre, and M. A. Armand, "Detecting OFDM signals in alpha-stable noise," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3571–3583, 2014.
- [12] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1601–1611, 1998.
- [13] J. Ni, J. Chen, and X. Chen, "Diffusion sign-error LMS algorithm: formulation and stochastic behavior analysis," *Signal Processing*, vol. 128, pp. 142–149, 2016.
- [14] W. Gao, Y. Xu, and J. Chen, "Stochastic analysis of diffusion dual sign algorithm for cyclostationary white Gaussian inputs and nonstationary system," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 7, pp. 1354–1358, Jul. 2020.
- [15] W. Gao and J. Chen, "Transient analysis of signed LMS algorithms with cyclostationary colored Gaussian inputs," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 12, pp. 3562–3566, 2020.
- [16] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Principe, "Generalized correntropy for Robust_newline adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3376–3387, 2016.
- [17] W. Ma, B. Chen, H. Zhao, G. Gui, J. Duan, and J. C. Principe, "Sparse least logarithmic absolute difference algorithm with correntropy-induced metric penalty," *Circuits, Systems, and Signal Processing*, vol. 35, no. 3, pp. 1077–1089, 2016.

- [18] B. Chen, L. Xing, B. Xu, H. Zhao, N. Zheng, and J. C. Principe, "Kernel risk-sensitive loss: definition, properties and application to robust adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2888–2901, 2017.
- [19] V. C. Gogineni and S. Mula, "Logarithmic cost based constrained adaptive filtering algorithms for sensor array beamforming," *IEEE Sensors Journal*, vol. 18, no. 14, pp. 5897–5905, 2018.
- [20] K. Xiong and S. Wang, "Robust least mean logarithmic square adaptive filtering algorithms," *Journal of the Franklin Institute*, vol. 356, no. 1, pp. 654–674, 2019.
- [21] B. Binwei Weng and K. E. Barner, "Nonlinear system identification in impulsive environments," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2588–2594, 2005.
- [22] S. Zhao, B. Chen, and J. C. Principe, "Kernel adaptive filtering with maximum correntropy criterion," in *Proceedings of the The 2011 International Joint Conference on Neural Networks*, pp. 2012–2017, San Jose, CA, USA, August, 2011.
- [23] S. Wang, J. Feng, and C. K. Tse, "Kernel affine projection sign algorithms for combating impulse interference," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 60, no. 11, pp. 811–815, 2013.
- [24] L. Lu, H. Zhao, and B. Chen, "Improved-variable-forgetting-factor recursive algorithm based on the logarithmic cost for Volterra system identification," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 6, pp. 588–592, 2016.
- [25] Y. He, F. Wang, J. Yang, H. Rong, and B. Chen, "Kernel adaptive filtering under generalized maximum correntropy criterion," in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 1738–1745, Vancouver, BC, Canada, July, 2016.
- [26] X. Luo, J. Deng, W. Wang, J.-H. Wang, and W. Zhao, "A quantized kernel learning algorithm using a minimum kernel risk-sensitive loss criterion and bilateral gradient technique," *Entropy*, vol. 19, no. 7, p. 365, 2017.
- [27] X. Li, Q. Lu, Y. Dong, and D. Tao, "Robust subspace clustering by Cauchy loss function," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 7, pp. 2067–2078, 2019.
- [28] N. Guan, T. Liu, Y. Zhang, D. Tao, and L. S. Davis, "Truncated Cauchy non-negative matrix factorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 246–259, 2019.
- [29] W. Shi, K. Xiong, and S. Wang, "Multikernel adaptive filters under the minimum Cauchy kernel loss criterion," *IEEE Access*, vol. 7, pp. 120 548–120 558, 2019.
- [30] Q. Wu, Y. Li, Y. V. Zakharov, W. Xue, and W. Shi, "A kernel affine projection-like algorithm in reproducing kernel Hilbert space," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 10, pp. 2249–2253, 2020.
- [31] W. Gao and J. Chen, "Kernel least mean p -Power algorithm," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 996–1000, Jul. 2017.
- [32] W. Ma, J. Duan, W. Man, H. Zhao, and B. Chen, "Robust kernel adaptive filters based on mean p -power error for noisy chaotic time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 58, pp. 101–110, 2017.
- [33] T. Zhang and S. Wang, "Nyström kernel algorithm under generalized maximum correntropy criterion," *IEEE Signal Processing Letters*, vol. 27, pp. 1535–1539, 2020.
- [34] W. Gao, M. Song, and J. Chen, "Tracking analysis of Gaussian kernel signed error algorithm for time-variant nonlinear systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 10, pp. 2289–2293, 2020.
- [35] S. Wang, W. Wang, K. Xiong, H. H. C. Iu, and C. K. Tse, "Logarithmic hyperbolic Cosine adaptive filter and its performance analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 4, pp. 2512–2524, 2021.
- [36] H. Zhang, B. Yang, L. Wang, and S. Wang, "General Cauchy conjugate gradient algorithms based on multiple random Fourier features," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1859–1873, 2021.
- [37] M. O. Sayin, N. D. Vanli, and S. S. Kozat, "A novel family of adaptive filtering algorithms based on the logarithmic cost," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4411–4424, 2014.
- [38] Y. Zou, S. C. Chan, and T. S. Ng, "A recursive least m-estimate (RLM) adaptive filter for robust filtering in impulse noise," *IEEE Signal Processing Letters*, vol. 7, no. 11, pp. 324–326, 2000.
- [39] B. Schölkopf, R. Herbrich, and R. Williamson, *A generalized representer theorem*, Vol. 81, Springer, Berlin, Heidelberg, 2000.
- [40] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, 2004.
- [41] W. Gao and J. Chen, "Transient performance analysis of zero-attracting Gaussian kernel LMS algorithm with pre-tuned dictionary," *IEEE Access*, vol. 7, pp. 135 770–135 779, 2019.
- [42] W. Gao, J. Chen, C. Richard, and J. Huang, "Online dictionary learning for kernel LMS," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2765–2777, 2014.