SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2013, Vol. 56 048201:1–048201:10
doi: 10.1007/s11432-011-4458-6

Unsupervised robust recursive least-squares algorithm for impulsive noise filtering

CHEN Jie[1,2], MA Tao[1,2*], CHEN WenJie[1,2] & PENG ZhiHong[1,2]

[1]*School of Automation, Beijing Institute of Technology, Beijing* 100081*, China;*
[2]*Key Laboratory of Complex System Intelligent Control and Decision (Beijing Institute of Technology), Ministry of Education, Beijing* 100081*, China*

Received September 14, 2010; accepted May 27, 2011; published online December 2, 2011

Abstract A robust recursive least-squares (RLS) adaptive filter against impulsive noise is proposed for the situation of an unknown desired signal. By minimizing a saturable nonlinear constrained unsupervised cost function instead of the conventional least-squares function, a possible impulse-corrupted signal is prevented from entering the filter's weight updating scheme. Moreover, a multi-step adaptive filter is devised to reconstruct the observed "impulse-free" noisy sequence, and whenever impulsive noise is detected, the impulse contaminated samples are replaced by predictive values. Based on simulation and experimental results, the proposed unsupervised robust recursive least-square adaptive filter performs as well as conventional RLS filters in "impulse-free" circumstances, and is effective in restricting large disturbances such as impulsive noise when the RLS and the more recent unsupervised adaptive filter fails.

Keywords unsupervised adaptive filtering, impulsive noise suppression, recursive least-squares algorithm

Citation Chen J, Ma T, Chen W J, et al. Unsupervised robust recursive least-squares algorithm for impulsive noise filtering. Sci China Inf Sci, 2013, 56: 048201(10), doi: 10.1007/s11432-011-4458-6

1 Introduction

The need to extract signals and other components from a time series is paramount in communication, control, and signal processing problems. An observed noisy sequence is often corrupted by noise that can be divided into two parts: the background noise and impulsive noise. Conventional adaptive filters [1] or robust filters [2,3] are either sensitive to impulsive noise or requiring prior knowledge of the desired signal that, however, sometimes is unavailable.

Impulsive noise occurs relatively infrequently and is non-stationary, and these qualities render impossible the prediction of the statistical properties of impulsive interference [4]. Rarely have models been able to approximate the most important characteristics of impulsive noise [5]. Under adverse conditions, impulsive events can spread over numerous data samples such that detection becomes difficult, and the performance of linear adaptive filters can deteriorate significantly [6]. Nonlinear techniques are often employed to reduce the adverse response from filters. The amplifier limiter is the simplest approach in combating such noise. Nevertheless, these static amplifier limiters can result in poor performance if the signal is non-stationary and changes rapidly [7].

*Corresponding author (email: matao@bit.edu.cn)

Median filtering is applied to the least-mean-squares (LMS) and recursive least-squares (RLS) algorithms, which prevents the filter weights from being affected by the adverse effects of impulsive noise. However, smearing of the original signal can occur [8]. Chan and Zou [4,9,10] proposed a recursive least M-estimate (RLM) algorithm for impulsive noise suppression. Unfortunately, this RLM algorithm is only suitable for system identification schemes in known desired signal conditions. Vega et al. [11–13] introduced a new framework for designing robust adaptive filters by restricting the weight updating process. Nevertheless, that framework also requires the desired signal and can only cope with time-invariant system identification cases.

To pave the way for a study of cases where both system and desired signal are unknown, the unsupervised adaptive filtering [14,15] seems to be useful. However, the probability distribution of the desired signal is required. Universal filtering based on linear least-squares prediction [16], the online gradient descent algorithm [17] and the minimum mean-squares-error (MMSE) [18,19] do not assume any stochastic mechanism of the desired and observed signals. It only makes assumptions about the noise, i.e. the noise is additive zero-mean, time-independent, bounded, and known variance. Nevertheless, these unsupervised form adaptive filters are very sensitive to impulsive noise.

According to the unsupervised adaptive filtering framework in [18], an unsupervised robust recursive least-squares (UR-RLS) filtering algorithm is proposed that renders large perturbations like impulsive noise unresponsive. The UR-RLS is first an extension of the RLS algorithm that is applicable if the desired signal is unavailable for scrutiny. A saturable nonlinear function is introduced to limit the adverse effects of the possible impulsive noise on the RLS weight updating process of the finite impulse response (FIR) filter. The UR-RLS also extends unsupervised adaptive filtering to cases when the input signal is contaminated by possible impulsive noise, for then the impulse-corrupted samples are replaced by predictive values generated from a multi-step adaptive predictor (MAP).

The paper is organized as follows: In Section 2, mathematical models for the impulsive environment and the filtering framework are described. The detailed form of the UR-RLS filter is in Section 3. Section 4 is devoted to the derivation of the MAP. Section 5 presents the simulation and experimental results, and concluding remarks are in Section 6.

2 Problem formulation and filter description

Consider the output of an additive channel

$$y_t = x_t + v_t, \quad t = 1, 2, \ldots, \tag{1}$$

where $\{y_t\}_{t\geqslant 1}$ is the observed noisy signal, and $\{x_t\}_{t\geqslant 1}$ denotes the real-valued clean (desired) signal to be estimated. The additive noise v_t could be divided into two parts: background noise b_t and impulsive noise n_t. The background noise $\{b_t\}_{t\geqslant 1}$ is assumed to be independent over t, and with a Gaussian distribution, i.e. $b_t \sim N(0, \sigma_b^2)$ for arbitrary t. Additionally, assume that $|x_t| \leqslant B_X$ for some positive real number $B_X < \infty$. Since $\{b_t\}_{t\geqslant 1}$ is assumed Gaussian, for given positive real number $B_B = 3\sigma_b^2$, $|b_t| \leqslant B_B$ exists with probability of 99.74%. Thus, the noise v_t can be expressed as $v_t = b_t + n_t$.

