Nonlinear Spline Adaptive Filtering Against Non-Gaussian Noise

Wenyan Guo[1] 🔘 · **Yongfeng Zhi**[1] 🔘

Abstract

In this paper, a generalized maximum Versoria criterion algorithm (GMVC) based on wiener spline adaptive filter, called SAF–GMVC, is proposed. The proposed algorithm is used for nonlinear system identification under non-Gaussian environment. To improve the convergence performance of the SAF–GMVC, the momentum stochastic gradient descent (MSGD) is introduced. In order to further reduce the steady-state error, the variable step-size algorithm is introduced, called as SAF–GMVC–VMSGD. Simulation results demonstrate that SAF–GMVC–VMSGD achieves better filtering effective against non-Gaussian noise.

Keywords Spline adaptive filter · Nonlinear system identification · Robust filtering · Generalized maximum Versoria criterion · Non-Gaussian noise

1 Introduction

Adaptive filtering (AF) algorithm is widely used in signal processing, which has been utilized for system identification [20], prediction model [5], interference elimination [19], inverse modeling [23]. Least mean square (LMS) is one of the most popular AF algorithms due to its simplicity and good performance [7]. However, it cannot deal with nonlinear structure. One solution is the spline adaptive filter (SAF), which uses spline interpolation to interpolate data from known nodes, formed SAF–LMS [14]. Some constructs have been proposed of the SAF algorithm, examples as: Wiener SAF [10], Hammerstein SAF [15], Sandwich 1 SAF, Sandwich 2 SAF [16].

✉ Yongfeng Zhi
yongfeng@nwpu.edu.cn

Wenyan Guo
wenyan0117@126.com

[1] Research & Development Institute in Shenzhen, Northwestern Polytechnical University, Shenzhen, China

However, the SAF–LMS performance degradation in the presence of the non-Gaussian environment, due to the use of the mean square error (MSE) criterion [14]. To achieve better robuster filtering performance, various nonlinear optimization criteria have been proposed. Examples include: the LMS adaptive scheme has been extended to the normalized version, called SAF–NLMS [6]. Peng et al. propose SAF–MCC algorithm, combining SAF with maximum correntropy criterion (MCC), which shown robust to large outliers [12]. The reference [24] proposed a novel algorithm, under the impulsive environment. That algorithm through introduce momentum in stochastic gradient descent, called SAF–ARC–MMSGD. The reference [4] proposed a sign normalised least mean square algorithm (SNLMS) based on Wiener spline adaptive filter, and the variable step-size scheme is introduced, SAF–VSS–SNLMS. Moreover, the weight update of the normalized subband spline adaptive filter algorithm is conducted using the principle of minimum disturbance [22]. Referring to [2,8], compared with the maximum correntropy criterion, the Versoria criterion has faster convergence speed and stronger robustness when applied to adaptive filtering. Huang et al. applied the Versoria function to the adaptive filtering to improve the robustness [8]. In order to further improve the robustness of the system, a robust weight-constraint decorrelation normalized maximum Versoria algorithm is designed [25]. Jain et al. applied the Versoria function to the research of kernel adaptive filtering [9]. Other researchers applied the maximum Versoria adaptive filtering algorithm to the study of power energy [1,18]. The filter algorithm based on Generalized maximum Versoria criterion (GMVC) has strong robustness, so this paper chooses GMVC algorithm as the adaptive strategy.

In this paper, we propose a novel robust wiener type spline adaptive filter algorithm, which is called SAF–GMVC–VMSGD. Through variable step size-momentum stochastic gradient descent update the loss function weight. Generalized maximum Versoria criterion as a loss function. Compared with the SAF-MCC algorithm, it does not require exponential operation. Under non-Gaussian noise, compare with SAF–MCC, simulation results demonstrate better efficiency of the proposed SAF–GMVC–VMSGD algorithm.

2 Wiener Spline Adaptive Filter

Wiener spline adaptive filter architecture consists of the cascade of a suitable number of linear filter, and a memoryless nonlinear function implemented by the spline interpolation scheme [17].

The input signal of SAF at instant n is $x(n)$, the input signal $x(n)$ passes through a FIR filter, obtaining an intermediate output $s(n)$. Input signal defined as $\mathbf{x}_n = \begin{bmatrix} x_n \ x_{n-1} \ \ldots \ x_{n-M+1} \end{bmatrix}^T$. M represents the order of tap weights of the FIR filter. Intermediate output $s(n)$ is defined as

$$s(n) = \mathbf{w}_n^T \mathbf{x}_n \tag{1}$$

where $\mathbf{w}_n = \begin{bmatrix} w_0 \ w_1 \ \ldots \ w_{M-1} \end{bmatrix}^T$ is the weight vector of FIR filter at instant n.

Interpolation functions are smooth parametric curves that interpolate by collecting defined control points in a lookup table (LUT). Cubic spline curve mainly includes

B spline and CR spline [3]. Since CR spline has the characteristic of passing all control nodes, Therefore, CR splines are the only objects considered in this article. The interpolation process starts with the calculation of normalized horizontal coordinates u and interval indexes i based on the LUT.

Let $\mathbf{Q}_i = \begin{bmatrix} q_{x,i}, q_{y,i} \end{bmatrix}^T$ for $i = 0, 1, 2, \ldots, N$ represent the $N + 1$ control points (knots), where the $x - axis$ value $q_{x,j}$ are uniformly distributed and order to $q_{x,N} > \cdots > q_{x,1} > q_{x,0}$. Then we can obtain i and u by the following equations

$$i = \left\lfloor \frac{s(n)}{\Delta x} \right\rfloor + \frac{N}{2} \tag{2}$$

$$u_n = \frac{s(n)}{\Delta x} - \left\lfloor \frac{s(n)}{\Delta x} \right\rfloor \tag{3}$$

where $\Delta x = q_{i+1} - q_i$ is the uniform space between knots, $\lfloor \cdot \rfloor$ is the floor operator.

To calculate the filter output $y(n)$, suppose that $y_n = \varphi(x_n)$ is the spline function to be estimated. Which is a spline interpolation of four adaptive control points of the ith interval contained in a LUT, which can be expressed

$$y_n = \varphi_i(u_n) = \mathbf{u}_n^T \mathbf{C} \mathbf{q_i}(n) \tag{4}$$

where the vector $\mathbf{u}_n$ is defined as $\mathbf{u}_n = [u^3, u^2, u, 1]^T$, the vector $\mathbf{q}_i(n)$ is defined by $\mathbf{q_i}(n) = [q_i, q_{i+1}, q_{i+2}, q_{i+3}]^T$, at the instant n. And $\mathbf{C}$ is the basis matrix of CR-spline defined [6]

$$\mathbf{C} = \frac{1}{2} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix} \tag{5}$$

The derivative of y_n with respect to u_n is calculated

$$\frac{\partial y_n}{\partial u} = \frac{\partial \varphi_i(u)}{\partial u} = \dot{\mathbf{u}}^T \mathbf{C} \mathbf{q_i}(n) \tag{6}$$

where $\dot{\mathbf{u}} = [3u^2, 2u, 1, 0]^T$.

