Partial-Update $L_\infty$-Norm Adaptive Filtering Algorithm with Sparse Updates

Seong-Eun Kim\textsuperscript{1}, Young-Seok Choi\textsuperscript{2}, Yun-Ki Han\textsuperscript{1}, and Woo-Jin Song\textsuperscript{1}

\textsuperscript{1}Department of Electrical and Computer Engineering, Pohang University of Science and Technology, Pohang, Kyungbuk, 790-784, Korea
\textsuperscript{2}Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
E-mail: wjsong@postech.ac.kr

Abstract: This paper provides a partial-update normalized sign least-mean square (NSLMS) algorithm with sparse updates. The proposed algorithm reduces the computational complexity compared to the conventional $L_\infty$-norm adaptive filtering algorithms by decreasing the frequency of updating the filter coefficients and updating only a part of the filter coefficients. And we develop a mean square analysis to present the convergence of the proposed algorithm. Experimental results show that the proposed algorithm has the good convergence performance with greatly reduced computational complexity.

1. Introduction

In the adaptive filtering algorithms, we have the choice of algorithms ranging from the simple least-mean square (LMS) algorithm to the more complex recursive least squares (RLS) algorithm with tradeoffs between the computational complexity and the convergence speed. Among of them, the normalized least-mean square (NLMS) algorithm is widely used due to its low computational complexity and simplicity. However, the usefulness of the NLMS algorithm may be diminished for a system designed with a large number of coefficients, which is caused by increasing complexity [1], [2].

The normalized sign least-mean square (NSLMS) algorithm, which is $L_\infty$-norm based adaptive filtering algorithm, results in the reduction of the computational complexity by updating the filter coefficients in a quantized form compared with the NLMS algorithm [3]. However, the NSLMS algorithm still is not sufficient to be used for the large number of filter coefficients.

As an alternative, a significant focus for adaptive filtering has studied to reduce the computational complexity by updating only a subset of the filter coefficients [4], [5]. In another view of the solution, the reduction of the frequency of updating the filter coefficients, which is sparse in time, can reduce the overall computational complexity. These approaches can be derived within the framework of set-membership filtering (SMF) employing a bounded error constraint on the filter output [4], [6], [7].

Motivated by these frameworks, we present a low-complexity $L_\infty$-norm based adaptive filtering algorithm which incorporates both the partial-update scheme and the sparse-updates. The proposed algorithm profits from the sparse updates related to the set-membership framework reducing the overall computational complexity, as well as from the reduced computational complexity obtained by the partial update scheme.

Throughout the paper, the following norms are adopted:

\begin{align*}
\| \cdot \|_\infty & \quad L_\infty\text{-norm of a vector} \\
\| \cdot \| & \quad \text{Euclidean norm (} L_2\text{-norm) of a vector} \\
\| \cdot \|_1 & \quad L_1\text{-norm of a vector.}
\end{align*}

2. Conventional Low-Complexity $L_\infty$-Norm Adaptive Filtering Algorithms

Consider a desire signal $d(i)$ which is obtained from the linear model

\begin{equation}
\label{eq:1}
d(i) = u_i^T w^o + v(i)
\end{equation}

where $w^o$ is an unknown vector to be identified with an adaptive filter $w_i$, $v(i)$ is a measurement noise, and $u_i$ denotes the $N \times 1$ input vector,

\begin{equation}
\label{eq:2}
u_i = [u(i) u(i-1) \cdots u(i-N+1)]^T.
\end{equation}

Then, the estimation error is given by

\begin{equation}
\label{eq:3}
e(i) = d(i) - u_i^T w_i.
\end{equation}

2.1 Partial-Update Normalized Sign LMS Algorithm

This subsection reviews the partial-update normalized sign LMS (PU-NSLMS) algorithm proposed in [5]. Let the $L$ coefficients to be updated at time instant $i$ be determined by an index set $T_L(i)$ defined by

\begin{equation}
\label{eq:4}
T_L(i) = \{t_1(i), t_2(i), \cdots, t_L(i)\}
\end{equation}

where \{\{t_k(i)\}_{k=1}^L\} is taken from the set $\{1, 2, \cdots, N\}$. The matrix $S_{T_L(i)}$ is a diagonal matrix having $L$ elements equal to one in the positions indicated by $T_L(i)$ and zeros elsewhere. From the coefficients selection matrix $S_{T_L(i)}$ and the error control procedure developed in [4], the PU-NSLMS algorithm is derived in [5]. The PU-NSLMS algorithm updates the filter coefficients corresponding to the $L$ largest elements of the vector $|u_i| = |u(i)| |u(i-1)| \cdots |u(i-N+1)|^T$.