The structure of the proposed UR-RLS filter as shown in Figure 1 is composed of three parts:

(a) a FIR filter: a linear transversal filter with order d.

(b) a weight updating algorithm: unsupervised weight updating scheme of the FIR filter.

(c) a predictor: estimate $\hat{y}_{t+k|t} (k \geqslant 1)$ by using observed noisy sequence $\{y_i\}_{i=1}^{t}$ and the estimated $\{\hat{y}_{t+i|t}\}_{i=1}^{k-1}$.

Remark 2.1. Comparing the predictor's output $\hat{y}_{t|t-1}$ with the observed y_t, we take the following algorithm:

$$\hat{y}_t = \begin{cases} y_t, & |y_t - \hat{y}_{t|t-1}| \leqslant \xi, \\ \hat{y}_{t|t-1}, & \text{otherwise.} \end{cases} \tag{2}$$

If $|\hat{y}_{t|t-1} - y_t|$ is greater than a certain threshold ξ, then the input to the FIR filter y_t will be replaced by $\hat{y}_{t|t-1}$ to avoid possibly intrusion of the impulse. A MAP is developed in case the impulse lasts more

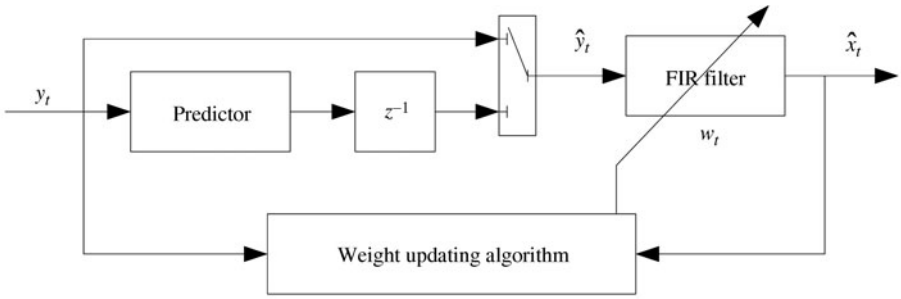

Figure 1 Scheme of the UR-RLS filter.

than one sampling period. A fixed ξ results in a tradeoff between noise sensitivity and signal reliability. If ξ is too small, the filter will behave well against impulsive noise, but those effective signals can also be eliminated. If ξ is large, some impulses might get into the filter. Consequently, a time varying $\{\xi_t\}$ sequence depending on the confidence level of the predictor is used.

3 Unsupervised weight updating algorithm

Let $\boldsymbol{w}_t = [w_t, w_{t-1}, \ldots, w_{t-(d-1)}]^{\mathrm{T}} \in \mathbb{R}^d$, and $\boldsymbol{y}_t = [y_t, y_{t-1}, \ldots, y_{t-(d-1)}]^{\mathrm{T}} \in \mathbb{R}^d$ be, respectively, the tap-weight vector and tap-input vector at time t, where $(\cdot)^{\mathrm{T}}$ is the usual transposition operator. Zeros are assigned to those elements of the vectors for which the indices are less than or equal to zero. In addition, we have $\boldsymbol{c} = [\sigma_b^2, 0, \ldots, 0]^{\mathrm{T}} \in \mathbb{R}^d$ is the background noise vector, $\|\cdot\|$ denoting the Euclidean norm for vectors, $|\cdot|$ the absolute value, and $(\cdot)^*$ signifying the optimal value.

Definition 3.1. For any $i, t \geqslant 1$ and $\boldsymbol{w}_t \in \mathbb{R}^d$, let $l_i(\boldsymbol{w}_t) \triangleq (y_i - \boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{y}_i)^2 + 2\boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{c}$. According to the cost function of conventional RLS [1], define the cost function of a UR-RLS as

$$L_\rho(\boldsymbol{w}_t) \triangleq \sum_{i=1}^{t} \lambda^{t-i}\rho(l_i(\boldsymbol{w}_t)) + \delta\lambda^t\|\boldsymbol{w}_t\|^2, \tag{3}$$

where the forgetting factor λ is a positive constant close to, but less than, unity; δ is a positive real number called the regularization parameter, and $\rho(\cdot)$ is a saturable nonlinear function. The term $\delta\lambda^t\|\boldsymbol{w}_t\|^2$ is the regularizing term which stabilizes the solution to the RLS problem by its smoothing effect.

Remark 3.2. The purpose of using a saturable nonlinear function $\rho(\cdot)$, instead of the one in [18], is to limit the adverse effect of the impulsive noise on the weight updating algorithm.

Figure 2 depicts the following nonlinear function:

$$\rho(l) = \begin{cases} l, & |l| < \zeta, \\ \mathrm{sign}(l)\zeta, & \text{otherwise}, \end{cases} \tag{4}$$

where ζ is the threshold parameter. A $|l|$ value smaller than ζ reduces the cost function to that for the universal FIR MMSE adaptive filtering (UAF) [18]. If $|l| \geqslant \zeta$, $\rho(\cdot)$ is equal to a constant which helps to suppress large perturbations. It can be seen that small value of ζ results in a good suppression of impulses, but values of ζ too small would also restrict the use of those "impulse-free" l. Alternatively, as for the analysis in Remark 2.1, filtering with large l does not behave well against the impact of impulses. In this paper, the threshold parameter ζ is continuously chosen according to the l_t sequence. This will be discussed later in this section. From the above discussion, it is apparent that the objective function $L_\rho(\boldsymbol{w}_t)$ defined in (3) can be used to smooth out the adverse effects of impulsive noise.