In this work, Wiener spline adaptive filter consists that the cascade of an adaptive FIR filter and cubic CR-spline. The basic block diagram of SAF is shown in Fig. 1.

With reference to Fig. 1, the priori error e_n is defined

$$e_n = d(n) - y_n \tag{7}$$

Based on the basic framework above, the adaptive strategy of SAF is derived by minimizing the instantaneous square error which is given [14].
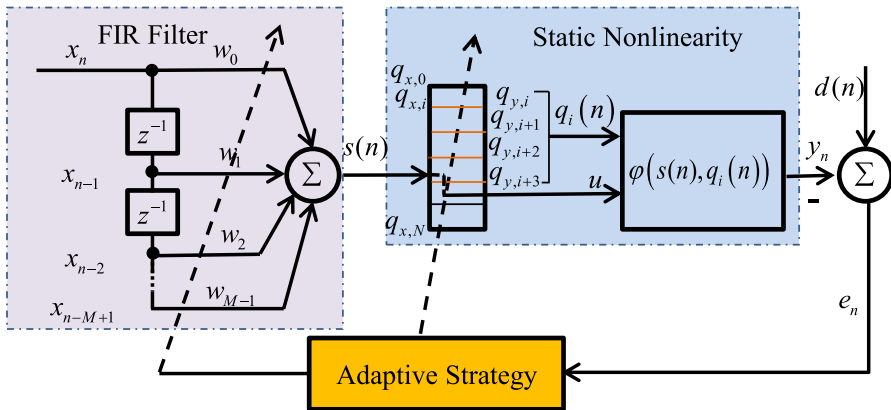
Fig. 1 Structure of Wiener spline adaptive filter

$$J(\mathbf{w}_n, \mathbf{q_i}(n)) = e_n^2 \tag{8}$$

Derivative of $J(\mathbf{w}_n, \mathbf{q_i}(n))$ as regards $\mathbf{w}_n$ and $\mathbf{q_i}(n)$, and using the stochastic gradient descent (SGD) method, the adaptive process of SAF can be expressed

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_w e_n \dot{\mathbf{u}}^{\mathbf{TCq_i}}(n)\mathbf{x_n}/\Delta x \tag{9}$$

$$\mathbf{q_i}(n+1) = \mathbf{q_i}(n) + \mu_q e_n \mathbf{C^T u_n} \tag{10}$$

where μ_w and μ_q represent the learning rates for the weights and for the control points, separately. The above traditional SAF method, called SAF–LMS [14].

3 SAF–GMVC–VMSGD Algorithm

In this part, in order to improve the robustness of SAF against non-Gaussian noise, we introduce the generalized Versoria criterion (GMVC) combines with the SAF, present the proposed algorithm SAF–GMVC. And then, to further improve the convergence performance, the variable step-size momentum stochastic gradient descent (VMSGD) algorithm is introduced, the algorithm called SAF–GMVC–VMSGD.

3.1 Generalized Versoria Function

In non-Gaussian noise environment, in order to more effectively use the system output error information to improve the convergence speed or steady-state error of the filter. Many error nonlinear adaptive filtering algorithms with saturation characteristics have been proposed. For example, adaptive filtering algorithm based on maximum correntropy criterion (MCC). When non-Gaussian noise appears, the system output error is very large. The error nonlinearity of MCC algorithm approaches the saturation
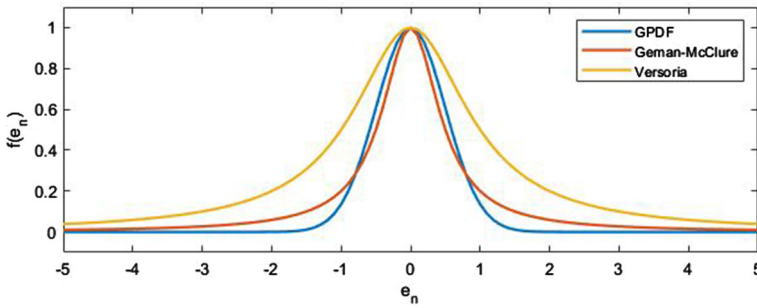
Fig. 2 Comparison of steepness of the GPDF, the Versoria function, the Geman–McClure function

value 0. At this time, the weight coefficient vector of MCC algorithm is almost not updated, thus suppressing the non-Gaussian noise. However, since MCC algorithms uses Gaussian probability density function (GPDF) as the cost function, it has a large computational complexity. An adaptive filtering algorithm with stronger robustness in non-Gaussian noise environment is proposed, namely maximum Versoria criterion (MVC) algorithm. Compared with MCC algorithm, this algorithm does not contain exponential operation and has lower steady-state error.

The cost function of the maximum correntropy criterion (MCC) algorithm is the GPDF, which can be written as[21]

$$f(e_n) = \exp(-\frac{e_n^2}{2\varpi^2}) \tag{11}$$

where $\varpi > 0$ is the kernel width. The GPDF is high calculation cost when it used in signal processing, especially in adaptive filtering.

For adaptive filtering, designing a suitable and high-efficiency cost framework is an important issue. In order to draw forth the proposed cost function, we first introduce the standard Versoria criterion, which is defined as [8,13,25],

$$f(e_n) = \frac{8a^3}{4a^2 + e_n^2} = \frac{2a}{1 + \left(\frac{e_n}{2a}\right)^2} \tag{12}$$

where e_n represent error, $a > 0$ is the radius of the generating circle of Versoria. The centroid of the generating circle is located at $(0, a)$. A larger value of the radius a leads to a steeper Versoria [8]. The loss function of Geman–McClure estimator defined as $f(e_n) = \frac{e_n^2}{\vartheta^2 + e_n^2}$, ϑ denotes the positive parameter which modulates the shape of the loss function [11].

As a comparison, Fig. 2 plots the GPDF ($\varpi = 0.5$), the Geman–McClure function ($\vartheta = 0.5$) and the standard Versoria function ($a = 0.5$). From this figure, we can observe that the Versoria function is less steep than the two other functions for the error e_n. This means that the error along the direction of the gradient ascent of the Versoria is faster to reach the optimal point than do the GPDF and the Geman–McClure function. From the perspective of the adaptive filter, the gradient ascent algorithm based on the
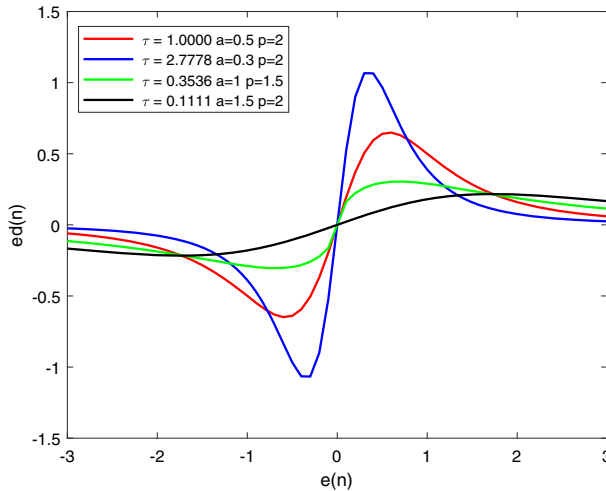
Fig. 3 Slope curves of the cost function with different parameters

Versoria function has a faster convergence rate to reach the same steady-state error than do the GPDF and the Geman–McClure function algorithm.