In the case, the index set $T_L(i)$ has the indices indicating the first $L$ largest maxima of $|u(i-k+1)|$ where $k = 1, 2, \cdots, N$. The update equations for the PU-NSLMS algorithm can be written as

\begin{equation}
\label{eq:5}
w_{i+1} = w_i + \frac{\mu e(i)}{\| S_{T_L(i)} u_i \|_1} \text{sign} \{ S_{T_L(i)} u_i \}.
\end{equation}

where $\mu$ is a step-size parameter. Stability of the PU-NSLMS algorithm is guaranteed, if $0 < \mu < \mu_{\text{max}}$ where [5]

\begin{equation}
\label{eq:6}
\mu_{\text{max}} = \left( \frac{2L}{N} \right) \left( \frac{2}{\pi} \xi^2 \right),
\end{equation}

905
and
\[ \zeta = \left( \frac{L}{N} \right) \left( \frac{E\{ \| S_{T_L(i)} u_i \|_1 \}}{E\{ \| u_i \|_1 \}} \right). \]  

(7)

2.2 Normalized Sign LMS Algorithm with Sparse Updates

In SMF, the filter coefficient vector \( \mathbf{w} \) is designed to achieve a specified bound on the magnitude of the estimation error [6]. The feasibility set \( \Theta \) is defined as the set of \( \mathbf{w} \) satisfying the estimation error bounded by \( \gamma \) as follows:

\[ \Theta = \bigcap_{(u,i) \in S} \{ \mathbf{w} \in \mathbb{R}^N : |d - \mathbf{u}^T \mathbf{w}| \leq \gamma \}. \]  

(8)

where \( S \) is a set containing all possible \( \{u,d\} \). In order to apply this concept to adaptive filters, the constraint set \( \mathcal{H}_i \) is defined as a set containing any vector \( \mathbf{w} \) for \( \{u_i,d(i)\} \) pair which is given by

\[ \mathcal{H}_i = \{ \mathbf{w} \in \mathbb{R}^N : |d(i) - \mathbf{u}_i^T \mathbf{w}| \leq \gamma \}. \]  

(9)

For the sparse updates of the filter coefficient vector, adaptive filter schemes employ the constraint set.

The low-complexity minimum \( L_\infty \)-norm adaptive algorithm with sparse updates is proposed in [7]. The update equation is given by

\[ \mathbf{w}_{i+1} = \mathbf{w}_i + \frac{\mu(i)e(i)}{\| u_i \|_1} \text{sign}\{u_i\}. \]  

(10)

It is shown [7] that \( \mu(i) = 1 - \gamma/\| e(i) \| \) through the geometric interpretation of (10).

3. Proposed Partial-Update NSLMS Algorithm with Sparse Updates

Our objective is to reduce the computational complexity while performing close to the NLMS family. The resulting algorithm benefit from both the partial update and the sparse updates.

Our approach is to seek a coefficient vector that minimize the change of the consecutive filter coefficient vectors in the \( L_\infty \)-norm sense, \( \| \mathbf{w}_{i+1} - \mathbf{w}_i \|_\infty \), subject to the constraint \( \mathbf{w}_{i+1} \in \mathcal{H}_i \) and the additional constraint of updating only \( L \) coefficients. Then, the constraint minimization problem can be written as

\[ \mathbf{w}_{i+1} = \arg \min_{\mathbf{w}} \| \mathbf{w} - \mathbf{w}_i \|_\infty \]  

subject to \( |d(i) - \mathbf{u}_i^T \mathbf{w}| \leq \gamma, \) \( S_{T_L(i)} (\mathbf{w} - \mathbf{w}_i) = 0. \)

where \( S_{T_L(i)} = I - S_{T_L(i)} \) is a complementary matrix which is used to represent that only \( L \) coefficients are updated.

