ON PRE-WHITENED SIGN ALGORITHMS

S. Ben Jebara and H. Besbes

Sup’Com, MASC Department, Cité Technologique des Communications, Ariana, TUNISIA.
E-mail: sofia.benjebara@supcom.rnu.tn, hichem.besbes@supcom.rnu.tn

ABSTRACT
This paper addresses the problem of improving convergence rate of sign algorithms. For such purpose, the idea of decorrelating signals used to pilot the adaptive algorithm is investigated. More precisely, two major modifications are carried: both input and error signals are filtered and the adaptation process is carried each two iterations. The novel algorithm is called “Pre-whitened Sign Algorithm”. To justify the idea, we develop an analytical formulation and we consider the particular case of the Dual Sign Algorithm. We prove that the proposed algorithm has low complexity and provides good convergence rate, with acceptable steady state performances. Simulations results are presented to support the theoretical analysis.

1. INTRODUCTION
The signed variants of stochastic gradient adaptive algorithms are proposed in order to improve many aspects: complexity, normalization, robustness, ... These variants include the Sign Algorithm (SA) [1], the Signed Regressor Algorithm (SRA) [2], the Sign-Sign Algorithm (SSA) [3], the Dual Sign Algorithm (DSA) [4]. The main idea is to use the polarity of the error and/or the input to update the filter coefficients.

The main drawback of signed algorithms is the slow convergence rate, especially when colored signals are used as inputs. We think that, as it was applied with basic LMS version [5, 6, 7], we can improve the convergence speed by decorrelating signals used to pilot the adaptive filter.

At our knowledge, there is no pre-whitened version of sign algorithms. We propose to develop a novel family of signed algorithms which is called “Pre-whitened Sign Algorithms”.

The main idea of the proposed algorithms is to use both input and error signals in the adaptation process and to carry the adaptation process each two iterations. The input is pre-whitened using an appropriate adaptive predictor and the error is filtered using the same pre-whitener.

The paper is organized as follows: in the next section, we will present the problem, the notations and the motivation of our work. In section 3, we present the main idea, we justify the pre-whitening approach and we illustrate in the case of the Dual Sign Algorithm. In section 4, we present an analytical analysis on the Pre-whitened Dual Sign Algorithm (PDSA), we present transient and steady states performances and compare the proposed algorithm to DSA. Finally, in section 5, some concluding points and perspectives are drawn.

2. PRELIMINARIES AND NOTATIONS

2.1. Sign algorithms
Before presenting the proposed idea, it is necessary to give a brief outline of the adaptation process. The classical block-diagram of adaptive identification system is depicted in Figure 1. The input/output equation of the system is given by:

\[ y(k) = F^T X(k) + n(k), \]

where \( F = [f_0, \ldots, f_{L-1}]^T \) is the system impulse response of length \( L \), \( X(k) = [x(k), \ldots, x(k-L+1)]^T \) is the input observation vector and \( n(k) \) is an additive white Gaussian noise.

The adaptive filter \( H(k) \) is governed by the following equations:

\[ e(k) = y(k) - H(k)^T X(k) \]
\[ H(k) = H(k) + \mu f(X(k), \ldots, X(k-m)) r(e(k), \ldots, e(k-l)), \]

where \( e(k) \) is the error signal, \( \mu \) is a positive step size, \( f(\cdot) \) and \( r(\cdot) \) are two functions characterizing the used algorithm.

The main sign algorithms are described by the following equations:

- **Sign Data Algorithm (SDA):**
  \[ f(X(k), X(k-m)) = \text{sign}(X[k]), \]
  \[ r(e(k), e(k-l)) = e(k). \]

- **Sign Algorithm (SA):**
  \[ f(X(k), X(k-m)) = X[k] \]
  \[ r(e(k), e(k-l)) = \text{sign}(e(k)). \]

- **Sign-Sign Algorithm (SSA):**
  \[ f(X(k), X(k-m)) = X[k] \]
  \[ r(e(k), e(k-l)) = \text{sign}(e(k)). \]

- **Dual Sign Algorithm (DSA):**
  \[ f(X(k), X(k-m)) = X[k] \]
  \[ r(e(k)) = \begin{cases} \text{sign}(e(k)) & |e(k)| \leq \tau \\ \gamma \text{sign}(e(k)) & |e(k)| > \tau \end{cases} \]
2.2. Performances criteria

The behavior of the algorithm is usually analyzed through the evolution of the deviation vector \( V(k) \) defined as follows:

\[
V(k) = H(k) - F.
\]  

(3)

The performances are deduced from the values of the Mean Square Deviation (MSD) and the Mean Square Error (MSE), defined as follows:

\[
\text{MSD}(k) = E\left\{ V(k)^T V(k) \right\}
\]

and

\[
\text{MSE}(k) = E\{e(k)^2\}.
\]

(4)

By defining:

\[
V(k) \triangleq E\left\{ V(k)^T V(k) \right\},
\]

(5)

and using the independence assumption between the observation vector \( X(k) \) and the deviation vector \( V(k) \), the MSD and MSE can be approximated by:

\[
\text{MSD}(k) = \text{trace}(V(k))
\]

\[
\text{MSE}(k) \approx \text{trace}(R_e V(k)) + P_n.
\]

(6)

where \( R_e \) is the auto-correlation matrix of the input vector \( X(k) \).