By minimizing the cost function $L_\rho(\boldsymbol{w}_t)$ in (3) with respect to $\boldsymbol{w}_t$, we get the following normal equation:

$$\boldsymbol{\Phi}_t\boldsymbol{w}_t^* = \boldsymbol{P}_t, \tag{5}$$

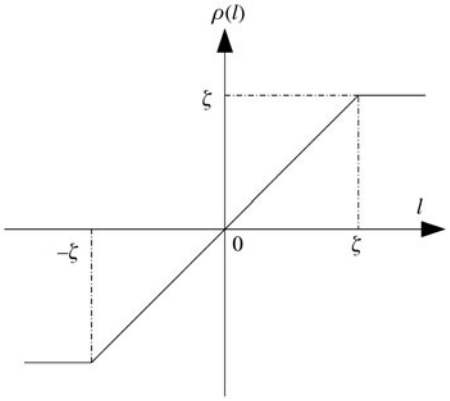

Figure 2 Saturable nonlinear function.

where

$$\boldsymbol{\Phi}_t = \delta\lambda^t \boldsymbol{I} + \sum_{i=1}^{t} \lambda^{t-i}\psi(l_i)\boldsymbol{y}_i\boldsymbol{y}_i^{\mathrm{T}} = \lambda\boldsymbol{\Phi}_{t-1} + \psi(l_t)\boldsymbol{y}_t\boldsymbol{y}_t^{\mathrm{T}}, \tag{6}$$

$$\boldsymbol{P}_t = \sum_{i=1}^{t} \lambda^{t-i}\psi(l_i)(\boldsymbol{y}_i\boldsymbol{y}_i - \boldsymbol{c}) = \lambda\boldsymbol{P}_{t-1} + \psi(l_t)(\boldsymbol{y}_t\boldsymbol{y}_t - \boldsymbol{c}), \tag{7}$$

in which $\psi(l) \triangleq \mathrm{d}\rho(l)/\mathrm{d}l$. Applying the matrix inversion lemma [1], the following RLS estimate algorithm is derived that solves the normal equation in (5):

$$\boldsymbol{\Phi}_t^{-1} = \lambda^{-1}(\boldsymbol{I} - K_t\boldsymbol{y}_t^{\mathrm{T}})\boldsymbol{\Phi}_{t-1}^{-1}, \tag{8}$$

$$K_t = \frac{\psi(l_t)\boldsymbol{\Phi}_{t-1}^{-1}\boldsymbol{y}_t}{\lambda + \psi(l_t)\boldsymbol{y}_t^{\mathrm{T}}\boldsymbol{\Phi}_{t-1}^{-1}\boldsymbol{y}_t}, \tag{9}$$

$$\boldsymbol{w}_t^* = \boldsymbol{\Phi}_t^{-1}\boldsymbol{P}_t = \boldsymbol{w}_{t-1}^* + \psi(l_t)\boldsymbol{\Phi}_t^{-1}[(\boldsymbol{y}_t - \boldsymbol{w}_{t-1}^{*\mathrm{T}}\boldsymbol{y}_t)\boldsymbol{y}_t - \boldsymbol{c}], \tag{10}$$

where K_t is the gain vector. If $|l_t| < \zeta$, we see that $\psi(l_t) = 1$, and eq. (10) becomes identical to the UAF, if λ and δ are equal to one. However, when $|l_t| \geqslant \zeta$ then $\psi(l_t) = 0$, and from (6) and (7), the impulse corrupted signal y_t does not appear in the expressions for $\boldsymbol{\Phi}_t$ and $\boldsymbol{P}_t$. At the same time, updating the weight vector is not possible. These properties make the proposed UR-RLS algorithm more robust in the UAF. Choices of forgetting factor λ and regularization parameter δ are identical to the conventional RLS algorithm [1]. For simplicity, λ is assumed to be 0.99, and δ is a large (small) positive constant for low (high) signal-to-noise ratio (SNR) circumstances.

Remark 3.3. As mentioned in Remark 3.2, the choice of the threshold parameter ζ can significantly affect the performance of the algorithm. Although the distribution of l_t is in general unknown, the threshold ζ in (4) can be chosen as

$$\zeta = k_\zeta \hat{\sigma}_{l_t}, \tag{11}$$

where k_ζ is the confidence factor, and $\hat{\sigma}_{l_t}$ is the estimated variance of "impulse-free" l_t. According to the experimental results, k_ζ can be chosen to be 3.0, which results in a 99% confidence in detecting and rejecting the impulse. We take the following formula to estimate $\hat{\sigma}_{l_t}$ [8]:

$$\hat{\sigma}_{l_t}^2 = \lambda_\sigma \hat{\sigma}_{l_{t-1}}^2 + (1 - \lambda_\sigma)c_1 \mathrm{med}(A_\rho(l_t)), \tag{12}$$

where $A_\rho(l_t) = \{\rho(l_t), \ldots, \rho(l_{t-N_l+1})\}$, $\mathrm{med}(\cdot)$ denotes the sample median operation, N_l is the length of the estimation window, λ_σ the forgetting factor, and $c_1 = 1.483(1 + 5/(N_l - 1))$ a finite sample correction factor [20]. Also $\hat{\sigma}_{l_0}^2$ is a relatively large value relevant to the bound of the observed "impulse-free" noisy signal y_t. According to our experimental results, $\hat{\sigma}_{l_0}^2 = 3(B_X + B_B)$ is big enough. Moreover, the forgetting factor could be chosen as $\lambda_\sigma = 0.99$.

4 Processing of input signal

In regard to the FIR filter outlined in Figure 1, the UR-RLS filter becomes

$$\hat{x}_t(\boldsymbol{y}^t) = \boldsymbol{w}_t^{*\mathrm{T}}\boldsymbol{y}_t. \tag{13}$$

It is not difficult to notice that although the weight updating algorithm of $\boldsymbol{w}_t^*$ in (10) is insensitive to large perturbations, the impulse corrupted $\boldsymbol{y}_t$ still degrades the performance of the filter. In the following part, a MAP is developed to restrict the adverse effects from a possible impulsive noise when filtering the input vector $\boldsymbol{y}_t$.