Following the procedure proposed in [3], we can obtain the filter coefficient vector \( \mathbf{w}_{i+1} \) by adjusting \( \mathbf{w}_i \) with a vector \( S_{T_L(i)} \Delta_i \) to minimize the error signal \( e(i) \). Let the a posteriori error \( r(i) \) define as

\[ r(i) = d(i) - (\mathbf{w}_i + S_{T_L(i)} \Delta_i)^T \mathbf{u}_i. \]  

(12)

In order to minimize the error signal, the \( r(i) = (1 - \mu(i)) \varepsilon(i) \) relation is introduced in [3] where \( \mu(i) \) is a parameter to be selected in the interval \((0, 1)\). Thus, we have

\[ \mu(i) \varepsilon(i) = \Delta_i^T S_{T_L(i)} u_i. \]  

(13)

Thus, the updating equation is obtained as follows:

\[ \mathbf{w}_{i+1} = \mathbf{w}_i + \frac{\mu(i)e(i)}{\| S_{T_L(i)} u_i \|_1} \text{sign}\{S_{T_L(i)} u_i\} \]  

(14)

where \( \mu(i) \) is data dependent and given by

\[ \mu(i) = \begin{cases} 1 - \gamma/\| e(i) \|, & \text{where } |e(i)| > \gamma \\ 0, & \text{otherwise} \end{cases} \]  

(15)

and \( T_L(i) = \{ t_k(i) \}_{k=1}^{L} \) is the same as that used in the PU-NSLMS algorithm.

Fig. 1 illustrated the geometric interpretation for \( N = 2 \) and \( L = 1 \). For a certain filter coefficient vector \( \mathbf{w}_i \), a vector that has same distance from \( \mathbf{w}_i \) in the \( L_\infty \)-norm sense, would be on the square which is centered at \( \mathbf{w}_i \). Thus, the updated filter coefficient vector \( \mathbf{w}_{i+1} \) can be any point on the square represented by the dashed line in Fig. 1. In addition, the direction of \( \text{sign}\{S_{T_L(i)} u_i\} \) is \((\pm 1, 0)\) or \((0, \pm 1)\). Assume that the direction of \( \text{sign}\{S_{T_L(i)} u_i\} \) is \((0, -1)\) in Fig. 1. Since \( \mathbf{w}_{i+1} \) is obtained to satisfy the constraint set \( \mathcal{H}_i \), if \( \mathbf{w}_{i+1} \notin \mathcal{H}_i \), \( \mathbf{w}_{i+1} \) belongs to the closest bounding hyperplane in \( \mathcal{H}_i \). Therefore, \( \mathbf{w}_{i+1} \) lies on the point presented in Fig. 1. However, we investigate the step-size \( \mu(i) \) to guarantee the stability of the proposed algorithm, since the step size \( \mu \) of the PU-NSLMS algorithm is limited to \( \mu_{\text{max}} \) depending on \( L \) in Section 2.

If \( \mathbf{w}_{i+1} \) gets closer to \( \mathbf{w}_{\text{NLMS},i+1} \) than \( \mathbf{w}_i \) at every iteration, then we conclude that proposed algorithm converges. In addition, it is acceptable that the convergence rate of the update equation improves when \( \mathbf{w}_{i+1} \) is updated to be closest to \( \mathbf{w}_{\text{NLMS},i+1} \). Along this line of thought, we consider following update strategies.

Let the angle \( \theta \) shown Fig. 1 denote the angle between the direction of \( \text{sign}\{S_{T_L(i)} u_i\} \) and \( u_i \). Let angle \( \theta_{\perp} \) define as the angle in the case of \( (\mathbf{w}_{i+1} - \mathbf{w}_i)\perp(\mathbf{w}_{i+1} - \mathbf{w}_i) \).
\( w_{NLMS,i+1} \). Then, \( \cos \theta_{i} \) is given by \( \cos \theta_{\perp} = \| w_i - w_{NLMS,i+1} \| / \| w_i - w_{NLMS,i+1} \| \) where the \( w_{NLMS,i+1} \) is given by \( w_{NLMS,i+1} = w_i + e(i)u_i / \| u_i \|^2 \) in the NLMS algorithm. If \( \theta > \theta_{\perp} \), increase the threshold \( \gamma \) temporarily at the iteration \( i \) to \( \gamma_i \) so as to be \( (w_i' - w_i) \perp (w_i' - w_{NLMS,i+1}) \), i.e., \( \mu \) is modified by \( \mu(i) = \| S_{T_L}(u_i) u_i^T \|_2^2 / L \| u_i \|^2 \).

4. Convergence Analysis

For the tractable analysis, certain assumptions are supposed about the statistics of the input data \( u(i) \) and the measurement noise \( v(i) \).