Generalized Versoria criterion can be regarded as [1,2,9,18],

$$f(e_n) = \frac{2a}{1 + \left|\frac{e_n}{2a}\right|^p} = \frac{2a}{1 + \tau |e_n|^p} \tag{13}$$

$p > 0$ is the shape parameter, and $\tau = (2a)^{-p}$. As a special case, it reduced to the standard Versoria function when $p = 2$.

3.2 SAF–GMVC

According to the generalized Versoria function in linear adaptive filtering show stronger robustness, and there is no exponential operation in the formula, thus reducing the computational complexity of the system. Used Versoria function as the cost function:

$$J(e_n) = J(\mathbf{w}_n, \mathbf{q_i}(n)) = E\left[\frac{1}{1 + \tau |e_n|^p}\right] \tag{14}$$

The gradient of Eq. (14) with respect to e_n, and denoting the result by $e_d(n)$,

$$e_d(n) = \nabla J(\mathbf{w}_n, \mathbf{q_i}(n)) = \frac{\tau p}{(1 + \tau |e_n|^p)^2} |e_n|^{p-1} \text{sign}(e_n) \tag{15}$$

where $E[\cdot]$ denotes the expectation operator. As can be seen from Eq. (14), when an non-Gaussian interference corrupted e_n appears, leads to $J(e_n) \to 0$, which plays the role of suppressing non-Gaussian interference. Eliminate the updating of weight vector based on the non-Gaussian interference.

Figure 3 shows the derivative curves for different parameters, the slope curves dropping gradually to zero, which will reduce the effect of non-Gaussian noise interference on cost function.

SAF combine with generalized Versoria cost function. Taking the derivative of Eq. (14) with respect to $\mathbf{w_n}$.

$$\mathbf{g_w}(n) = \frac{\partial J(e_n)}{\partial \mathbf{w}_n} = \frac{J(\mathbf{w}_n, \mathbf{q_i}(n))}{\partial \mathbf{w}_n} = -e_d(n)\frac{\partial \varphi_i(u)}{\partial u}\frac{\partial u}{\partial s(n)}\frac{\partial s(n)}{\partial \mathbf{w}_n} \tag{16}$$

Through Eqs. (1), (3), (6), can get

$$\mathbf{g_w}(n) = -e_d(n)\dot{\mathbf{u}}^{\mathbf{T}}\mathbf{Cq_i}(n)\mathbf{x_n}/\Delta x \tag{17}$$

Taking the derivative of Eq. (14) with respect to $\mathbf{q_i}(n)$,

$$\mathbf{g_q}(n) = \frac{\partial J(e_n)}{\partial \mathbf{q_i}(n)} = \frac{J(\mathbf{w_n}, \mathbf{q_i}(n))}{\partial \mathbf{q_i}(n)} = -e_d(n)\frac{\partial \varphi_i(u)}{\partial \mathbf{q_i}(n)} \tag{18}$$

Through Eq. (4), we can get

$$\mathbf{g_q}(n) = -e_d(n)\mathbf{C^T u_n} \tag{19}$$

Using gradient descent method can be to obtain the iterative learning rules of SAF–GMVC.

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_w e_d(n)\dot{\mathbf{u}}^{\mathbf{T}}\mathbf{Cq_i}(n)\mathbf{x_n}/\Delta x \tag{20}$$

$$\mathbf{q_i}(n+1) = \mathbf{q_i}(n) + \mu_q e_d(n)\mathbf{C^T u_n} \tag{21}$$

where the parameters μ_w and μ_q represent the learning rates for the weights and for the control points, respectively, for simplicity, incorporate the others constant values.

3.3 Variable Step-Size Momentum SGD Algorithm

In order to improve the convergence performance, the momentum term and variable step-size term are introduced. The specific operation is as follows. In order to increase the convergence speed, the momentum stochastic gradient descent algorithm is introduced, called SAF–GMVC–MMSGD. In this method, the exponential moving average value of the historical gradients are defined as the momentum term, and the actual modification of the weight vector at current instant depends on both the current gradient and the momentum term [24], which can be expressed

$$\mathbf{m}(n) = \sigma \mathbf{m}(n-1) + \mu \mathbf{g}(n) \tag{22}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mathbf{m}(n) \tag{23}$$

where $\mathbf{m}(n)$ represents the momentum term also the actual modification of the weight vector at instant n, $\mathbf{w}_n$ and $\mathbf{g}(n)$ are the weight vector and gradient vector, respectively, σ is the momentum parameter, and μ is the learning rate.

In order to further reduce the steady-state error, the variable step-size algorithms is introduced, named SAF–GMVC–VMSGD. $\mu_w(n)$ and $\mu_q(n)$ represent the learning rates for the weights and for the control points, respectively.

$$\mu_w(n) = \rho\mu_w(n-1) + (1-\rho)\min\left[\hat{e}_o^2(n), \mu_w(n-1)\right] \tag{24}$$

$$\mu_q(n) = \rho\mu_q(n-1) + (1-\rho)\min\left[\hat{e}_o^2(n), \mu_q(n-1)\right] \tag{25}$$

where ρ is the forgetting factor approaching one. $\hat{e}_o^2(n)$ is the estimate of squared value of the non-Gaussian noise error [4].

$$\hat{e}_o^2(n) = \lambda\hat{e}_o^2(n-1) + c_1(1-\lambda)med(\gamma_n) \tag{26}$$

where λ is forgetting factor close to but smaller than one, $c_1 = 1.483(1 + \frac{5}{N_w - 1})$ is a finite correction factor [4] and N_w is the data window. $\gamma_n = \left[e_n^2, e_{n-1}^2, \ldots, e_{n-N_w+1}^2\right]$ and $med(\cdot)$ denotes the median operator.

$\mu(n)$ bring Eq. (22), we can get

$$\mathbf{m}(n) = \sigma\mathbf{m}(n-1) + \mu(n)\mathbf{g}(n) \tag{27}$$

Due to the momentum term will cause oscillations at the neighborhood of minimum value, two algorithms are introduced. First, the decay technique about momentum factor σ.

$$\sigma_n = c_0\sigma_{n-1} \tag{28}$$

where c_0 is the decay coefficient $(0.9995 \sim 1)$. When $c_0 = 1$, the decay technique will be ineffective.