Definition 4.1. For any $\boldsymbol{u} \in \mathbb{R}^d$, define the cost function of k-step ($k \geqslant 1$) MAP at time t,

$$J_{\boldsymbol{u}}(t,k) \triangleq \sum_{i=1}^{t-k} \beta^{t-k-i}(\boldsymbol{y}_{i+k} - \boldsymbol{u}_{t,k}^{\mathrm{T}}\boldsymbol{y}_i)^2 + \alpha\beta^{t-k}\|\boldsymbol{u}_{t,k}\|^2, \tag{14}$$

where $\boldsymbol{u}_{t,k}$ denotes the weight vector of MAP, and as in Definition 3.1, $0 < \alpha < \infty$ and $0 < \beta \leqslant 1$ are the regularization parameter and the forgetting factor, respectively.

By minimizing the cost function $J_{\boldsymbol{u}}(t,k)$ with respect to $\boldsymbol{u}_{t,k}$, we have

$$\boldsymbol{\Psi}_{t,k}\boldsymbol{u}_{t,k}^* = \boldsymbol{Q}_{t,k}, \tag{15}$$

where

$$\boldsymbol{\Psi}_{t,k} = \alpha\beta^{t-k}\boldsymbol{I} + \sum_{i=1}^{t-k}\beta^{t-k-i}\boldsymbol{y}_i\boldsymbol{y}_i^{\mathrm{T}} = \beta\boldsymbol{\Psi}_{t,k-1} + \boldsymbol{y}_{t-k}\boldsymbol{y}_{t-k}^{\mathrm{T}}, \tag{16}$$

$$\boldsymbol{Q}_{t,k} = \sum_{i=1}^{t-k}\beta^{t-k-i}\boldsymbol{y}_{i+k}\boldsymbol{y}_i = \beta\boldsymbol{Q}_{t,k-1} + \boldsymbol{y}_t\boldsymbol{y}_{t-k}. \tag{17}$$

Thus, from the matrix inverse lemma, the following RLS algorithm is derived by solving (15):

$$\boldsymbol{\Psi}_{t,k}^{-1} = \beta^{-1}(\boldsymbol{I} - K_{t,k}\boldsymbol{y}_{t-k}^{\mathrm{T}})\boldsymbol{\Psi}_{t,k-1}^{-1}, \tag{18}$$

$$K_{t,k} = \frac{\beta^{-1}\boldsymbol{\Psi}_{t,k-1}^{-1}\boldsymbol{y}_{t-k}}{1 + \beta^{-1}\boldsymbol{y}_{t-k}^{\mathrm{T}}\boldsymbol{\Psi}_{t,k-1}^{-1}\boldsymbol{y}_{t-k}}, \tag{19}$$

$$\boldsymbol{u}_{t,k}^* = \boldsymbol{\Psi}_{t,k}^{-1}\boldsymbol{Q}_{t,k} = \boldsymbol{u}_{t,k-1}^* + (\boldsymbol{y}_t - \boldsymbol{u}_{t,k-1}^{*\mathrm{T}}\boldsymbol{x}_{t-k})K_{t,k}, \tag{20}$$

where $K_{t,k}$ is the gain vector. Consequently, the output of the k-step ($k \geqslant 1$) MAP is

$$\hat{\boldsymbol{y}}_{t+k|t} = \boldsymbol{u}_{t,k}^{*\mathrm{T}}\boldsymbol{y}_t. \tag{21}$$

Assume that the "impulse-free" time-series y_t could be described by an autoregressive model with order p (AR(p)):

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t, \tag{22}$$

where ε_t is independent of time, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, and $p \leqslant d$. Denote $\boldsymbol{a} = [a_1 \cdots a_p\, 0 \cdots 0]^{\mathrm{T}} \in \mathbb{R}^d$, then $\boldsymbol{y}_t = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t$.

For any $\boldsymbol{a} \in \mathbb{R}^d$, define

$$\Gamma(\boldsymbol{a}) = \begin{bmatrix} \boldsymbol{0} & 0 \\ \boldsymbol{I}_{d-1} & \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{a}^{\mathrm{T}} \\ \boldsymbol{0} \\ \vdots \end{bmatrix} \in \mathbb{R}^{d\times d},$$

where $\boldsymbol{I}_{d-1}$ is the identity matrix. Therefore,

$$\boldsymbol{y}_{t+k} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{y}_{t+k-1} + \boldsymbol{\varepsilon}_{t+k} = \boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}\boldsymbol{y}_t + \boldsymbol{a}^{\mathrm{T}}\underbrace{\sum_{i=0}^{k-2}[\boldsymbol{\Gamma}(\boldsymbol{a})]^i\begin{bmatrix}\varepsilon_{t+k-1-i}\\ 0\\ \vdots\end{bmatrix}}_{\triangleq\Xi_{k-2}} + \boldsymbol{\varepsilon}_{t+k}. \tag{23}$$

Theorem 4.2. Given the MAP in (21), then, the prediction error variance becomes

$$\begin{aligned}&E[(\boldsymbol{y}_{t+k}-\hat{\boldsymbol{y}}_{t+k|t})^{\mathrm{T}}(\boldsymbol{y}_{t+k}-\hat{\boldsymbol{y}}_{t+k|t})]\\&=\boldsymbol{y}_t^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}-\boldsymbol{u}_{t,k}^{*\mathrm{T}})^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}-\boldsymbol{u}_{t,k}^{*\mathrm{T}})\boldsymbol{y}_t\\&\quad+\boldsymbol{a}^{\mathrm{T}}\sum_{i=0}^{k-2}[\boldsymbol{\Gamma}(\boldsymbol{a})]^i\begin{bmatrix}\sigma_\varepsilon^2&0&\cdots\\0&0&\cdots\\\vdots&\vdots&\ddots\end{bmatrix}([\boldsymbol{\Gamma}(\boldsymbol{a})]^i)^{\mathrm{T}}\boldsymbol{a}+\sigma_\varepsilon^2.\end{aligned}\tag{24}$$