Assumption 1: \( u(i) \) and \( v(i) \) is a stationary, zero-mean, white Gaussian with a variance \( \sigma_u^2 \) and \( \sigma_v^2 \), respectively. Moreover, \( v(i) \) is statistically independent of \( u(i) \).

Let coefficient error define as \( \bar{w}_i = w_i - w^* \). The following is the often realistic assumption.

Assumption 2: \( \bar{w}_i \) is independent of \( u_i \).

Assumption 3: The filter is updated with the probability \( P_s = P | e(i) > \gamma, \) and \( P | e(i) < \gamma = P | e(i) < -\gamma \).

(14) and (15) are rewritten with \( e(i) = u_i^T \bar{w}_i + v(i) \) by

\[
\bar{w}_{i+1} = w_i + \frac{\mu(i) (-u_i^T \bar{w}_i + v(i))}{\| S_{T_L}(u_i) u_i^T \|_1} \text{sign}(S_{T_L}(u_i) u_i^T) \tag{16}
\]

where

\[
\mu(i) = \begin{cases} 
1 - \gamma / e(i), & \text{where } |e(i)| > \gamma \\
0, & \text{otherwise}
\end{cases}
\]

Thus, the coefficient error vector equation is given by

\[
\bar{w}_{i+1} = \left[ I - P_{e,i} \mu(i) \text{sign}(S_{T_L}(u_i) u_i^T) \right] \bar{w}_i \\
+ P_{e,i} \mu(i) \frac{v(i) \text{sign}(S_{T_L}(u_i) u_i^T)}{\| S_{T_L}(u_i) u_i^T \|_1} \tag{18}
\]

And, the excess MSE \( \varepsilon_{i+1} \) is defined as

\[
\varepsilon_{i+1} = \sigma_u^2 E \left\{ \bar{w}_{i+1}^T \bar{w}_{i+1} \right\} \tag{19}
\]

\[
\varepsilon_{i+1} = \varepsilon_i - \sigma_u^2 E \left[ P_{e,i} \mu(i) \bar{w}_i \text{sign}(S_{T_L}(u_i) u_i^T) + \text{sign}(S_{T_L}(u_i) u_i^T) \bar{w}_i \right] \frac{1}{\| S_{T_L}(u_i) u_i^T \|_1} \]

\[
+ \sigma_u^2 E \left[ P_{e,i}^2 \mu^2(i) \bar{w}_i^T \text{sign}(S_{T_L}(u_i) u_i^T) \text{sign}(S_{T_L}(u_i) u_i^T) \bar{w}_i \right] \frac{1}{\| S_{T_L}(u_i) u_i^T \|_2^2} \]

\[
+ \sigma_u^2 E \left[ P_{e,i}^2 \mu^2(i) \text{sign}(S_{T_L}(u_i) u_i^T) \text{sign}(S_{T_L}(u_i) u_i^T) \bar{w}_i \right] \frac{1}{\| S_{T_L}(u_i) u_i^T \|_2^2} = \varepsilon_i + \sigma_u^2 (-\rho_1 + \rho_2 + \rho_3) \tag{20}
\]

\[
\rho_1 = P_{e,i} E \left[ \frac{\| S_{T_L}(u_i) u_i^T \|_1}{\| u_i \|^2} \bar{w}_i^T \text{sign}(S_{T_L}(u_i) u_i^T) + \text{sign}(S_{T_L}(u_i) u_i^T) \bar{w}_i \right] \tag{21}
\]

From (18) and (19), the excess MSE is obtained by (20), shown at the bottom of this page. In the worst case, \( \mu(i) = \| S_{T_L}(u_i) u_i^T \|_1^2 / L \| u_i \|^2 \), \( \rho_2 \) and \( \rho_3 \) are given by

\[
\rho_2 = P_{e,i}^2 E \left[ \frac{\| S_{T_L}(u_i) u_i^T \|_1^2}{\| u_i \|^4} \bar{w}_i^T \frac{\text{sign}(S_{T_L}(u_i) u_i^T)}{\| u_i \|^2} \right] \tag{22}
\]

and,

\[
\rho_3 = P_{e,i}^2 E \left[ \frac{\| S_{T_L}(u_i) u_i^T \|_1^2}{\| u_i \|^4} v^2(i) \right] \tag{23}
\]

, respectively. \( \rho_1 \) is represented at the bottom of this page.

