Signal Processing 102 (2014) 304–312

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A variable step-size sign algorithm for channel estimation

Yuan-Ping Li *, Ta-Sung Lee, Bing-Fei Wu

Department of Electrical and Computer Engineering, National Chiao Tung University, 1001, Ta-Hsueh Road, Hsinchu 30010, Taiwan

A R T I C L E I N F O

Article history:
Received 2 November 2013
Received in revised form
17 February 2014
Accepted 21 March 2014
Available online 28 March 2014

Keywords:
Adaptive filters
Channel estimation
Impulsive noise
Least mean square
Sign algorithm
System identification

A B S T R A C T

This paper proposes a new variable step-size sign algorithm (VSSA) for unknown channel estimation or system identification, and applies this algorithm to an environment containing two-component Gaussian mixture observation noise. The step size is adjusted using the gradient-based weighted average of the sign algorithm. The proposed scheme exhibits a fast convergence rate and low misadjustment error, and provides robustness in environments with heavy-tailed impulsive interference.

1. Introduction

In recent years, the variable step-size (VSS) techniques have been adopted in the least-mean-square (LMS) algorithm for improving the convergence rate [1–9]. A VSS technique was proposed in [4] by applying the squared instantaneous error to control the step size. A variable step-size LMS (VSLMS) algorithm using the weighted average of the gradient vector was proposed in [5] and a variable step size normalized version (VSSNLMS) was proposed in [6]. A modified version of [4] using the noise resilient variable step size was presented in [7]. A quotient form LMS algorithm of filtered version of the quadratic error for system identification application was proposed in [8]. The LMS algorithm, which is applied to the sparse channel estimation, using an l_1-norm penalty to the cost function was proposed in [9]. The channel estimation is done by an adaptive filter, the weight vector of which is $\mathbf{w}_i=[w_{0,i},\ldots,w_{N-1,i}]^T$ with a tap length of N, and is updated based on the error e_i, which is given by

$$e_i=d_i-\mathbf{w}_i^T\mathbf{x}_i \tag{1}$$

and

$$d_i=y_i+n_i=\mathbf{w}_{\text{opt}}^T\mathbf{x}_i+n_i, \tag{2}$$

where $(\cdot)^T$, d_i, $\mathbf{x}_i$, y_i, n_i, and $\mathbf{w}_{\text{opt}}$ denote the vector transpose operator, the desired signal, the input signal vector $\mathbf{x}_i=[x_i,\ldots,x_{i-N+1}]^T$, the output of the unknown system, the system noise, and the optimal Wiener weight, respectively, at time index i. The algorithm for updating the weight of the LMS adaptive filter with a fixed step size μ is given as $\mathbf{w}_{i+1}=\mathbf{w}_i+\mu e_i\mathbf{x}_i$, where $e_i\mathbf{x}_i$ is the gradient vector. This is because the cost function using $(1/2)e_i^2$ is minimized according to the weights. The mathematical formulas used in these VSLMS algorithms to update the step size μ_i are summarized in Table 1. A common problem in these algorithms is that their convergence performance can be degraded by the presence of heavy-tailed impulsive interference. Because the energy of the instantaneous error is used as the cost function of the LMS algorithm [1–9] and the error signal is sensitive to impulsive noise, this will make these LMS-type algorithms prone to considerable degradation in several practical applications. Furthermore, because the error signal is used as an

* Corresponding author.
E-mail addresses: leochan87@hotmail.com (Y.-P. Li), tslee@mail.nctu.edu.tw (T.-S. Lee), bwu@cssp.cn.nctu.edu.tw (B.-F. Wu).

http://dx.doi.org/10.1016/j.sigpro.2014.03.030
0165-1684/

Table 1
Summary and complexity of the step-size updates of some existing VSLMS algorithms.

Algorithm	Update equations of the step size	The number of *mults* (*adds*)
VSS [4]	$\mu_i=\alpha\mu_{i-1}+\gamma e_i^2$	$2N+4$ $(2N+1)$
VSLMS [5]	$\begin{cases}\hat{\mathbf{p}}_i=\beta\hat{\mathbf{p}}_{i-1}+e_{i-1}\mathbf{x}_{i-1}\\ \mu_i=\mu_{i-1}+\gamma e_i\mathbf{x}_i^T\hat{\mathbf{p}}_i\end{cases}$	$5N+3$ $(4N)$
VSSNLMS [6]	$\begin{cases}\hat{\mathbf{p}}_i=\beta\hat{\mathbf{p}}_{i-1}+(1-\beta)\frac{\mathbf{x}_i}{\Vert\mathbf{x}_i\Vert^2}e_i\\ \mu_i=\mu_s\Vert\hat{\mathbf{p}}_i\Vert^2/\Vert\mathbf{x}_i\Vert^2\left(\frac{\sigma_n^2}{N\sigma_x^2}+\Vert\hat{\mathbf{p}}_i\Vert^2\right)\end{cases}$	$6N+6$ $(5N-1)$
Proposed	$\begin{cases}\hat{\mathbf{p}}_i=\beta\hat{\mathbf{p}}_{i-1}+(1-\beta)\mathrm{sgn}(e_i)\mathbf{x}_i\\ \mu_i=\alpha\mu_{i-1}+\gamma_s\Vert\hat{\mathbf{p}}_i\Vert^2\end{cases}$	$5N+2$ $(4N)$

Note: the parameters represented by the same symbols in different algorithms are not necessarily related. The complexities of various algorithms include computation of the filter output and updates of the tap weights and step-size parameters (*mults* and *adds* denote the multiplications and additions, respectively).

Table 2
Summary and complexity of the step-size updates of some existing variable step-size sign algorithms.

Algorithm	Update equations of the step size	The number of *mults* (*adds*)
DSA [13]	$\begin{cases}r(e_i)=\begin{cases}\mathrm{sgn}(e_i),\ \vert e_i\vert\le\tau\\ L\ \mathrm{sgn}(e_i),\ \vert e_i\vert>\tau\end{cases}\\ \mu_i=\mu r(e_i)\end{cases}$	$2N+1$ $(2N)$
NRMN [14]	$\begin{cases}\lambda_i=2\mathrm{erfc}[\vert d_i\vert/\hat{\sigma}_{d,i}]\\ \hat{\sigma}_{d,i}=\sqrt{\frac{1}{N-K_w-1}\mathbf{o}_i^T\mathbf{T}\mathbf{o}_i}\\ \mu_i=\frac{2A}{[2\lambda_i+[1-\lambda_i]\sqrt{2/\pi}(\sigma_b^2+\sigma_\eta^2)^{-1/2}]N\sigma_x^2}\end{cases}$	Greater than $3N-K_w+4$ $(3N-K_w+2)$
APSA [15]	$\mu_i=\mu/\sqrt{\Vert\mathbf{x}_i\Vert^2}$	$3N$ $(3N-1)$
MVSS-APSA [16]	$\begin{cases}\beta_i=\lambda\beta_{i-1}+(1-\lambda)\vert e_{i-1}\vert\\ \mu_i=\alpha\mu_i+(1-\alpha)\min\left(\frac{\Vert e_{i-1}\vert-\beta_i\vert}{\sqrt{\Vert\mathbf{x}_{i-1}\Vert^2}},\ \mu_{i-1}\right)\end{cases}$	$3N+4$ $(3N+2)$
Proposed	$\begin{cases}\hat{\mathbf{p}}_i=\beta\hat{\mathbf{p}}_{i-1}+(1-\beta)\ \mathrm{sgn}(e_i)\mathbf{x}_i\\ \mu_i=\alpha\mu_{i-1}+\gamma_s\Vert\hat{\mathbf{p}}_i\Vert^2\end{cases}$	$5N+2$ $(4N)$