Second, the direction compare of the momentum term and the current gradient by their inner product. If the inner product is non-negative, we use the momentum method. In reverse, it means the momentum term and current gradient orient to opposite directions. We only choose current gradient [24]. Expressed as

$$\mathbf{m_w}(n) = \begin{cases} \mu_w(n)\mathbf{g_w}(n), & \langle\mathbf{m_w}(n-1), \mathbf{g_w}(n)\rangle < 0 \\ \sigma_n\mathbf{m_w}(n-1) + \mu_w(n)\mathbf{g_w}(n), & \langle\mathbf{m_w}(n-1), \mathbf{g_w}(n)\rangle \geq 0 \end{cases} \tag{29}$$

$$\mathbf{m_q}(n) = \begin{cases} \mu_q(n)\mathbf{g_q}(n), & \langle\mathbf{m_q}(n-1), \mathbf{g_q}(n)\rangle < 0 \\ \sigma_n\mathbf{m_q}(n-1) + \mu_q(n)\mathbf{g_q}(n), & \langle\mathbf{m_q}(n-1), \mathbf{g_q}(n)\rangle \geq 0 \end{cases} \tag{30}$$

where $\mathbf{g_w}(n)$ and $\mathbf{g_q}(n)$ are defined at Eqs. (17), (19), $\mathbf{m_w}(n)$ and $\mathbf{m_q}(n)$ represent the momentum terms at instant n with regard to w and q_i. The updating equations can be rewritten as

$$\mathbf{w_{n+1}} = \quad \mathbf{w_n} - \mathbf{m_w}(n) \tag{31}$$

$$\mathbf{q_i}(n+1) = \quad \mathbf{q_i}(n) - \mathbf{m_q}(n) \tag{32}$$

3.4 The Proposed SAF–GMVC–VMSGD Algorithm Brief

The proposed algorithm SAF–GMVC–VMSGD, combine the generalized Versoria cost function is adopted to increase the robustness against non-Gaussian noises, with the variable step-size momentum stochastic gradient descent (VMSGD) is introduced to further improve the convergence performance. The proposed algorithm is summarized in Algorithm 1.

Algorithm 1 Summary of the SAF–GMVC–VMSGD algorithm

1: Initialize: $w_0, q_0, \mu_w(0), \mu_q(0), \tau, p, a, \sigma_0, c_0, m_w(0), m_q(0), N_w, e_0^2(0), \lambda, \rho$
2: **for** $n = 1, 2, \ldots$ **do**
3: $\quad s(n) = w_n^T x_n$
4: $\quad i = \left\lfloor \frac{s(n)}{\Delta x} \right\rfloor + \frac{N}{2}$
5: $\quad u_n = \frac{s(n)}{\Delta x} - \left\lfloor \frac{s(n)}{\Delta x} \right\rfloor$
6: $\quad \mathbf{u_n} = \left[u^3, u^2, u, 1 \right]^T$
7: $\quad \mathbf{q_i}(n) = \left[q_i, q_{i+1}, q_{i+2}, q_{i+3} \right]^T$
8: $\quad y_n = \varphi_i(u_n) = \mathbf{u_n^T C q_i}(n)$
9: $\quad e_n = d_n - y_n$
10: $\quad e_d(n) = \tau p \frac{1}{(1+\tau|e_n|^p)^2} |e_n|^{p-1} sign(e_n)$
11: $\quad \mathbf{g_w}(n) = - e_d(n) \dot{\mathbf{u}}^T \mathbf{C q_i}(n) \mathbf{x_n}/\Delta x$
12: $\quad \mathbf{g_q}(n) = -e_d(n) \mathbf{C^T u_n}$
13: $\quad \sigma_n = c_0 \sigma_{n-1}$
14: $\quad \hat{e}_o^2(n) = \lambda \hat{e}_o^2(n-1) + c_1(1-\lambda) med(\gamma_n)$
15: $\quad \mu_w(n) = \rho \mu_w(n-1) + (1-\rho) \min \left[\hat{e}_o^2(n), \mu_w(n-1) \right]$
16: $\quad \mu_q(n) = \rho \mu_q(n-1) + (1-\rho) \min \left[\hat{e}_o^2(n), \mu_q(n-1) \right]$
17: $\quad$ **if** $\langle \mathbf{m_w}(n-1), \mathbf{g_w}(n) \rangle < 0$ **then** $\mathbf{m_w}(n) = \mu_w(n)\mathbf{g_w}(n)$
18: $\quad$ **else** $\mathbf{m_w}(n) = \sigma_n \mathbf{m_w}(n-1) + \mu_w(n)\mathbf{g_w}(n)$
19: $\quad \mathbf{w_{n+1}} = \mathbf{w_n} - \mathbf{m_w}(n)$
20: $\quad$ **if** $\langle \mathbf{m_q}(n-1), \mathbf{g_q}(n) \rangle < 0$ **then** $\mathbf{m_q}(n) = \mu_q(n)\mathbf{g_q}(n)$
21: $\quad$ **else** $\mathbf{m_q}(n) = \sigma_n \mathbf{m_q}(n-1) + \mu_q(n)\mathbf{g_q}(n)$
22: $\quad \mathbf{q_i}(n+1) = \mathbf{q_i}(n) - \mathbf{m_q}(n)$

---

4 Convergence Analysis

In this section, we analyzed the convergence property of the proposed algorithm in nonlinear identification system. As shown in Fig. 4, w^* is the weight vector and q^* is the control point, of real system. We take $|e_{n+1}| < |e_n|$ as a condition for judging whether
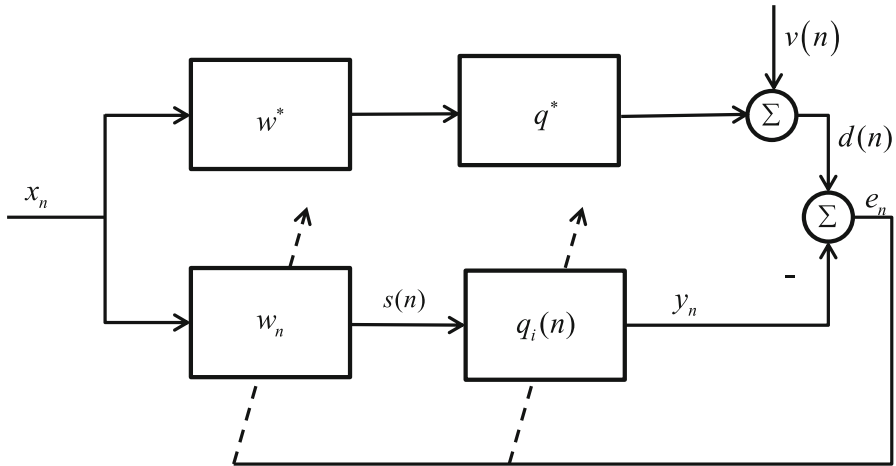
Fig. 4 Block diagram of the system identification using Wiener SAF

the algorithm convergence. The convergence analysis of the proposed algorithm can be implemented in two separate phases with regard to **w** and **q_i**.