Proof. Since ε_t is white noise with variance σ_ε^2, the prediction error variance becomes

$$\begin{aligned}&E[(\boldsymbol{y}_{t+k}-\hat{\boldsymbol{y}}_{t+k|t})^{\mathrm{T}}(\boldsymbol{y}_{t+k}-\hat{\boldsymbol{y}}_{t+k|t})]\\&=E[(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{y}_{t+k-1}+\boldsymbol{\varepsilon}_{t+k}-\boldsymbol{u}_{t,k}^{*\mathrm{T}}\boldsymbol{y}_t)^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{y}_{t+k-1}+\boldsymbol{\varepsilon}_{t+k}-\boldsymbol{u}_{t,k}^{*\mathrm{T}}\boldsymbol{y}_t)]\\&=E[(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}\boldsymbol{y}_t+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Xi}_{k-2}+\boldsymbol{\varepsilon}_{t+k}-\boldsymbol{u}_{t,k}^{*\mathrm{T}}\boldsymbol{y}_t)^{\mathrm{T}}\cdot(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}\boldsymbol{y}_t+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Xi}_{k-2}+\boldsymbol{\varepsilon}_{t+k}-\boldsymbol{u}_{t,k}^{*\mathrm{T}}\boldsymbol{y}_t)]\\&=E[\boldsymbol{y}_t^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}-\boldsymbol{u}_{t,k}^{*\mathrm{T}})^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}-\boldsymbol{u}_{t,k}^{*\mathrm{T}})\boldsymbol{y}_t]\\&\quad+E[(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Xi}_{k-2})^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Xi}_{k-2})]+E[\boldsymbol{\varepsilon}_{t+k}^{\mathrm{T}}\boldsymbol{\varepsilon}_{t+k}]\\&=\boldsymbol{y}_t^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}-\boldsymbol{u}_{t,k}^{*\mathrm{T}})^{\mathrm{T}}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{\Gamma}(\boldsymbol{a})]^{k-1}-\boldsymbol{u}_{t,k}^{*\mathrm{T}})\boldsymbol{y}_t\\&\quad+\boldsymbol{a}^{\mathrm{T}}\sum_{i=0}^{k-2}[\boldsymbol{\Gamma}(\boldsymbol{a})]^i\begin{bmatrix}\sigma_\varepsilon^2&0&\cdots\\0&0&\cdots\\\vdots&\vdots&\ddots\end{bmatrix}([\boldsymbol{\Gamma}(\boldsymbol{a})]^i)^{\mathrm{T}}\boldsymbol{a}+\sigma_\varepsilon^2.\end{aligned}$$

Remark 4.3. From Theorem 4.2, as the prediction length k increases, the error accumulates. Usually, the a prior predictor ($k=1$) is used to generate the comparison sequence with the observed samples. The multi-step form of the predictor ($k>1$) works only if the corrupted impulse lasts more than one sampling period.

As mentioned in Remark 2.1, a time-varying ξ_t determines whether the observed signal is corrupted by impulsive noise or not. Given $\hat{y}_{t|t-1}$ from the MAP, the "impulse-free" estimation error becomes $e_t=y_t-\hat{y}_{t|t-1}$. More precisely, the error signal is assumed, for simplicity, to be a Gaussian- distributed random process. Suppose that the variance of e_t can be estimated, then, the probability $\Pr\{|e_t|>\xi\}=\theta_\xi=1-\mathrm{erfc}\{\xi/(\sqrt{2}\hat{\sigma}_{\xi_t}^2)\}$, where $\mathrm{erfc}(r)=\frac{2}{\sqrt{\pi}}\int_0^r \mathrm{e}^{-x^2}\mathrm{d}x$ is the complementary error function [9], because of the Gaussian assumption of e_t. By choosing different values of θ_ξ, we have a different confidence in detecting the impulsive noise that appears at the input signal. For example, if θ_ξ is set to 0.01, we have 99% confidence to detect and reject the impulse. Therefore, the threshold ξ in (2) can be chosen as [10]

$$\xi=k_\xi\hat{\sigma}_{\xi_t}=2.576\hat{\sigma}_{\xi_t},\tag{25}$$

where $k_\xi=2.576$, and $\hat{\sigma}_{\xi_t}$ is the estimated variance of the "impulse-free" estimation error. To estimate $\hat{\sigma}_{\xi_t}$, the following formula can be used

$$\hat{\sigma}_{\xi_t}^2=\lambda_\xi\hat{\sigma}_{\xi_{t-1}}^2+(1-\lambda_\xi)c_1\mathrm{med}(A(e_t)),\tag{26}$$

where $A(e_t)=\{e_t^2,\ldots,e_{t-N_e+1}^2\}$, N_e is the length of the estimation window, λ_ξ is the forgetting factor, and $\mathrm{med}(\cdot)$ and c_1 are the same as that in (12). Similar to λ_ζ, the forgetting factor λ_ξ could be chosen as 0.99. Since the MAP could not get a good estimation of $\hat{y}_t$ at the very beginning, the initial value of $\hat{\sigma}_{\xi_0}$ is set to be $3(B_X+B_B)$.