Note: the parameters of $\mathbf{T}$ and $\mathbf{o}_i$ in the NRMN algorithm [14] are set according to $\mathbf{T}=\mathrm{Diag}[1,\ldots,1,0,\ldots,0]$ and $\mathbf{o}_i=O([d_i,\ldots,d_{i-N+1}]^T)$. The $\mathbf{o}_i$ contains the most recent samples of d_i, ordered from the smallest to the largest absolute value ($O(.)$ denotes this ordering).

estimate of the step size, gradient-based algorithms are also sensitive to impulsive noise.

The sign algorithm (SA) [1–3,10–17], is now receiving attention in the adaptive filtering area because of the simplicity of its implementation. This algorithm can perform efficiently in the presence of impulsive interference. SA is more suitable for this application than LMS because it has a lower computational requirement and is resistant to the presence of impulsive interference. Based on the advantages of SA, several studies have used adaptive algorithms to reduce the detrimental effects of impulse noise. A robust mixed norm (RMN) algorithm using the weighted averaging of the l_1 and l_2 norms of error was proposed in [11] and its normalized version (NRMN) was introduced in [14]. A dual sign algorithm (DSA) operates between two sign algorithms with a large step-size parameter for increasing the convergence speed and a small one for reducing the steady-state error [12,13]. An affine projection sign algorithm (APSA) [15] using an l_1-norm optimization criterion has been proposed without involving any matrix inversion to achieve robustness against impulsive noise. A modified variable step-size APSA (MVSS-APSA) was proposed in [16] in order to obtain a fast convergence rate and small misalignment error when compared to APSA. A similar MVSS-APSA method applied to a subband adaptive filter was proposed in [17]. In [18], a variable sign-sign Wilcoxon algorithm was developed for the system identification application and performs efficiently in the presence of impulsive noise. The mathematical formulas used in these sign algorithms for updating the step size are summarized in Table 2.

This paper proposes a new framework based on scaling in the conventional SA cost function, using a critical factor γ to $\gamma|e_i|$ $(\gamma>0)$; hence, its gradient vector is $\gamma\ \mathrm{sgn}(e_i)\mathbf{x}_i$ and weight update is $\mathbf{w}_{i+1}=\mathbf{w}_i+\gamma\ \mathrm{sgn}(e_i)\mathbf{x}_i$. Similar to the step size, the parameter γ determines the convergence time and level of misadjustment of the algorithm. When the convergence speed of the SA is enhanced using a large step size, the convergence performance exhibits a substantial chattering phenomenon. The loss of information in the sign error signals occurs because they provide only positive or negative polarities, similar to a switching mode with a substantial chattering phenomenon in a control effect. To overcome this disadvantage, γ can be treated as a variable instead of a fixed value, thus compensating for the loss of information in the sign error signals. Therefore, the algorithm can converge quickly by maintaining γ as a large value in the early stages of the adaptive process and using a small γ value at the steady state to ensure accurate convergence. Therefore, estimating a smooth sign gradient vector, $\hat{\mathbf{p}}_i$, using a weighted average with a smoothing factor β $(0<\beta<1)$ was proposed so that

$$\hat{\mathbf{p}}_i=\beta\hat{\mathbf{p}}_{i-1}+(1-\beta)\text{sgn}(e_i)\mathbf{x}_i. \tag{3}$$

When using $\gamma_s\|\hat{\mathbf{p}}_i\|^2$ $(\gamma_s>0)$ instead of γ in the recursive operation, the proposed variable step-size sign algorithm (VSSA) becomes

$$\mu_i=\alpha\mu_{i-1}+\gamma_s\|\hat{\mathbf{p}}_i\|^2, \tag{4}$$

$$\mathbf{w}_{i+1}=\mathbf{w}_i+\mu_i\,\text{sgn}(e_i)\mathbf{x}_i, \tag{5}$$

where $\|\cdot\|^2$ denotes the squared Euclidean norm operation.

The behavior in (3) and (4) corresponds to low-pass filtering, which effectively reduces the noise content. The gradient vector can be regarded as a criterion of optimal performance because it always points in the direction of the greatest rate of decrease during the adaptive process toward the bottom of the error performance surface. Thus, based on these advantages, the most favorable option is to apply the weighted average of the sign gradient vector in (3) and the recursive operation in (4) to determine the step size of the adaptive algorithm. The simulation results show that the proposed VSSA achieved faster convergence, a lower misadjustment error, and lower complexity than did the gradient-based VSLMS. In addition, it provided robustness in environments exhibiting heavy-tailed impulsive interference.

2. Derivation and analysis of proposed algorithm

2.1. Modification for impulse noise

The convergence behavior of (5) has been studied in [1–3,10], and is based on Gaussian inputs and independent additive Gaussian observation noise. To extend this to a two-component Gaussian mixture for the observation noise, similar assumptions are used in the convergence analysis. The input signal is white noise, with a zero mean and variance σ_x^2. Therefore, the autocorrelation matrix of the input signals is $\mathbf{R}=E(\mathbf{x}_i\mathbf{x}_i^T)=\sigma_x^2\mathbf{I}$. Consider that a contaminated Gaussian impulse noise n_i [12] is defined as follows:

$$n_i=b_i+\omega_i\eta_i, \tag{6}$$

where b_i and η_i are each zero-mean, independent, white Gaussian sequences with variances σ_b^2 and $\sigma_\eta^2=K\sigma_b^2$ $(K\gg 1)$; ω_i is a Bernoulli random process, an independent sequence of zeros and ones with $Pr[\omega_i=1]=p_r$ and $Pr[\omega_i=0]=1-p_r$. Thus, the probability density function (pdf) of n_i is given by

$$p_{n_i}(n_i)=(1-p_r)N(0,\sigma_b^2)+p_rN(0,(K+1)\sigma_b^2), \tag{7}$$

$$\sigma_n^2=E(n_i^2)=\sigma_b^2+p_r\sigma_\eta^2=(1-p_r)\sigma_b^2+p_r[(K+1)\sigma_b^2] \tag{8}$$

If $p_r=0$ or 1, then n_i is a zero-mean Gaussian random variable.