4.1 Convergence Analysis of w

Taking a first-order Taylor series expansion of the error e_{n+1} at instant n.

$$e_{n+1} = e_n + \frac{\partial e_n}{\partial \mathbf{w}_n^T} \Delta \mathbf{w}_n + h.o.t \tag{33}$$

where h.o.t. represents the high-order term of Taylor series. The second term can be expressed as

$$\frac{\partial e_n}{\partial \mathbf{w}_n^T} = -\frac{\partial y_n}{\partial u} \frac{\partial u}{\partial s(n)} \frac{\partial s(n)}{\partial \mathbf{w}_n^T} = -\dot{\mathbf{u}}_n^{\mathbf{TCq_i}}(n)\mathbf{x_n}^T / \Delta x \tag{34}$$

$$\Delta \mathbf{w}_n = \mathbf{w}_{n+1} - \mathbf{w}_n \tag{35}$$

Then reference to Eqs. (29), (31), (35) can be expressed as

$$\Delta \mathbf{w_n} = \begin{cases} -\mu_w(n)\mathbf{g_w}(n), & \langle \mathbf{m_w}(n-1), \mathbf{g_w}(n)\rangle < 0 \\ -\sigma_n\mathbf{m_w}(n-1) - \mu_w(n)\mathbf{g_w}(n), & \langle \mathbf{m_w}(n-1), \mathbf{g_w}(n)\rangle \geq 0 \end{cases} \tag{36}$$

Consider the case of $\langle \mathbf{m_w}(n-1), \mathbf{g_w}(n)\rangle \geq 0$, substituting Eqs. (34), (36) into Eq. (33) and ignoring the higher order terms, then we can get

$$e_{n+1} = e_n + [\sigma_n\mathbf{m_w}(n-1) + \mu_w(n)\mathbf{g_w}(n)]\dot{\mathbf{u}}_n^{\mathbf{TCq_i}}(n)\mathbf{x_n}^T / \Delta x \tag{37}$$

Bring Eq. (17) into Eq. (37),

$$e_{n+1} = e_n \left[1 - \frac{\sigma_n \mathbf{g_w^T}(n)\mathbf{m_w}(n-1) + \mu_w(n)\mathbf{g_w^T}(n)\mathbf{g_w}(n)}{e_d(n)e_n} \right] \tag{38}$$

$|e_{n+1}| < |e_n|$ ensure the convergence of the algorithm. The following relation holds

$$\left| 1 - \frac{\sigma_n \mathbf{g_w^T}(n)\mathbf{m_w}(n-1) + \mu_w(n)\mathbf{g_w^T}(n)\mathbf{g_w}(n)}{e_d(n)e_n} \right| < 1 \tag{39}$$

It is derive that

$$0 < \frac{\sigma_n \mathbf{g_w^T}(n)\mathbf{m_w}(n-1) + \mu_w(n)\mathbf{g_w^T}(n)\mathbf{g_w}(n)}{e_d(n)e_n} < 2 \tag{40}$$

$\mathbf{g_w^T}(n)\mathbf{m_w}(n-1)$ and $\mathbf{g_w^T}(n)\mathbf{g_w}(n)$ is non-negative, e_n and $e_d(n)$ are of the same sign, the bound of the learning rate under convergence condition is

$$0 < \mu_w(n) < \frac{2e_d(n)e_n - \sigma_n \mathbf{g_w^T}(n)\mathbf{m_w}(n-1)}{\mathbf{g_w^T}(n)\mathbf{g_w}(n)} \tag{41}$$

Consider the case of $\langle \mathbf{m_w}(n-1), \mathbf{g_w}(n) \rangle < 0$, substituting Eqs. (34), (36) into Eq. (33) and ignoring the higher order terms, then we can get.

$$e_{n+1} = e_n + \mu_w(n)\mathbf{g_w}(n)\dot{\mathbf{u}}_n^{\mathbf{TCq_i}}(n)\mathbf{x_n}^T / \Delta x \tag{42}$$

Bring Eq. (17) into Eq. (42), i.e.,

$$e_{n+1} = e_n \left[1 - \frac{\mu_w(n)\mathbf{g_w^T}(n)\mathbf{g_w}(n)}{e_d(n)e_n} \right] \tag{43}$$

The bound of the learning rate is

$$0 < \mu_w(n) < \frac{2e_d(n)e_n}{\mathbf{g_w^T}(n)\mathbf{g_w}(n)} \tag{44}$$

The learning rate should satisfy

$$0 < \mu_w(n) < \min\left(\frac{2e_d(n)e_n}{\mathbf{g_w^T}(n)\mathbf{g_w}(n)}, \frac{2e_d(n)e_n - \sigma_n \mathbf{g_w^T}(n)\mathbf{m_w}(n-1)}{\mathbf{g_w^T}(n)\mathbf{g_w}(n)} \right) \tag{45}$$

4.2 Convergence Analysis of $\mathbf{q_i}$

The convergence analysis of $\mathbf{q_i}$ is similar to convergence analysis of $\mathbf{w}$. Taking a first-order Taylor series expansion of the error e_{n+1} at instant n, we can get

$$e_{n+1} = e_n + \frac{\partial e_n}{\partial \mathbf{q_i^T}(n)} \Delta \mathbf{q_i}(n) + h.o.t \tag{46}$$

where h.o.t. represents the high order terms of the first-order Tylor series expansion. The second term can be expressed as

$$\frac{\partial e_n}{\partial \mathbf{q_i^T}(n)} = -\frac{\partial y_n}{\partial \mathbf{q_i^T}(n)} = -\mathbf{u_n^T C} \tag{47}$$

$$\Delta \mathbf{q_i}(n) = \mathbf{q_i}(n+1) - \mathbf{q_i}(n) \tag{48}$$

Then, reference to Eqs. (30), (32), (48) can be expressed as

$$\Delta \mathbf{q_i}(n) = \begin{cases} -\mu_q(n)\mathbf{g_q}(n), & \langle \mathbf{m_q}(n-1), \mathbf{g_q}(n) \rangle < 0 \\ -\sigma_n \mathbf{m_q}(n-1) - \mu_q(n)\mathbf{g_q}(n), & \langle \mathbf{m_q}(n-1), \mathbf{g_q}(n) \rangle \geq 0 \end{cases} \tag{49}$$

Consider the case of $\langle \mathbf{m_q}(n-1), \mathbf{g_q}(n) \rangle < 0$, substituting Eqs. (47), (49) and (19) into Eq. (46) and ignoring the higher order terms, then we can get

$$e_{n+1} = e_n \left[1 - \frac{\mu_q(n)\mathbf{g_q^T}(n)\mathbf{g_q}(n)}{e_d(n)e_n} \right] \tag{50}$$

Satisfy the convergence condition $|e_{n+1}| < |e_n|$, obtained

$$\left| 1 - \frac{\mu_q(n)\mathbf{g_q^T}(n)\mathbf{g_q}(n)}{e_d(n)e_n} \right| < 1 \tag{51}$$