2.3. Motivations

To point out the effect of input correlation on the rate of convergence, we present in Figure 2 the evolution of the residual error for different values of input correlation. In this simulation context, we consider the following case: the input is a first order autoregressive process AR(1):

\[
x(k) = \rho x(k-1) + g(k),
\]

(7)

where \( g(k) \) is a Gaussian white noise. We have chosen three values of \( \rho \), namely, \( \rho = 0, 0.5 \) and \( \rho = 0.8 \). The system parameters are \( L = 10, F = [1, 0, 10, -6, -1, 4, 0.1, 5, -2, -0.1]^T \), \( P_n = 0.1 \) and \( P_e = 10 \). The DSA parameters are \( \tau = 2, \gamma = 0.3 \) and \( E = 10^{-5} \). The simulation results are averaged over 200 independent runs using 3000 samples.

Figure 2 shows that, for the same steady state, the convergence time increases when the input correlation increases. Best convergence rate is obtained for white input. The signal pre-whitening approach can then be applied in order to enhance the convergence rate of the classical DSA algorithm.

3. PRE-WHITENED SIGN ALGORITHMS

3.1. Main idea

In this section, we try to develop a pre-whitened version of an existing algorithm given by:

\[
H(k + 1) = H(k) + \mu f(X(k)) r(e(k)).
\]

(8)

Without loss of generality, let consider the input signal modeled by an AR(1). By filtering the error signal as follows:

\[
e^f(k) = e(k) - \rho e(k-1),
\]

(9)

we can show easily that \( e^f(k) \) satisfies the following relationship:

\[
e^f(k) = \frac{[F - H(k)]^T G(k) + n^f(k)}{\rho [H(k) - H(k-1)]^T X(k-1)}
\]

or

\[
e^f(k) = \frac{X(k) - \rho X(k-1)}{\rho}
\]

(10)

where \( G(k) = X(k) - \rho X(k-1) \) is the observation vector of the white signal \( g(k) \). The term \( [F - H(k)]^T G(k) + n^f(k) \) corresponds to the residual error when a white input is used and when the additive noise is the filtered version of \( n(k) \):

\[
n^f(k) = n(k) - \rho n(k-1).
\]

(11)

It is interesting to note that since \( n(k) \) is a white noise, \( n^f(k) \) and \( n^f(k-2) \) are decorrelated.

The term \( \rho [H(k) - H(k-1)]^T X(k-1) \) of \( e^f(k) \) corresponds to an augmented error due to the adaptation procedure.

If we assume that the adaptive filter is constant during iterations \( k \) and \( k-1 \), and it is equal to \( H(k-1) \), the filtered error will be equal to:

\[
e^f(k) = -V(k)^T G(k) + n^f(k).
\]

(12)

At this point, we can note that the filtered error is equivalent to the classical error obtained when the filter and the algorithm are excited by the white input \( G(k) \) and when the additive noise is \( n^f(k) \). The filtered input \( G(k) \) leads to better convergence rate. However, during steady state, the error is amplified because the filtered noise power is greater than the non filtered additive noise:

\[
E(n^f(k)) = 1 + \rho^2 E(n^2) > E(n^2).
\]

As a result, we can conclude that by updating the adaptive filter, one each two iterations, using a function of the whitened input signal and a function of the filtered error, we will watch the same behavior as exciting the filter by a white signal with an amplified additive noise.

We remark that this idea was carried with the LMS algorithm and was shown to give good performances [7].

3.2. The Pre-whitened Dual Sign Algorithm (PDSA)

In this work, we will carry out our analysis based on the DSA. The adaptation process of Pre-whitened Dual Sign Algorithm (PDSA) is defined as follows:

\[
\begin{aligned}
H(2k + 1) &= H(2k) \\
H(2k + 2) &= H(2k) + \mu r^f(2k) X^f(2k),
\end{aligned}
\]

(13)

where

\[
r^f(2k) = \begin{cases} 
\text{sign} \{e^f(2k)\} & |e^f(2k)| \leq \tau_f \\
\gamma_f \text{sign} \{e^f(2k)\} & |e^f(2k)| > \tau_f,
\end{cases}
\]

(14)
where $\gamma_f$ and $\tau_f$ are the new parameters characterizing the PD SA.