2.2. Mean and mean-squared behavior

Let $\mathbf{v}_i=\mathbf{w}_i-\mathbf{w}_{\text{opt}}$, and $\mathbf{K}_i=E(\mathbf{v}_i\mathbf{v}_i^T)$ denotes the second moment matrix of $\mathbf{v}_i$. Eq. (2) can be inserted into (1), therefore, the error can be further represented as

$$e_i=n_i-\mathbf{v}_i^T\mathbf{x}_i \tag{9}$$

Taking the expectation in (1) and conditioned on $\mathbf{v}_i$ yields a mean squared error (MSE) of

$$E(e_i^2|\mathbf{v}_i)\approx E(e_i^2)=\sigma_{e,i}^2 \tag{10}$$

Substituting (9) in (5), taking the expectation, and using the condition in which μ_i is statistically independent of $\mathbf{x}_i$, $\mathbf{v}_i$, and e_i, the weight error vector of VSSA satisfies

$$E(\mathbf{v}_{i+1})=E(\mathbf{v}_i)+E(\mu_i)E[\text{sgn}(e_i)\mathbf{x}_i] \tag{11}$$

The second moment $\mathbf{K}_i$ of the weight error vector can be evaluated recursively as

$$\mathbf{K}_{i+1}=\mathbf{K}_i+E(\mu_i)E[\text{sgn}(e_i)(\mathbf{v}_i\mathbf{x}_i^T+\mathbf{x}_i\mathbf{v}_i^T)]+E(\mu_i^2)\mathbf{R} \tag{12}$$

According to Appendix A, The weight-error vector and the second moment $\mathbf{K}_i$ can be obtained as follows from (11) and (12), respectively:

$$E(\mathbf{v}_{i+1})=\left\{\mathbf{I}-E(\mu_i)\sqrt{\frac{2}{\pi}}\left[\frac{1-p_r}{\sqrt{\sigma_b^2+\text{tr}(\mathbf{RK}_i)}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\text{tr}(\mathbf{RK}_i)}}\right]\mathbf{R}\right\}E(\mathbf{v}_i), \tag{13}$$

$$\mathbf{K}_{i+1}=\mathbf{K}_i-E(\mu_i)\sqrt{\frac{2}{\pi}}(\mathbf{K}_i\mathbf{R}+\mathbf{RK}_i)\times\left\{\frac{1-p_r}{\sqrt{\sigma_b^2+\text{tr}(\mathbf{RK}_i)}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\text{tr}(\mathbf{RK}_i)}}\right\}+E(\mu_i^2)\mathbf{R} \tag{14}$$

Assuming the initial condition $\hat{\mathbf{p}}_0=\mathbf{0}$ and using the expectation of the squared norm of (3), the following is obtained using Lemma 1

$$\begin{aligned}E(\|\hat{\mathbf{p}}_i\|^2)&=(1-\beta)^2\sum_{k=1}^{i}\sum_{m=1}^{i}\beta^{i-k}\beta^{i-m}E[\text{sgn}(e_k)\text{sgn}(e_m)\mathbf{x}_k^T\mathbf{x}_m]\\&=(1-\beta)^2\left[\sum_{\substack{k=1\\k\neq m}}^{i}\sum_{\substack{m=1\\m\neq k}}^{i}\beta^{i-k}\beta^{i-m}\frac{2}{\pi}\frac{E(e_ke_m\mathbf{x}_k^T\mathbf{x}_m)}{\sigma_{e,k}\sigma_{e,m}}+\sum_{\substack{k=1\\k=m}}^{i}\beta^{2(i-k)}E\left(\|\mathbf{x}_k\|^2\right)\right]\\&\approx(1-\beta)^2\left\{\sum_{\substack{k=1\\k\neq m}}^{i}\sum_{\substack{m=1\\m\neq k}}^{i}\beta^{2i-k-m}\frac{2}{\pi}\left[\frac{1-p_r}{\sqrt{\sigma_b^2+\text{tr}(\mathbf{RK}_k)}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\text{tr}(\mathbf{RK}_k)}}\right]\right.\end{aligned}$$

$$\times\left[\frac{1-p_r}{\sqrt{\sigma_b^2+\mathrm{tr}(\mathbf{RK}_m)}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\mathrm{tr}(\mathbf{RK}_m)}}\right]\times E(\mathbf{v}_k^T\mathbf{x}_k\mathbf{x}_k^T\mathbf{x}_m\mathbf{x}_m^T\mathbf{v}_m)+\sum_{\substack{k=1\\k=m}}^{i}\beta^{2(i-k)}E(\|\mathbf{x}_k\|^2)\Bigg\}, \quad (15)$$

where $\sigma_{e,k}$ and $\sigma_{e,m}$ are the standard deviations of the error sequences. Note that the last line on the right-hand side of (15) corresponds to the effect of impulsive noise. Similarly, the expectation of the recursion in (4) can be obtained as follows:

$$E(\mu_i)=\gamma_s\sum_{k=1}^{i}\alpha^{i-k}E(\|\hat{\mathbf{p}}_k\|^2) \quad (16)$$

Eqs. (13)–(16) show the transient behavior of the VSSA. To analyze the steady-state performance, the following standard assumptions were made: (1) the white Gaussian noise n_i is statistically stationary, and is uncorrelated and independent of the input signal $\mathbf{x}_i$ with a distribution of $N(0,\sigma_x^2)$ and (2) when the step size is small at the steady state, the excess error simultaneously converges to a value much smaller than the value of the noise signal; therefore, $e_i\approx n_i$. For the time-index s, the system is assumed to be at the steady state when $i\geq s$, and the error signals are assumed to be uncorrelated when $k\neq m$, (15) is

$$\lim_{i\to\infty}E(\|\hat{\mathbf{p}}_i\|^2)\approx(1-\beta)^2\sum_{k=s}^{i}\beta^{2(i-k)}N\sigma_x^2. \quad (17)$$

Hence, when $i\to\infty$, (17) can be further simplified as

$$E(\|\hat{\mathbf{p}}_\infty\|^2)\approx\frac{1-\beta}{1+\beta}N\sigma_x^2. \quad (18)$$

Following the same procedure, when $i\to\infty$, and by substituting (18) into (16), (16) can be simplified as

$$E(\mu_\infty)\approx\frac{\gamma_s}{1-\alpha}\cdot\frac{1-\beta}{1+\beta}\cdot N\sigma_x^2. \quad (19)$$

Using (10), based on the Gaussian assumption in [12], allows showing $\sigma_{e,i}^2$ as a mixture of two Gaussian variables with parameters p_r and $1-p_r$ and their respective variances $(K+1)\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)$ and $\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)$. Because input $\mathbf{x}_i$ is white ($\mathbf{R}=\sigma_x^2\mathbf{I}$), using Lemma 1 in Appendix A and the standard assumption in [1–3,10,12], the MSE in (10) is derived as follows:

$$\sigma_{e,i}^2=(1-p_r)\sigma_b^2+p_r[(K+1)\sigma_b^2]+\sigma_x^2\mathrm{tr}(\mathbf{K}_i). \quad (20)$$