To evaluate \( \rho_1 \), it is to compute the elements of matrix \( A = E \left[ \| S_{T_L}(u_i) u_i \| \right] \), and matrix \( B \) is given by

\[
B = \left( u_i \text{sign}(S_{T_L}(u_i) u_i) + \text{sign}(S_{T_L}(u_i) u_i) u_i^T \right) \tag{24}
\]

By the Assumption 1, the off diagonals of matrix \( A \) will average to zero. The diagonal will be an average over the \( L \) largest elements only. Let \( p_i \) denote the probability for representing that one of the \( L \) largest components contribute to the \( i \)th element on the diagonal. Moreover, let \( \left( y_{1} \right)_{N_{i}} \) define elements of the input vector \( u_i \) sorted in magnitude such that \( y_1 \leq y_2 \leq \cdots \leq y_{N_i} \). For a given \( L \), the diagonal elements of \( A \) can be calculated as follows:

\[
A_{m,m} = 2 \sum_{k=0}^{L-1} E \left[ \| S_{T_L}(u_i) u_i \| \right] p_m \left[ y_{N_i-k+1} \right] \]

\[
= 2 E \left[ \| S_{T_L}(u_i) u_i \| \right] \left( \frac{1}{N} \sum_{k=0}^{L-1} p_m \left[ y_{N_i-k+1} \right] \right) \tag{25}
\]

where input signals are i.i.d., \( p_m = 1/N \). Assuming \( N \) large such that \( \| u_i \|^2 \) can be considered a reasonable estimate of \( NE \left[ u_i^2 \right] \). In addition, by assumption 1, 2, we may rewrite \( \rho_1, \rho_2 \) and \( \rho_3 \) as follows:

\[
\rho_1 \approx 2 P_{e,i} E \left[ \| S_{T_L}(u_i) u_i \| \right] \left( \frac{1}{L^2 \sigma_u^2} \left( \| u_i \|^2 \right) \right) \tag{26}
\]

\[
\rho_2 \approx 2 P_{e,i}^2 E \left[ \| S_{T_L}(u_i) u_i \| \right] \left( \frac{1}{L^2 \sigma_u^2} \left( \| u_i \|^2 \right) \right) \tag{27}
\]

\[
\rho_3 \approx 2 P_{e,i}^2 E \left[ \| S_{T_L}(u_i) u_i \| \right] \left( \frac{1}{L^2 \sigma_u^2} \left( \| u_i \|^2 \right) \right) \tag{28}
\]
Substituting (26), (27), and (28) into (20) results in

$$\varepsilon_{t+1} = \left[ 1 - P_{e,i}(2 - P_{e,i}) \frac{E \left[ \| S_{T_L(i)} u_i \|_2^2 \right]}{LN^2 \sigma_u^2} \right] \varepsilon_t + P_{e,i}^2 \frac{E \left[ \| S_{T_L(i)} u_i \|_2^2 \right]}{LN^2 \sigma_u^2} \sigma_v^2$$

(29)

Since $P_{e,i}(2 - P_{e,i}) \frac{E \left[ \| S_{T_L(i)} u_i \|_2^2 \right]}{LN^2 \sigma_u^2} < 1$, (29) is always stable.

5. Experimental Results

We illustrate the performance of the proposed algorithm by carrying out computer simulations in a channel identification scenario. The unknown channel is the impulse response of echo path with 128 taps. The adaptive filter and the unknown channel are assumed to have the same number of taps. The input signal is randomly generated with zero mean and unit variance. The signal-to-noise ratio (SNR) was set to 30 dB. The mean square error (MSE), $E \left[ \| e(t) \|_2^2 \right]$, is taken and averaged over 2000 independent trials. Fig. 2 compares the MSE curves of the NSLMS, the SM-NSLMS [7], the PUNLSMS [5], the NLMS, the SM-PU-NLMS [4], and the proposed algorithm with $L = 16$. To get similar the steady-state error level, we set the step-size of the NLMS, the NSLMS, SM-PU-NLMS, SM-NSLMS, and the proposed algorithm with $L = 16$.

6. Conclusion

We have presented an efficient $L_\infty$-norm based adaptive filtering algorithm whose computational complexity is reduced not only by the sparse updates but by the partial updates. Analysis for the excess MSE was carried out for white input signal, and the convergence of the proposed algorithm was provided. Experimental results demonstrate that the proposed algorithm has the good convergence performance with reduced computational complexity.

7. Acknowledgement

This work was supported by the Brain Korea (BK) 21 Program funded by the Ministry of Education and HY-SDR Research Center at Hanyang University under the ITRC Program of MIC, Korea.

References