5 Simulation and experimental results

5.1 Linear, stochastic signal

Consider the AR(1) clean signal [18]:

$$x_t=\gamma x_{t-1}+z_t,\quad \gamma=0.9,\quad t=1,2,\ldots,\tag{27}$$

where $\{z_t\}_{t\geqslant 1}$ is independent and identically distributed (i.i.d.), $z_t \sim \mathcal{N}(0,1)$, and $x_0 \sim \mathcal{N}(0, \frac{1}{1-\gamma^2})$ to assure the stationarity of $\{x_t\}_{t\geqslant 1}$. The noisy signal $\{y_t\}_{t\geqslant 1}$ is obtained by passing the clean signal through the additive channel (1), where the background noise $\{b_t\}_{t\geqslant 1}$ is a Gaussian sequence, independent of $\{x_t\}_{t\geqslant 1}$. If σ_x^2 and σ_b^2 are the power of the clean signal and the background noise, respectively, then the signal-to-background-noise ratio (SBNR) is defined as

$$\text{SBNR} = 10 \lg \left[\frac{\sigma_x^2}{\sigma_b^2}\right]. \tag{28}$$

In the simulation, σ_b^2 is chosen in such a way that SBNR=10 or SBNR=-5 dB. The measure of performance considered is the mean-square-error (MSE) of the estimation error in dB form, i.e., 10lg(MSE).

The impulsive noise, n_t, can be modeled as $n_t = p_t i_t$, where the impulse generating process i_t is a random process representing an ever-present impulse component with variance σ_i^2 and p_t is a switching sequence of ones and zeros. If p_t is a one (zero) then there is (is not) an impulse present in the t-th sample. To model an i.i.d. impulse noise the switching sequence, p_t, is typically chosen to be a Bernoulli random process with $P(p_t = 1) = \epsilon$.

Unless otherwise specified, the following parameters will be used: adaptive FIR filter order $d = 5$; sequence length $n = 10000$; the forgetting factors $\lambda = 0.99$, $\beta = 0.99$, $\lambda_\sigma = 0.99$, $\lambda_\xi = 0.99$; regularization parameters $\delta = 1$, $\alpha = 1$ (10 dB) or $\delta = 100$, $\alpha = 100$ (-5 dB); impulse variance $\sigma_i^2 = 300$, $\epsilon = 0.1$; the threshold proportionality factor $k_\zeta = 3.0$, $k_\xi = 2.576$, and the length of variance estimation windows $N_l = 10$, $N_e = 10$. We take the conventional RLS algorithm [1] and the UAF [18] as comparisons.

Figure 3 shows the average MSE results of 50 independent experiments in both median SBNR (10 dB) and low SBNR (-5 dB). Impulses occur as a random Bernoulli process as described above between $n = 4001$ and $n = 4500$. From Figure 3(a) and Figure 3(b), it is not difficult to see that the proposed UR-RLS has the same performance as UAF, which has been shown in (10), when there is no impulsive noise. Similar to the conclusions in [18], the unsupervised adaptive filters, UAF and UR-RLS, perform as well as the conventional RLS transversal filter after a learning process during "impulse-free" circumstances. However, the impulsive noise deteriorates the performance for both RLS and UAF, whereas only a slight disturbance occurs for UR-RLS.

The UAF has the worst performance under impulsive circumstances, as the impulses affect both the weight updating algorithm and the filtering process. For the uncorrupted desired signal, x_t partially helps suppress the impulsive noise in the weight updating process; hence the RLS FIR filter performs much better than UAF. With the help of (2) and (4), the adverse effects of impulses are partially eliminated from the UR-RLS.

5.2 Nonlinear, stochastic signal

Consider the following nonlinear clean signal [18]:

$$x_t = 0.1x_{t-1} - 0.5\cos(3x_{t-1}) + 0.4\sin(x_{t-2}) + 0.1x_{t-2} + z_t, \quad t = 2, 3, \ldots \tag{29}$$

and parameter values and testing conditions are the same as that in Subsection 5.1. Again, the conventional RLS algorithm and the UAF are used as comparisons.

Figure 4 shows the average MSE results of 50 independent experiments of conventional RLS, UAF and the proposed UR-RLS algorithm for the nonlinear signal (29). The median SBNR condition (10 dB) in Figure 4(a) again gives the similar results as for the AR(1) signal in Figure 3(a) as well as the low SBNR case (-5 dB) in Figures 3(b) and 4(b).

5.3 Choice of the thresholds

This experiment evaluates an identical estimation of thresholds in (2) and (4). The signal and testing conditions are chosen to be the same as that in Subsection 5.2, and SBNR=-5 dB.

Figure 5(a) presents the estimated threshold of the saturable nonlinear function in (4). It is easy to see that the threshold is big enough to let those "impulse-free" l_t pass through the nonlinear channel.

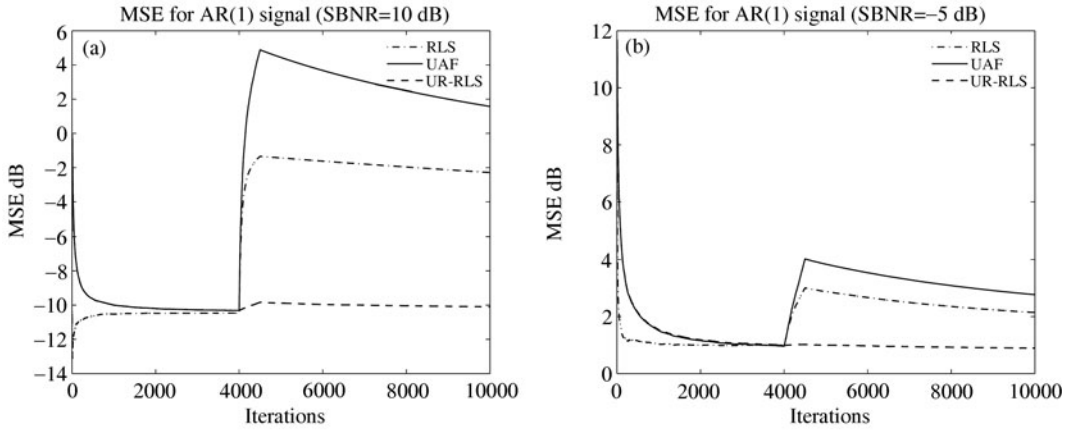

Figure 3 MSE results of the various algorithms in the presence of the impulses for AR(1) signal: RLS, UAF, and UR-RLS. Impulses in the observed noisy signal appear between $n = 4001$ and $n = 4500$.