Observing the MSE given in (20), it is only necessary to study a recursion for k_i=tr($\mathbf{K}_i$). Taking the trace of both sides of (14) yields

$$k_{i+1}=k_i-E(\mu_i)\sigma_x^2\sqrt{\frac{8}{\pi}}\left\{\frac{1-p_r}{\sqrt{\sigma_b^2+\sigma_x^2k_i}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\sigma_x^2k_i}}\right\}k_i+E(\mu_i^2)N\sigma_x^2. \quad (21)$$

Assuming the adaptive filter has converged when $i\to\infty$, the following is obtained

$$\left\{\frac{1-p_r}{\sqrt{\sigma_b^2+\sigma_x^2k_\infty}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\sigma_x^2k_\infty}}\right\}k_\infty=\sqrt{\frac{\pi}{8}}E(\mu_\infty)N. \quad (22)$$

Assuming $\sigma_x^2k_\infty\ll\sigma_b^2$ when the system has converged to a steady state and its step size is sufficiently small, (22) can be approximated as

$$k_\infty\approx\sqrt{\frac{\pi}{8}}E(\mu_\infty)N\left(\frac{1-p_r}{\sigma_b}+\frac{p_r}{\sqrt{K+1}\sigma_b}\right)^{-1}. \quad (23)$$

The excess MSE (EMSE) defined as $\xi_{\mathrm{excess}}=\mathrm{tr}(\mathbf{RK}_\infty)=\sigma_x^2k_\infty$ is

$$\xi_{\mathrm{excess}}\approx\sqrt{\frac{\pi}{8}}E(\mu_\infty)N\sigma_x^2\left(\frac{1-p_r}{\sigma_b}+\frac{p_r}{\sqrt{K+1}\sigma_b}\right)^{-1}. \quad (24)$$

Hence, with the EMSE in (24), the VSSA produces a lower impact on the impulsive interference than does the LMS algorithm (shown in Appendix B). Substituting (19) into (24), the EMSE for the proposed VSSA becomes

$$\xi_{\mathrm{excess}}\approx\sqrt{\frac{\pi}{8}}N^2\sigma_x^4\left[\frac{\gamma_s(1-\beta)}{(1-\alpha)(1+\beta)}\right]\left(\frac{1-p_r}{\sigma_b}+\frac{p_r}{\sqrt{K+1}\sigma_b}\right)^{-1}. \quad (25)$$

According to [1–3,10], to guarantee the stability of the MSE, α, β, and γ_s can be determined by

$$0<E(\mu_\infty)\approx\frac{\gamma_s}{1-\alpha}\cdot\frac{1-\beta}{1+\beta}\cdot N\sigma_x^2<\frac{\sqrt{\frac{\pi}{2}\{(1-p_r)\sigma_b^2+p_r[(K+1)\sigma_b^2]\}}}{N\sigma_x^2}, \quad (26)$$

$$0<\gamma_s<\frac{(1-\alpha)(1+\beta)}{(1-\beta)N^2\sigma_x^4}\sqrt{\frac{\pi}{2}\{(1-p_r)\sigma_b^2+p_r[(K+1)\sigma_b^2]\}}. \quad (27)$$

Because $K\gg 1$, the right-hand side of (25) becomes

$$\left(\frac{1-p_r}{\sigma_b}+\frac{p_r}{\sqrt{K+1}\sigma_b}\right)^{-1}=\sigma_b\left(1-p_r+\frac{p_r}{\sqrt{K+1}}\right)^{-1}\approx\sigma_b(1-p_r)^{-1}. \quad (28)$$

In most cases, (28) can be simplified to σ_b when $p_r\leq 0.1$. Hence, the EMSE in (25) can be further simplified as

$$\xi_{\mathrm{excess}}\approx\sqrt{\frac{\pi}{8}}N^2\sigma_x^4\left[\frac{\gamma_s(1-\beta)}{(1-\alpha)(1+\beta)}\right]\sigma_b,\quad p_r\leq 0.1. \quad (29)$$

It can be observed in (29) that the EMSE for the proposed VSSA depends on the standard deviation of the system noise and the variance of the input vector when $p_r\leq 0.1$. The heavy-tailed impulsive noise σ_η^2 ($=K\sigma_b^2$) can be completely neglected. In addition, the proposed algorithm also performed well when verified with p_r=0.5 (not shown here). When using the EMSE, (25) can be determined

Table 3
Simulation parameters of the variable step-size sign algorithms for the channel estimation problem.

Algorithm	Parameters	SNR = 10 dB		
		White Gaussian inputs	Third-order inputs	First-order inputs
SA	μ	0.00002	0.00006	0.000227
DSA [13]	μ, τ, L	0.00002, 1, 16	0.00006, 1, 16	0.000227, 1, 16
NRMN [14]	A, K_w	0.001, 5	0.001, 5	0.001, 5
APSA [15]	μ	0.00015	0.00025	0.00035
MVSS-APSA [16]	α, β, μ_0	0.99, 0.9999999, 0.5	0.99, 0.9999999, 0.5	0.99, 0.9999999, 0.5
Proposed	α, β, γ_s	0.99, 0.9999, 0.00016	0.99, 0.9999, 0.00137	0.99, 0.9999, 0.0203

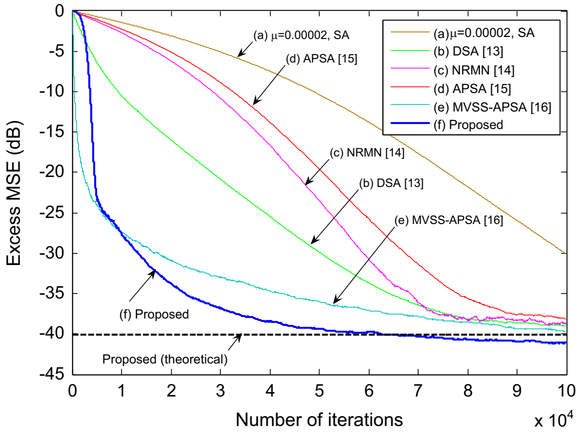

Fig. 1. Comparison of the EMSE for various adaptive sign algorithms (white Gaussian inputs, 10 dB SNR, and no impulsive noise ($p_r = 0$)).

according to p_r as follows:

$$\xi_{\text{excess}} \approx \begin{cases} \sqrt{\frac{\pi}{8}} N^2 \sigma_x^4 \left[\frac{\gamma_s(1-\beta)}{(1-\alpha)(1+\beta)}\right] \sigma_b, \ p_r = 0 \\ \sqrt{\frac{\pi}{8}} N^2 \sigma_x^4 \left[\frac{\gamma_s(1-\beta)}{(1-\alpha)(1+\beta)}\right] \left(\frac{1-p_r}{\sigma_b} + \frac{p_r}{\sqrt{K+1}\sigma_b}\right)^{-1}, \ 0 < p_r < 1 \\ \sqrt{\frac{\pi}{8}} N^2 \sigma_x^4 \left[\frac{\gamma_s(1-\beta)}{(1-\alpha)(1+\beta)}\right] (\sqrt{K+1}\sigma_b), \ p_r = 1 \end{cases} \quad (30)$$