The other case is similar to the above $\mathbf{w}$ convergence analysis process, the convergence condition of the learning rate $\mathbf{q_i}$,

$$0 < \mu_q(n) < \min \left(\frac{2e_d(n)e_n}{\mathbf{g_q^T}(n)\mathbf{g_q}(n)}, \frac{2e_d(n)e_n - \sigma_n \mathbf{g_q^T}(n)\mathbf{m_q}(n-1)}{\mathbf{g_q^T}(n)\mathbf{g_q}(n)} \right) \tag{52}$$

5 Numerical Simulations

Several numerical simulations are performed, observing the convergence speed and the steady-state error, to verify the effectiveness of the proposed algorithm. The input signal x_n is generated by the following relationship.

$$x_n = r x_{n-1} + \sqrt{1 - r^2} \xi(n) \tag{53}$$

where $\xi(n)$ is a zero mean white Gaussian noise with unitary variance, and $r \in [0, 1)$ is a correlation factor that determines the correlation between adjacent input signal x_n values.

The algorithm performance is measured by use of mean square error (MSE),

$$\mathrm{MSE} = 10 \lg [e_n]^2 \tag{54}$$

In this paper, the weight vectors of the FIR filter in all SAF type related algorithms are typically initialized to $\mathbf{w}_0 = [1, 0, \dots, 0]^T$. In addition, an independent white

Table 1 Parameters of three kinds of algorithms in simulation

Algorithm	Parameters
SAF–GMVC	$M = 5, r = 0.1, c_0 = 0.99$
SAF–GMVC–MMSGD	$M = 5, r = 0.1, c_0 = 0.99, \sigma_0 = 0.9$
SAF–GMVC–VMSGD	$M = 5, r = 0.1, c_0 = 0.99, \sigma_0 = 0.9, \rho = \lambda = 0.99$

Fig. 5 The profiles curve of nonlinearity using SAF–GMVC–VMSGD (SNR = 30 dB) (Color figure online)
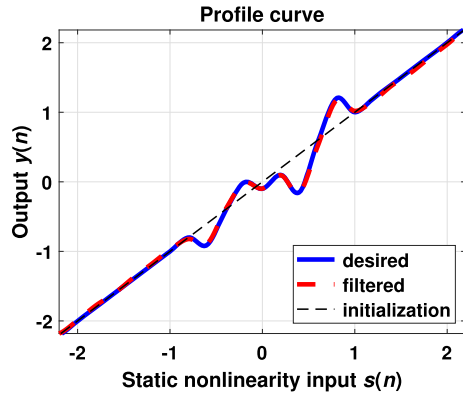

Gaussian noise $v(n)$ with different signal to noise ratio (SNR), that is added to the output of the real unknown system. In the simulation, according to the theoretical boundary of the step size and the simulation experiment value, the step size range of the experiment simulation is obtained $0 < \mu_w < 0.056, 0 < \mu_q < 1.65$. The initialized learning rate $\mu_w = \mu_q = 0.01$. The control points of CR-spline are initialized as a straight line with unitary slope, which is the same as [14] [6] [24]. The following experiment results are obtained by 100 independent Monte Carlo trials.

### 5.1 With White Gaussian Noise

Assuming that the wiener system to be identified consists of a FIR filter with $w^* = [0.6, -0.4, 0.25. - 0.15, 0.1]$, with a uniform interval $\Delta x = 0.2$, a nonlinear spline function interpolated by a 23 knots vector [6]. Set $q^* = [-2.2, -2.0, \ldots, -0.8,$ $-0.91, 0.42, -0.01, -0.1, 0.1, -0.15, \ldots, 2.0, 2.2], N_w = 11, a = 1.6, p = 1.9$. The parameters of all kinds of algorithms are listed in Table 1.

The profiles curve of nonlinearity using the proposed algorithm are shown in Fig. 5. It can be seen that the proposed algorithm (the red dashed line) can accurately converge to profiles of nonlinearity (the solid blue line) of the system to be identified. Figure 6 shows the MSE curves of SAF–GMVC–VMSGD under different correlation factors $r = 0.1, r = 0.3, r = 0.5, r = 0.8, r = 0.9, r = 0.98$. When the input signal under different correlation factors $r = 0.1, r = 0.3, r = 0.5, r = 0.8, r = 0.9$, show the better performance of the convergence speed and the steady-state error. When $r = 0.98$, the steady-state error not better than other factors. Figure 7 the proposed algorithms are compared with each other (SAF–GMVC,
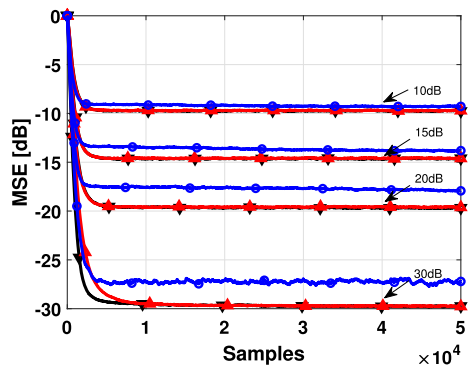
Fig. 6 MSE of SAF–GMVC–VMSGD under different correlation factors r (SNR = 30 dB)
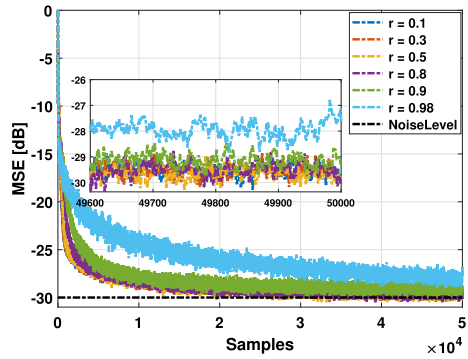
Fig. 7 MSE curves of proposed algorithms compared with each other (SNR = 30 dB)
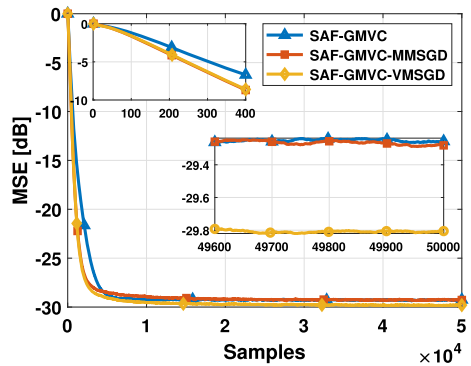
Fig. 8 Compare the three proposed algorithms with the different SNR (Color figure online)