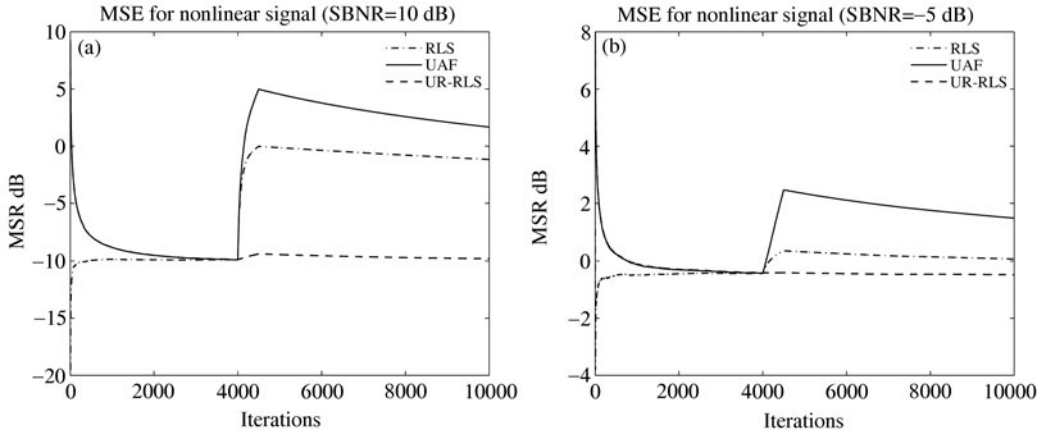

Figure 4 MSE results of the various algorithms in the presence of the impulses for nonlinear signal: RLS, UAF, and UR-RLS. Impulses in the observed noisy signal appear between $n = 4001$ and $n = 4500$.

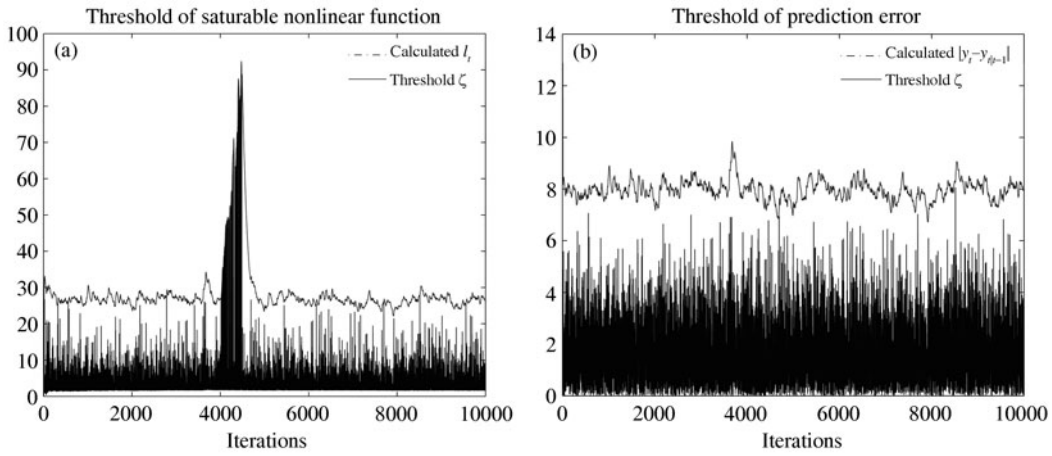

Figure 5 Thresholds estimation of saturable nonlinear function and prediction error ($k_\zeta = 3.0$, $k_\xi = 2.576$).

When there is impulsive noise, the threshold restricts the impulse corrupted signal extremely well, ensuring that those impulses will not affect the weight updating process of the FIR filter. Figure 5(b) gives the prediction error bound. Assumed to be a Gaussian process, the "impulse-free" input signal does not get affected by the predicted $\hat{y}_{t|t-1}$ sequence with high probability. Since those impulse corrupted samples are replaced by predicted signals, the FIR filter input vector also is not contaminated by the impulses.

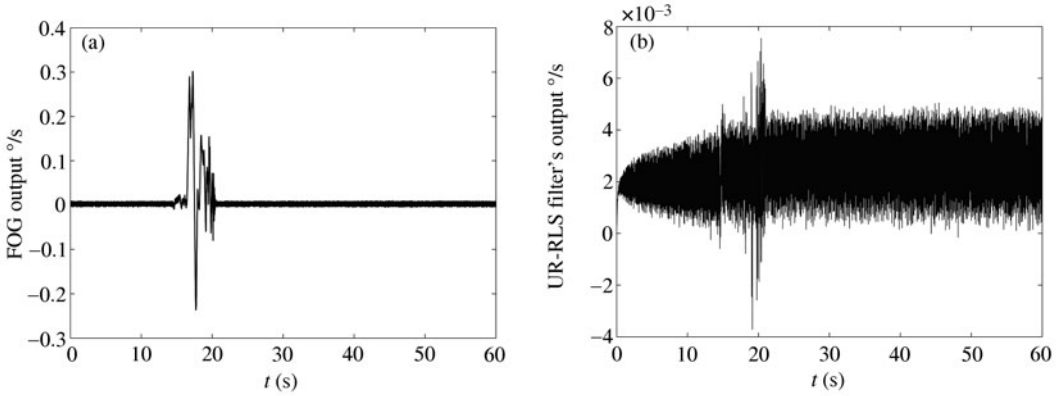

Figure 6 Initial output of FOG in a north finder in adverse environment and the UR-RLS filter's output. Notice the longitudinal axis of (b) has been greatly expanded compared with (a).