3. Simulation results and discussion

The performance of the proposed algorithm was evaluated by carrying out computer simulations in a channel estimation scenario, using an adaptive filter with a length of 25 taps (the same as that of the unknown channel) to demonstrate the validity of the analysis. The input signal was obtained through three Gaussian distributed signals by directly passing a white zero-mean Gaussian random sequence (white Gaussian inputs) or filtering the same Gaussian random sequence through a third-order low-pass filter (third-order inputs) $G_1(z) = 0.44/(1-1.5z^{-1}+z^{-2}-0.25z^{-3})$ or a first-order system $G_2(z) = 1/(1-0.9z^{-1})$ (first-order inputs). The desired signal was generated by adding the contaminated Gaussian impulsive noise to the output of the system. The impulse response of the system was normalized as $\mathbf{w}_{\text{opt}}^T \mathbf{w}_{\text{opt}} = 1$, and the input signal was scaled so that the output power was $\sigma_y^2 = 1$. The measurement noise b_i was added to y_i such that SNR = 10 dB and 0 dB according to the calculation of the signal-to-noise ratio (SNR) [SNR = $10\log_{10}(\sigma_y^2/\sigma_b^2)$]. A strong impulsive interference with the Bernoulli-Gaussian distribution $(\omega_i \eta_i)$, where η_i was a white Gaussian random sequence in which $\sigma_\eta^2 = 100{,}000\sigma_y^2$ when SNR = 10 dB and 0 dB, and ω_i was a Bernoulli process with the probability of $Pr[\omega_i = 1] = p_r$, was also added to y_i. The results obtained in this study were averaged from over 200 independent trials. The simulation parameters of the various sign algorithms are shown in Table 3, according to the original papers. Although the studies of the step size for NRMN [14], APSA [15], and MVSS-APSA [16] had been carried out, there were no general guidelines for the selection of the step size in these proposed methods. Manual adjustment of each parameter was needed to achieve good performance. The input signals were generated using direct white Gaussian inputs, $G_1(z)$, and $G_2(z)$ for Figs. 1–3, Figs. 4 and 5, and Figs. 6 and 7, respectively, when SNR = 10 dB. For SNR = 0 dB, the performance comparison of the EMSE curves is similar to the case of SNR = 10 dB, so we only show the comparison with white Gaussian inputs (Fig. 8).

Fig. 1 shows a comparison of the EMSE curves of the proposed algorithm with those of other adaptive sign algorithms at a 10 dB SNR, without impulsive noise ($p_r = 0$). The theoretical value of the steady-state EMSE is also included. The proposed VSSA converged faster with the same steady-state error compared with SA using a fixed step size of $\mu = 0.00002$, DSA [13], NRMN [14], and APSA [15] using one projection order. Although MVSS-APSA [16] (also using one projection order) had a higher initial convergence speed, the proposed VSSA showed a lower steady-state error. Because MVSS-APSA starts with a large step size, it converges fast initially. It should be noted that the theoretical value of the steady-state EMSE is slightly biased from the simulation results because of the approximations and assumptions made in the steady-state performance analysis. Fig. 2 shows the step size of the proposed algorithm in (a), the estimates of $||\hat{\mathbf{p}}_i||^2$ with impulsive noise of $p_r = 0$ in (b), and the estimates of $||\hat{\mathbf{p}}_i||^2$ with $p_r = 0.1$ in (c). Estimates of $||\hat{\mathbf{p}}_i||^2$ and the step size were close to their respective theoretical values of the steady state according to (18) and (19), which are represented by a dashed line. Fig. 3 shows a comparison of the EMSE curves of the proposed VSSA with those of other adaptive sign algorithms at a 10 dB SNR, with impulsive noise of $p_r = 0.1$. Moreover, the change in the coefficient

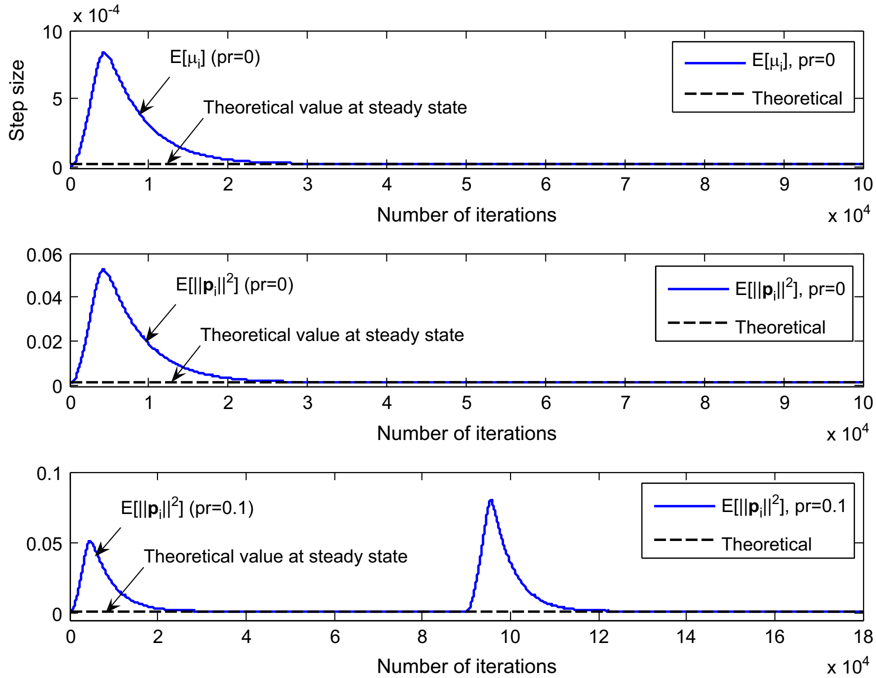

Fig. 2. (a) Estimates of the step size for the proposed method. (b) Estimates of $||\hat{\mathbf{p}}_i||^2$ with $p_r = 0$ for the proposed method and with $p_r = 0.1$ in (c) when the channel is changed. The dashed lines indicate the theoretical $||\hat{\mathbf{p}}_i||^2$ and μ_i at the steady state (white Gaussian inputs at 10 dB SNR).

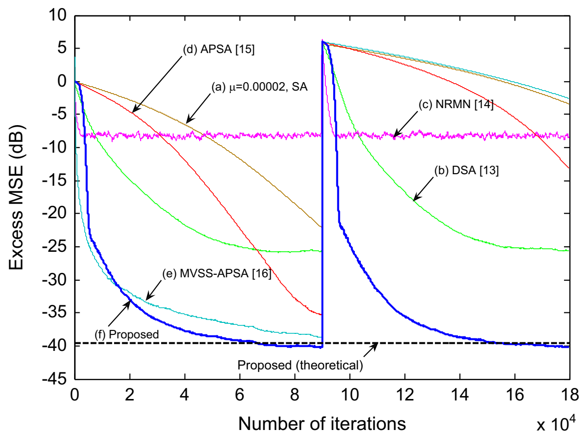

Fig. 3. Comparison of the EMSE for various adaptive sign algorithms (white Gaussian inputs, 10 dB SNR, and with impulsive noise of p_r=0.1).