SAF–GMVC–MMSGD, SAF–GMVC–VMSGD). SAF–GMVC–MMSGD compare with SAF–GMVC improved the convergence speed, which much the same steady-state error. SAF–GMVC–VMSGD compare with SAF–GMVC–MMSGD reduced the steady-state error, which have the same convergence speed. The experiment is performed to verify the performance of the proposed algorithms with the different signal-to-noise ratio (SNR) of 30 dB, 20 dB, 15 dB, 10 dB. In Fig. 8 the proposed algorithms are compared with each other under different SNR, $a = 0.32$, $p = 2.5$, the blue curve represents the SAF–GMVC, the red curve represents the SAF–GMVC–
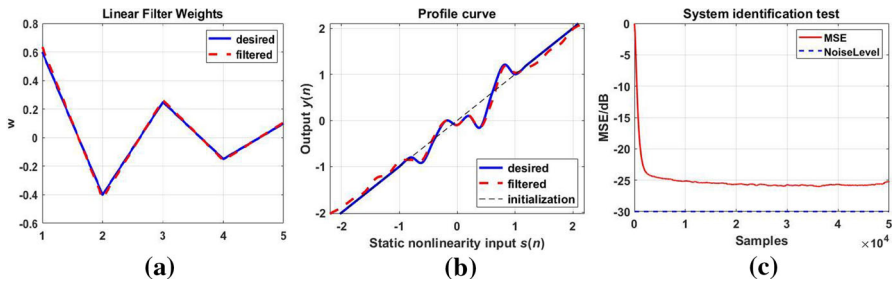
Fig. 9 The simulation of the step size boundary

MMSGD, the black curve represents the SAF–GMVC–VMSGD. With different SNR, the proposed SAF–GMVC–VMSGD algorithms shown the convergence performance better under different SNR, especially SNR = 30 dB. When SNR = 30 dB, $\mu_w = 0.056$, $\mu_q = 1.65$, the simulation results of the step size theoretical boundary are shown in Fig. 9. (a) Represents the tracking of filter weight value. Slight deviation appears in the tracking of the first and second weight values. The filtering of the other weights is consistent with the desired. (b) Represents the tracking of control points. Deviation in the front and back of the control points. The filtering of the middle section control points is consistent with the desired. (c) The system identifies the MSE curve of the test. The convergence value of MSE curve is far from the level line of noise. The tail of the MSE curve shows an upward trend.

5.2 With Non-Gaussian Noise

The experiment is performed to further verify the performance of the proposed algorithm under non-Gaussian noise environment. The experiment include two non-Gaussian noises types, impulse noise and Cauchy noise. The impulse noise is generated by an alpha-stable distribution, with $\alpha \in (0, 2]$ is a characteristic exponent representing the stability index which determines the strength of impulse, $\beta \in [-1, 1]$ is a symmetry parameter, $\iota > 0$ is a dispersion parameter, and ϱ is a location parameter [12]. which can be expressed as

$$f(t) = \exp\left\{j\varrho t - \iota|t|^{\alpha}\left[1 + j\beta\,\text{sign}(t)S(t, \alpha)\right]\right\} \tag{55}$$

where

$$S(t, \alpha) = \begin{cases} \tan\left(\frac{\alpha\pi}{2}\right), & \alpha \neq 1 \\ \left(\frac{2}{\pi}\right)\log|t|, & \alpha = 1 \end{cases} \tag{56}$$

The parameters of alpha-stable distribution are set as follows, $\alpha = 1.6$, $\beta = 0$, $\iota = 0.05$, $\varrho = 0$.

The experimental parameters are the same as those in Table 1, $a = 0.32$, $p = 2.5$, $N_w = 11$, SNR = 30 dB. Figure 10 shows the MSE curves of different algorithms under impulsive noise environment. Specially, the proposed algorithm SAF–GMVC–VMSGD performs better than other algorithms due to the fast convergence speed

Fig. 10 MSE curves of six different algorithms with impulsive noises
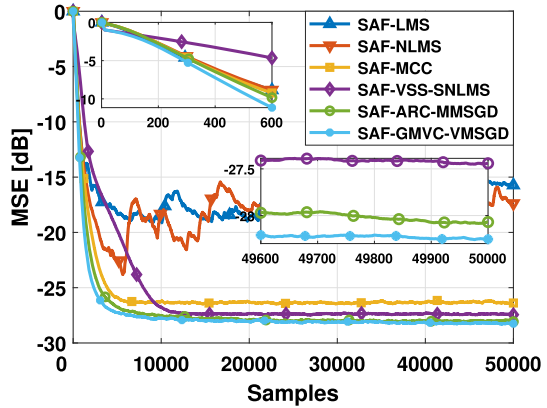
**Fig. 11** MSE curves of SAF–GMVC–VMSGD with different p, a under impulsive noise
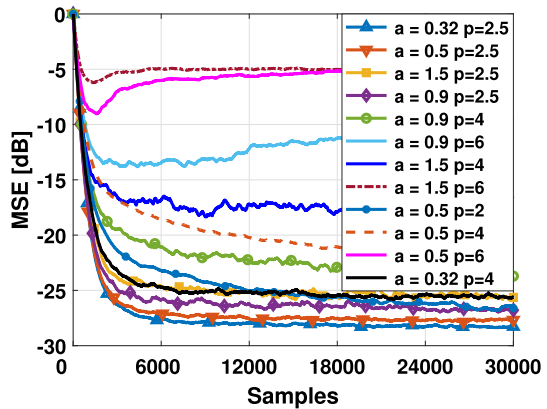
and small steady-state error. Figure 11 MSE curves of SAF–GMVC–VMSGD with different factors p, a under impulsive noise. In the same shape factor $p = 2.5$, which is shown that the larger parameter a, the steady-state error worse. However, $p = 4$, which is shown that the steady-state error not better with the factor a smaller. To sum up, when the parameter p is selected suitable, the steady-state error will vary with factor a. No matter how the steady-state error changes, the convergence speed is faster. In the above experiment simulations, using different parameters also getting the better performance, show the robustness of the nonlinearity identified system.

The Cauchy noise also described by an alpha-stable distribution, with $\alpha = 1$, $\beta = 0$, $r = 0.9$, other parameters are same to the impulsive noise. This experiment result is obtained by 150 independent Monte Carlo trials. Figure 12 shows the MSE curves of different algorithms under Cauchy noise environments. The proposed SAF–GMVC–VMSGD algorithm shows better convergence performance. The experimental results show that the proposed SAF–GMVC–VMSGD algorithm performs well in the above two non-Gaussian noise type.