5.4 De-noising of fiber optic gyroscope signal

A fiber optic gyroscope (FOG) is used to sense the rotation of the earth, which is less than 0.0042°/s, in a four position north finder (used in seeking the direction of north). The angle random walk (ARW) is $0.005°/\sqrt{h}$, sampling period T=10 ms, testing time is 60 s at each position, and the FIR filter order $d = 5$. The SBNR is estimated according to the ARW value. Other initial values are the same as that in Subsection 5.1.

The north finder is placed on a tripod, and the true value of the north is 1369.2 mil (82.152°). The total sampling time scale is 60 s at each position; that is, it takes about 4 minutes to get the north value. Figure 6(a) shows the initial output of the FOG in the north finder at one position, where someone has passed by when the FOG had been working. Human's activities create large disturbance compared with the measured valued of the earth's rotation. The SBNR is approximately −3.06 dB. Using this signal to calculate the north value, we get a value of 1298.6 mil (77.916°) which causes a misalignment of 70.6 mil. Figure 6(b) presents the output of the UR-RLS adaptive filter. It is not difficult to see that the large disturbance is almost eliminated from the initial FOG output, and the noise level is reduced, i.e., SBNR=6.13 dB. This impulse-eliminated signal results in an estimation of 1368.1 mil (82.086°) with a misalignment of only 1.1 mil.

6 Conclusion

An unsupervised robust adaptive filter by a least-squares means has been devised to implement impulsive noise suppression. The weight updating algorithm follows from the optimization of a certain cost function subject to a time-dependent saturable nonlinear function constraint on the mismatch of the estimation. Within the saturable nonlinear constraint, the adverse effect of impulsive noise gets partially eliminated from the weight updating process. A predictor has been introduced to replace possible impulse-contaminated samples of the input to the filter. A multi-step form of the predictor has also been developed for situations when the impulses last more than one sampling period. Simulation and experimental results showed that the UR-RLS adaptive filter performs as well as the conventional RLS adaptive filter in "impulse-free" circumstances without requiring prior knowledge of the desired signals rate of convergence and precision. The results also provided evidence of the proposed algorithm's good performance and robustness against impulsive noise.

Acknowledgements

This work was supported by National Science Fund for Distinguished Young Scholars of China (Grant No. 60925011).

References

1 Haykin S. Adaptive Filter Theory. 4th ed. New Jersey: Prentice Hall, 2002
2 Sun J, Chen J, Liu G P, et al. Delay-dependent robust H_∞ filter design for uncertain linear systems with time-varying delay. Circ Syst Signal Pr, 2009, 28: 763–779
3 Shi P, Mahmoud M, Nguang S K, et al. Robust filtering for jumping systems with mode-dependent delays. Signal Process, 2006, 86: 140–152
4 Chan S C, Zou Y. A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: fast algorithm and convergence performance analysis. IEEE Trans Signal Proces, 2004, 52: 975–991
5 Xia T, Wan Q, Wang X, et al. 2-D DOAs estimation in impulsive noise environments using joint diagonalization fractional lower-order spatio-temporal matrices. Sci China Ser F-Inf Sci, 2008, 51: 1585–1593
6 Weng B, Kenneth E B. Nonlinear system identification in impulsive environments. IEEE Trans Signal Proces, 2005, 53: 2588–2594
7 Seong R K, Efron A. Adaptive robust impulse noise filtering. IEEE Trans Signal Proces, 1995, 43: 1855–1866
8 Jayant N. Average- and median-based smoothing techniques for improving digital speech quality in the presence of transmission errors. IEEE Trans Commun, 1976, 24: 1043–1045
9 Zou Y, Chan S C, Ng T S. A recursive least M-estimate (RLM) adaptive filter for robust filtering in impulsive noise. IEEE Signal Proc Let, 2000, 7: 324–326
10 Zou Y, Chan S C, Ng T S. Least mean M-estimate algorithms for robust adaptive filtering in impulse noise. IEEE Trans Circuits II, 2000, 47: 1564–1569
11 Vega L R, Rey H, Benesty J, et al. A new robust variable step-size NLMS algorithm. IEEE Trans Signal Proces, 2008, 56: 1878–1893
12 Vega L R, Rey H, Benesty J, et al. A family of robust algorithms exploiting sparsity in adaptive filters. IEEE Trans Audi Speech Lang Pr, 2009, 17: 572–581
13 Vega L R, Rey H, Benesty J, et al. A fast robust recursive least-squares algorithm. IEEE Trans Signal Proces, 2009, 57: 1209–1216
14 Awate S P, Whitaker R T. Unsupervised, information-theoretic, adaptive image filtering for image restoration. IEEE Trans Pattern Anal, 2006, 28: 364–376
15 Makov U. Approximations to unsupervised filters. IEEE Trans Automat Contr, 1980, 25: 842–847
16 Zeitler G C, Singer A C. Universal linear least-squares prediction in the presence of noise. In: Proceedings of IEEE/SP 14th Workshop on Statistic Signal Process. New York: IEEE Press, 2007. 611–614
17 Zinkevich M. Online convex programming and generalized infinitesimal gradient ascent. In: Lane T, ed. Proceedings of the 20th International Conference on Machine Learning. Menlo Park: AAAI Press, 2003. 928–936
18 Moon T, Weissman T. Universal FIR MMSE filtering. IEEE Trans Signal Proces, 2009, 57: 1068–1083
19 Weissman T, Ordentlich E, Weinberger M J, et al. Universal filtering via prediction. IEEE Trans Inform Theory, 2007, 53: 1253–1264
20 Rousseeuw P J, Leroy A M. Robust Regression and Outlier Detection. New York: Wiley, 1987