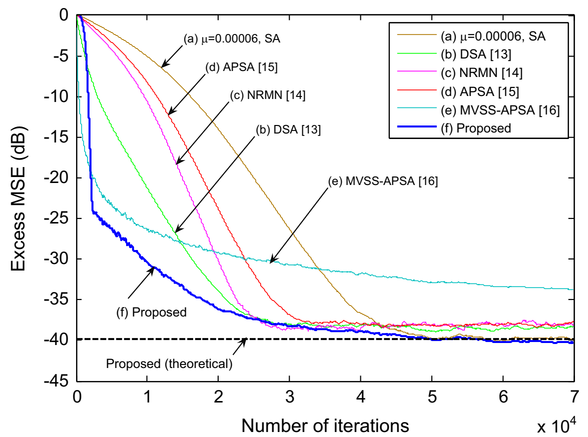

Fig. 4. Comparison of the EMSE for various adaptive sign algorithms (third-order inputs, 10 dB SNR, and no impulsive noise (p_r=0)).

values (all multiplied by −1) was abrupt when the channel was changed. As observed in Fig. 3, the proposed method converged quickly and had a low misadjustment error. The proposed VSSA performed well and was robust to the heavy-tailed impulsive interference. Figs. 4 and 5 (third-order inputs) and Figs. 6 and 7 (first-order inputs) are the simulated results, with a different input signal generated by $G_1(z)$ and $G_2(z)$. Similar result to that shown in Fig. 1 (10 dB SNR) is observed in Fig. 8 (0 dB SNR). In Fig. 8, DSA used μ=0.00002, τ=3, and L=8; NRMN used A=0.0007 and K_w=5; the step size of APSA was set to μ=0.0003 (using one projection order); MVSS-APSA used α=0.99, λ=0.9999999, μ_0=0.5, and one projection order; the proposed VSSA used α=0.99, β=0.9999, and γ_s=0.0005. These parameters were chosen to obtain the best performance and to achieve the same steady-state error for each of the compared algorithms. The proposed VSSA performed well at a 10 dB or 0 dB SNR, with heavy-tailed impulsive noises.

Methods using the technique based on the weighted average of the gradient vector were introduced in [5,6]. The gradient vector is initially large and converges into a small value at the steady state, so it can be used as a performance index for convergence. However, this leads to a performance degradation of the LMS-type algorithms [5,6] when impulsive interference is present (see Appendix B). Similarly, the experimental results in [4] are sensitive to high-level noise because the instantaneous

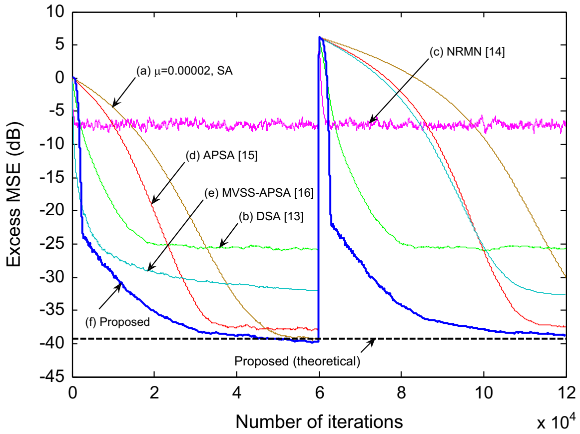

Fig. 5. Comparison of the EMSE for various adaptive sign algorithms (third-order inputs, 10 dB SNR, and with impulsive noise of $p_r=0.1$).

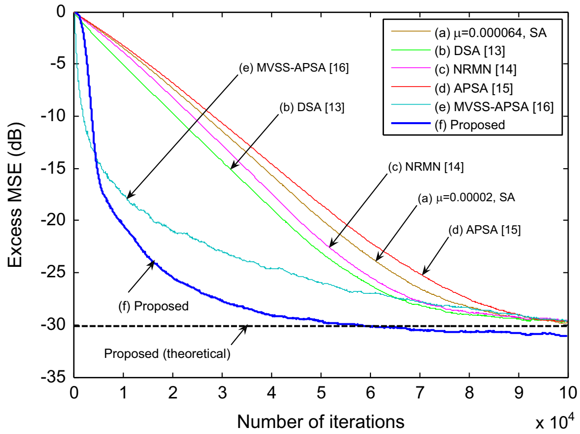

Fig. 8. Comparison of the EMSE for various adaptive sign algorithms (white Gaussian inputs, 0 dB SNR, and no impulsive noise ($p_r=0$)).

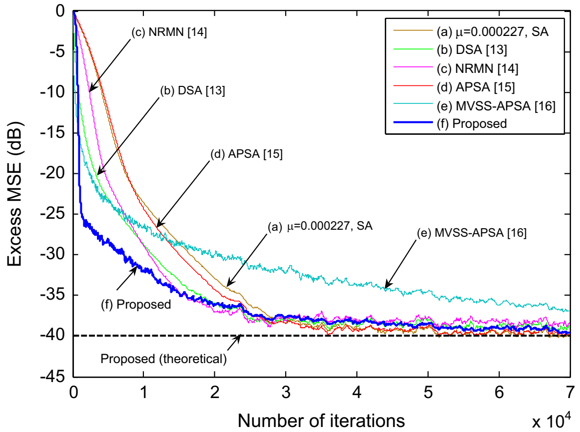

Fig. 6. Comparison of the EMSE for various adaptive sign algorithms (first-order inputs, 10 dB SNR, and no impulsive noise ($p_r=0$)).

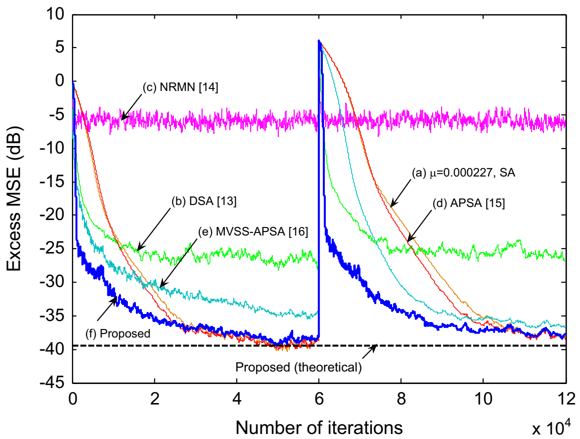

Fig. 7. Comparison of the EMSE for various adaptive sign algorithms (first-order inputs, 10 dB SNR, and with impulsive noise of $p_r=0.1$).

error value is used and could, therefore, be contaminated by the noise.