Fig. 12 MSE curves of proposed algorithms under Cauchy noise environments
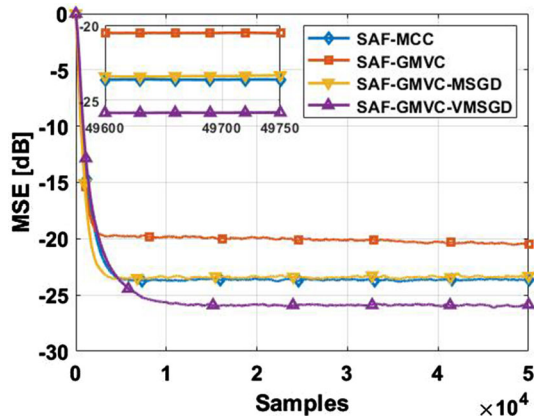
6 Conclusion

This paper proposed a novel Wiener SAF type robust adaptive filtering algorithm, named SAF–GMVC–VMSGD, for nonlinear system identification. The algorithm substitutes a generalized Versoria as cost function, improving the robustness against non-Gaussian noises. Meanwhile, it adopts the variable step-size momentum SGD to further improve the convergence performance. It can be concluded from the numerical simulations that: (1) Compared with the other Wiener SAF type algorithm, the proposed algorithm has the fastest convergence speed and the smallest steady-state error both in the with and without of non-Gaussian noise. (2) The suitable factors lead to stronger robustness in nonlinear identification system. (3) The proposed algorithm not only suitable for Wiener SAF type, but also suitable for other kinds of SAF type. The future work may be devoted to giving the steady-state analysis of the proposed algorithm.

Data Availability The data sets generated during and analyzed during the current study are available from the author on reasonable request.

References

1. P. Bansal, A. Singh, Control of multilevel inverter as shunt active power filter using maximum Versoria criterion. In *2019 International Conference on Power Electronics, Control and Automation (ICPECA)* (IEEE, 2019), pp. 1–6. https://doi.org/10.1109/ICPECA47973.2019.8975392
2. S.S. Bhattacharjee, M.A. Shaikh, K. Kumar, N.V. George, Robust constrained generalized correntropy and maximum Versoria criterion adaptive filters. IEEE Trans. Circuits Syst. II Express Briefs (2021). https://doi.org/10.1109/TCSII.2021.3063491
3. E. Catmull, R. Rom, A class of local interpolating splines. Comput. Aided Geom. Des. **74**, 317–326 (1974). https://doi.org/10.1016/B978-0-12-079050-0.50020-5

4. L. Chang, Z. Zhi, T. Xiao, Sign normalised spline adaptive filtering algorithms against impulsive noise. Signal Process. **148**, 234–240 (2018)
5. X. Chu, L. Zhao, D. Huang, The study of method about diagnosis prediction based on adaptive filtering and hmm. In *Proceedings of the 33rd Chinese Control Conference* (IEEE, 2014), pp. 3229–3232. https://doi.org/10.1109/ChiCC.2014.6895470
6. S. Guan, Z. Li, Normalised spline adaptive filtering algorithm for nonlinear system identification. Neural Process. Lett. **46**(2), 595–607 (2017)
7. S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Inc, Hoboken, 1996)
8. F. Huang, J. Zhang, S. Zhang, Maximum Versoria criterion-based robust adaptive filtering algorithm. IEEE Trans. Circuits Syst. II Express Briefs **64**(10), 1252–1256 (2017)
9. S. Jain, R. Mitra, V. Bhatia, Kernel adaptive filtering based on maximum Versoria criterion. In *2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)* (IEEE, 2018), pp. 1–6. https://doi.org/10.1109/ANTS.2018.8710152
10. C. Liu, Z. Zhang, Set-membership normalised least M-estimate spline adaptive filtering algorithm in impulsive noise. Electron. Lett. **54**(6), 393–395 (2018)
11. Q. Liu, Y. He, Robust Geman–McClure based nonlinear spline adaptive filter against impulsive noise. IEEE Access **8**, 22571–22580 (2020). https://doi.org/10.1109/ACCESS.2020.2969219
12. S. Peng, Z. Wu, X. Zhang, B. Chen, Nonlinear spline adaptive filtering under maximum correntropy criterion. In *TENCON 2015–2015 IEEE Region 10 Conference* (IEEE, 2015), pp. 1–5
13. S. Radhika, F. Albu, A. Chandrasekar, Steady state mean square analysis of standard maximum Versoria criterion based adaptive algorithm. IEEE Trans. Circuits Syst. II Express Briefs **68**(4), 1547–1551 (2020). https://doi.org/10.1109/TCSII.2020.3032089
14. M. Scarpiniti, D. Comminiello, R. Parisi, A. Uncini, Nonlinear spline adaptive filtering. Signal Process. **93**(4), 772–783 (2013)
15. M. Scarpiniti, D. Comminiello, R. Parisi, A. Uncini, Hammerstein uniform cubic spline adaptive filters: learning and convergence properties. Signal Process. **100**, 112–123 (2014)
16. M. Scarpiniti, D. Comminiello, R. Parisi, A. Uncini, Novel cascade spline architectures for the identification of nonlinear systems. IEEE Trans. Circuits Syst. I Regul. Pap. **62**(7), 1825–1835 (2015)
17. M. Scarpiniti, D. Comminiello, R. Parisi, A. Uncini, Spline adaptive filters: theory and applications. In *Adaptive Learning Methods for Nonlinear System Modeling* (Elsevier, 2018), pp. 47–69. https://doi.org/10.1016/B978-0-12-812976-0.00004-X
18. A. Sharma, B.S. Rajpurohit, Maximum Versoria criteria based adaptive filter algorithm for power quality intensification. In *2020 IEEE 9th Power India International Conference (PIICON)* (IEEE, 2020), pp. 1–5. https://doi.org/10.1109/PIICON49524.2020.9112943
19. R.H. Strandberg, M.A. Soderstrand, H.H. Loomis, Elimination of narrow-band interference using adaptive sampling rate notch filters. In *Conference Record of the Twenty-Sixth Asilomar Conference on Signals, Systems Computers* (IEEE, 1992), pp. 861–865. https://doi.org/10.1109/ACSSC.1992.269150
20. E. Szopos, M. Topa, N. Toma, System identification with adaptive algorithms. In *2005 IEEE 7th CAS Symposium on Emerging Technologies: Circuits and Systems for 4G Mobile Wireless Communications* (IEEE, 2005), pp. 64–67. https://doi.org/10.1109/EMRTW.2005.195681
21. W. Wang, H. Zhao, X. Zeng, K. Doğançay, Steady-state performance analysis of nonlinear spline adaptive filter under maximum correntropy criterion. IEEE Trans. Circuits Syst. II Express Briefs **67**(6), 1154–1158 (2019). https://doi.org/10.1109/TCSII.2019.2929536
22. P. Wen, J. Zhang, S. Zhang, B. Qu, Normalized subband spline adaptive filter: algorithm derivation and analysis. Circuits Syst. Signal Process. **40**(5), 2400–2418 (2021)
23. B. Widrow, Adaptive inverse control. In *Adaptive Systems in Control and Signal Processing* (Elsevier, 1987), pp. 1–5
24. L. Yang, J. Liu, R. Yan, X. Chen, Spline adaptive filter with arctangent-momentum strategy for nonlinear system identification. Signal Process. **164**, 99–109 (2019)
25. Z. Zhang, S. Zhang, J. Zhang, Robust weight-constraint decorrelation normalized maximum Versoria algorithm. In *2019 Ninth International Workshop on Signal Design and its Applications in Communications (IWSDA)* (IEEE, 2019), pp. 1–4. https://doi.org/10.1109/IWSDA46143.2019.8966128