The performance of DSA [13] is determined by the values of transition thresholds and selection of two step-size parameters. It is similar to the hard-switching from one step size to another. The step size always maintains a large value when the heavy-tailed impulsive interference exists and this will lead to performance degradation. The cost function of NRMN [14] minimized according to a convex mixture of the first and second error norms, is mainly controlled by a time varying mixing parameter. If the parameter estimate tends to a large value, the NRMN algorithm is similar to the LMS algorithm and this will make the algorithm prone to considerable degradation in the presence of heavy-tailed impulsive noise. When the parameter estimate is a small value, NRMN will be similar to SA and hence converge slow. Although APSA [15] could speed up under colored input conditions, it is practically similar to SA and this makes its convergence speed lower in Gaussian input environments. In [16], when compared to APSA, the MVSS-APSA algorithm is derived based on the minimization of mean-square deviation to calculate the optimum step size and to ensure an improved performance in terms of convergence rate and misalignment. However, MVSS-APSA uses a decreasing property rule to control the step size. It always chooses the minimum value between the adjacent step sizes, so tracking capability will be degraded when the channel is changed.

From a robustness perspective, an approach to improving the performance of the family of LMS algorithms to examine the step size is using the squared norm of the sign gradient vector to enhance the dynamic range of the step size between the maximum and minimum allowable values of μ instead of using a fixed value. The squared norm of the sign gradient vector can cover the overall tracking process during adaptation, providing tracking capability when the channel is changed because the proposed VSSA uses instantaneous gradient vectors, and always points in the direction of the greatest rate of decrease during the adaptive process toward the bottom of the error performance surface. Furthermore, the recursive operation in (3) and (4), when applying the smoothing factors of α and β, is similar to low-pass filtering, which effectively reduces the noise content. This ensures that the proposed algorithm not only enhances the convergence rate and reduces the complexity, but also exhibits a low misadjustment error, and is robust against strong impulsive disturbances. The simulation results demonstrate that the proposed method performs well and is robust in low SNR, high impulsive interference, and colored input conditions. Regarding the complexity of various adaptive schemes (Tables 1 and 2), the proposed approach requires $5N+2$ multiplications and $4N$ additions per filter output for computing.

4. Conclusion

This paper introduces a new algorithm, known as VSSA, which uses the squared Euclidean norm of the sign gradient vector's weighted-averaging as a criterion for the convergence performance. The proposed VSSA combines the benefits of the gradient-based algorithm and SA. The gradient-based algorithm makes the proposed algorithm converge fast with colored input signals and simultaneously the SA guarantees its robustness against impulsive interference. Analyses and computer simulations confirm that the proposed algorithm improves the performance of conventional SA by offering a fast convergence rate, a lower misadjustment error, and a lower complexity when compared to other gradient-based VSLMS algorithms. The proposed algorithm also exhibits high robustness against strong impulsive interferences.

Acknowledgment

This work was in part funded by the Aiming for the Top University and Elite Research Center Development Plan, NSC 101-2221-E-009–093-MY2, and the MediaTek Research Center at National Chiao Tung University.

Appendix A. Proof of (13) and (14)

The following lemma is needed to verify (13) and (14):

Lemma 1. *Let u_1 and u_2 be jointly Gaussian zero-mean random variables with variances σ_1^2 and σ_2^2, and let $y=u_2+n$ and n with the pdf given in (7) be independent of u_1 and u_2. Let $z_1=u_2+h_1$ and $z_2=u_2+h_2$, where h_1 with variance $\sigma_{h_1}^2=\sigma_b^2$ and h_2 with $\sigma_{h_2}^2=(K+1)\sigma_b^2$, be zero-mean Gaussian variables independent of u_1 and u_2. Therefore,*

$$E[\mathrm{sgn}(y)u_1]=\sum_{k=1}^{2}\varepsilon_k E[\mathrm{sgn}(z_k)u_1], \tag{A.1}$$

where $\varepsilon_1=1-p_r$ *and* $\varepsilon_2=p_r$. Using (12), the second moment $\mathbf{K}_i$ of the weight error vector in (13) is necessary to calculate $E[\mathrm{sgn}(e_i)\mathbf{v}_i\mathbf{x}_i^T]$ and $E[\mathrm{sgn}(e_i)\mathbf{x}_i\mathbf{v}_i^T]$. Thus, $E[\mathrm{sgn}(e_i)\mathbf{v}_i\mathbf{x}_i^T]$ can be written as

$$E[\mathrm{sgn}(e_i)\mathbf{v}_i\mathbf{x}_i^T]=E\{E[\mathrm{sgn}(e_i)\mathbf{v}_i\mathbf{x}_i^T|\mathbf{v}_i]\} \tag{A.2}$$

Furthermore, using Price's theorem [19] and Refs. [1–3,10,12], the following result is obtained

$$E\left[\mathrm{sgn}(e_i)\mathbf{x}_i^T\right]=\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_{e,i}}E(\mathbf{x}_i^T e_i) \tag{A.3}$$

Using Lemma 1 and (A1)–(A3), $E[\mathrm{sgn}(e_i)\mathbf{v}_i\mathbf{x}_i^T|\mathbf{v}_i]$ can be written as

$$\begin{aligned}E\left[\mathrm{sgn}(e_i)\mathbf{v}_i\mathbf{x}_i^T|\mathbf{v}_i\right]&=\mathbf{v}_i\sqrt{\frac{2}{\pi}}\sum_{k=1}^{2}\frac{\varepsilon_k}{\sigma_{e_{k,i}}}E(\mathbf{x}_i^T e_{k,i}|\mathbf{v}_i)\\&=\mathbf{v}_i\sqrt{\frac{2}{\pi}}\left\{\frac{1-p_r}{\sqrt{\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}E(\mathbf{x}_i^T e_{1,i}|\mathbf{v}_i)+\frac{p_r}{\sqrt{(K+1)\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}E(\mathbf{x}_i^T e_{2,i}|\mathbf{v}_i)\right\},\end{aligned} \tag{A.4}$$

where $e_i=-\mathbf{v}_i^T\mathbf{x}_i+n_i$ and $e_{k,i}=-\mathbf{v}_i^T\mathbf{x}_i+h_{k,i}$ [k=1, 2 and $h_{1,i}$ with variance $\sigma_{h_1}^2=\sigma_b^2$ and $h_{2,i}$ with $\sigma_{h_2}^2=(K+1)\sigma_b^2$]. Taking the expectation with respect to $\mathbf{v}_i$ and with $E[\mathbf{x}_i^T e_i|\mathbf{v}_i]=-\mathbf{v}_i^T\mathbf{R}$, the following is obtained

$$E\left[\mathrm{sgn}(e_i)\mathbf{v}_i\mathbf{x}_i^T\right]=-\sqrt{\frac{2}{\pi}}\mathbf{K}_i\mathbf{R}\left\{\frac{1-p_r}{\sqrt{\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}+\frac{p_r}{\sqrt{(K+1)\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}\right\} \tag{A.5}$$

$E[\mathrm{sgn}(e_i)\mathbf{x}_i\mathbf{v}_i^T]$ can be derived using the same procedure:

$$E\left[\mathrm{sgn}(e_i)\mathbf{x}_i\mathbf{v}_i^T\right]=-\sqrt{\frac{2}{\pi}}\mathbf{RK}_i\left\{\frac{1-p_r}{\sqrt{\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}+\frac{p_r}{\sqrt{(K+1)\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}\right\} \tag{A.6}$$

Hence, we have

$$\begin{aligned}&E\left[\mathrm{sgn}(e_i)(\mathbf{v}_i\mathbf{x}_i^T+\mathbf{x}_i\mathbf{v}_i^T)\right]\\&=-\sqrt{\frac{2}{\pi}}(\mathbf{K}_i\mathbf{R}+\mathbf{RK}_i)\left\{\frac{1-p_r}{\sqrt{\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\mathrm{tr}(\mathbf{RK}_i)}}\right\}\end{aligned} \tag{A.7}$$

Similarly, (11) can be derived as

$$\begin{aligned}&E(\mathbf{v}_{i+1})=E(\mathbf{v}_i)+E(\mu_i)E[\mathrm{sgn}(e_i)\mathbf{x}_i]\\&=\left\{\mathbf{I}-E(\mu_i)\sqrt{\frac{2}{\pi}}\left[\frac{1-p_r}{\sqrt{\sigma_b^2+\mathrm{tr}(\mathbf{RK}_i)}}+\frac{p_r}{\sqrt{[(K+1)\sigma_b^2]+\mathrm{tr}(\mathbf{RK}_i)}}\right]\mathbf{R}\right\}E(\mathbf{v}_i)\end{aligned} \tag{A.8}$$

Appendix B. Derivation of excess MSE for LMS algorithm

In this appendix, the LMS algorithm using a fixed step size of μ was derived based on the two-component Gaussian mixture observation noise given in (7) and (8). According to the standard assumptions used in [1–4, 7–10,12], the weight-error vector and its second moment $\mathbf{K}_i$ can be evaluated recursively as

$$E(\mathbf{v}_{i+1})=[\mathbf{I}-\mu\mathbf{R}]E(\mathbf{v}_i) \tag{B.1}$$

and

$$\mathbf{K}_{i+1}=\mathbf{K}_i-\mu(\mathbf{RK}_i+\mathbf{K}_i\mathbf{R})+\mu^2[2\mathbf{RK}_i\mathbf{R}+\mathbf{R}\mathrm{tr}(\mathbf{RK}_i)]+\mu^2\sigma_n^2\mathbf{R} \tag{B.2}$$

Observing the MSE given in (20), it is only necessary to study a recursion for k_i=tr($\mathbf{K}_i$). Taking the trace of both sides of (B.2) yields

$$k_{i+1}=k_i-2\mu\sigma_x^2k_i+\mu^2(N+2)\sigma_x^4k_i+\mu^2N\sigma_x^2\sigma_n^2 \tag{B.3}$$

By substituting (8) into (B.3), assuming the adaptive filter has converged when $i \to \infty$, the following is obtained:

$$k_\infty = \frac{\mu N}{2-\mu\sigma_x^2(N+2)}\{(1-p_r)\sigma_b^2 + p_r[(K+1)\sigma_b^2]\} \tag{B.4}$$

The EMSE [defined as $\xi_{\text{excess}} = \text{tr}(\mathbf{RK}_\infty) = \sigma_x^2 k_\infty$ and with $\mathbf{R} = \sigma_x^2 \mathbf{I}$] is

$$\xi_{\text{excess}} = \frac{\mu N \sigma_x^2}{2-\mu\sigma_x^2(N+2)}\{(1-p_r)\sigma_b^2 + p_r[(K+1)\sigma_b^2]\} \tag{B.5}$$

It can be observed in (B.5) that the EMSE for the LMS algorithm depends on the power of the impulsive noise and the input power. Hence, the LMS that uses the energy of the instantaneous error as its cost function is sensitive to impulsive noise, making it prone to substantial degradation in several practical applications.

References

[1] B. Farhang-Boroujeny, Adaptive Filters: Theory and Applications, Wiley, New York, 1998.
[2] A.H. Sayed, Adaptive Filters, John Wiley & Sons, New York, NY, USA, 2008.
[3] P.S.R. Diniz, Adaptive Filtering: Algorithms and Practical Implementation, third ed. Springer, New York, 2008.
[4] R.H. Kwong, E.W. Johnston, A variable step size LMS algorithm, IEEE Trans. Signal Process. 40 (7) (July 1992) 1633–1642.
[5] W.P. Ang, B. Farhang-Boroujeny, A new class of gradient adaptive step-size LMS algorithms, IEEE Trans. Signal Process. 49 (4) (April 2001) 805–810.
[6] H.C. Shin, A.H. Sayed, W.J. Song, Variable step-size NLMS and affine projection algorithms, IEEE Signal Process. Lett. 11 (2) (February 2004) 132–135.
[7] M.H. Costa, J.C.M. Bermudez, A noise resilient variable step-size LMS algorithm, Signal Process. 88 (March 2008) 733–748.
[8] S. Zhao, Z. Man, S. Khoo, H.R. Wu, Variable step-size LMS algorithm with a quotient form, Signal Process. 89 (1) (January 2009) 67–76.
[9] K. Shi, P. Shi, Convergence analysis of sparse LMS algorithms with l_1-norm penalty based on white input signal, Signal Process. 90 (12) (December 2010) 3289–3293.
[10] V.J. Mathews, S.H. Cho, Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm, IEEE Trans. Acoust. Speech Signal Process. 35 (4) (April 1987) 450–454.
[11] J. Chambers, A. Avlonitis, A robust mixed-norm adaptive filter algorithm, IEEE Signal Process. Lett. 4 (2) (February 1997) 46–48.
[12] S.C. Bang, S. Ann, I. Song, Performance analysis of the dual sign algorithm for additive contaminated-Gaussian noise, IEEE Signal Process. Lett. 1 (12) (December 1994) 196–198.
[13] V.J. Mathews, Performance analysis of adaptive filters equipped with the dual sign algorithm, IEEE Trans. Signal Process. 39 (1) (January 1991) 85–91.
[14] E.V. Papoulis, T. Stathaki, A normalized robust mixed-norm adaptive algorithm for system identification, IEEE Signal Process. Lett. 11 (1) (January 2004) 173–176.
[15] T. Shao, Y.R. Zheng, J. Benesty, An affine projection sign algorithm robust against impulsive interferences, IEEE Signal Process. Lett. 17 (4) (February 2010) 173–176.
[16] S. Zhang, J. Zhang, Modified variable step-size affine projection sign algorithm, Electron. Lett. 49 (20) (September 2013) 1264–1265.
[17] J. Shin, J. Yoo, P. Park, Variable step-size sign subband adaptive filter, IEEE Signal Process. Lett. 20 (2) (February 2013) 173–176.
[18] S. Dash, M.N. Mohanty, Variable sign-sign Wilcoxon algorithm: a novel approach for system identification, Int. J. Electr. Comput. Eng. 2 (4) (August 2012) 481–486.
[19] R. Price, A useful theorem for nonlinear devices having Gaussian inputs, IRE Trans. Inf. Theory 4 (June 1958) 69–72.