The filtered signals are defined by:

$$
\begin{cases}
  x^f(k) = x(k) - p(k)x(k-1) \\
  e^f(k) = e(k) - p(k)e(k-1).
\end{cases}
$$

(15)

The coefficient $p(k)$ estimates the first order correlation of $x(k)$. Since the input statistics are unknown, we use an adaptive algorithm, for example, the NLMS algorithm to estimate the unknown value of $p$:

$$
p(k) = p(k-1) + \mu_p \frac{x^f(x-1)x(k-1)}{x(k-1)^2 + \epsilon_p}.
$$

(16)

where $\mu_p$ is the predictor step size and $\epsilon_p$ is a regularization parameter.

We note that we limit our algorithm to a one tap predictor for two reasons. The first one is to reduce the computational requirements imposed by the additional pre-whitening filter and the second reason is to reduce the noise enhancement due to the filtering process.

4. PDSA PERFORMANCES EVALUATION

4.1. Analytical evaluation

The behavior of the deviation vector is described by the following relationship:

$$
V(2k+2) = V(2k) + \mu_f r^f(2k)X^f(2k).
$$

(17)

The deviation matrix follows the recurrent relationship:

$$
\begin{align*}
\mathcal{V}(2k+2) &= \mathcal{V}(2k) + \mu_f^2 E \{ r^f(2k)^2 X^f(2k)X^f(2k)^T \} + \mu_f E \{ r^f(2k) \} \{ V(2k)X^f(2k)^T + X^f(2k)V(2k)^T \}.
\end{align*}
$$

(18)

The analysis of the proposed algorithm is inspired from the work developed by Mathews [8] on the dual algorithm, it uses the following assumptions:

- the filtered signal is Gaussian,
- the filtered signal $x^f(k)$ is white,
- the deviation vector $V(k)$ is independent of the input signal $X(k)$,
- the step size is too small.

Using the above mentioned assumptions, and applying the Price theorem [9], we get:

$$
E \{ r^f(2k)^2 X^f(2k)X^f(2k)^T \} =
\sqrt{\frac{2 \pi}{\tau_f}} \left( \gamma_f - 1 \right) R_{x^f} \mathcal{V}(2k) R_{x^f} \frac{\tau_f}{E \{ r^f(2k)^2 \}^2} e^{- \frac{\tau_f^2}{2 E \{ r^f(2k)^2 \}^2}}
+ \left( \gamma_f - 1 \right) e(r) \mathcal{V}(2k) R_{x^f}.
$$

(19)

$$
E \{ r^f(2k) \} \{ V(2k)X^f(2k)^T + X^f(2k)V(2k)^T \} =
- \sqrt{\frac{2 \pi}{\tau_f}} \left( \gamma_f - 1 \right) R_{x^f} \mathcal{V}(2k) R_{x^f} \left( 1 + \gamma_f - 1 \right) e^{- \frac{\tau_f^2}{2 E \{ r^f(2k)^2 \}^2}}
$$

(20)

where $erf(.)$ is the error function defined by:

$$
erf(s) = \sqrt{\frac{2}{\pi}} \int_0^s e^{-t^2} dt.
$$

The power of the filtered error is expressed by:

$$
E \{ r^f(2k)^2 \} = \left( 1 - \rho^2 \right) P_x \text{trace}(\mathcal{V}(2k)) + \left( 1 + \rho^2 \right) P_n.
$$

(21)

Using the three previous expressions, we obtain:

$$
\mathcal{V}(2k+2) = \mathcal{V}(2k) + \mu_f^2 \left( \gamma_f - 1 \right) \frac{\tau_f}{E \{ r^f(2k)^2 \}^2} e^{- \frac{\tau_f^2}{2 E \{ r^f(2k)^2 \}^2}} \mathcal{V}(2k)
+ \left( \gamma_f - 1 \right) erf \left( \frac{\tau_f}{2 E \{ r^f(2k)^2 \}^2} \right) \left( 1 - \rho^2 \right) P_x I_L
- \mu_f \sqrt{\frac{2 \pi}{\tau_f}} \left( 1 - \rho^2 \right) P_n \left( 1 + \gamma_f - 1 \right) e^{- \frac{\tau_f^2}{2 E \{ r^f(2k)^2 \}^2}} \mathcal{V}(2k),
$$

(22)

where $I_L$ is the $L \times L$ identity matrix.

The behavior of the algorithm is described by resolving iteratively equations (21) and (22).

We note that by replacing filtered parameters, in equations (19) to (22) by the original ones, we get the performances of the classical Dual Sign Algorithm [8].

4.2. Validity of the theoretical expression

Figure 3 compares the theoretical and the simulation curves of $\text{MSE}(k)$ for two values of step size $\mu_{x^f} = 0.002$ and $\mu_{x^f} = 0.003$. The PDSA parameters are $\gamma_f = 2^4$ and $\tau_f = 0.5$. The simulation context is the same as previous. This figure shows that simulation curves closely match the corresponding theoretical ones, validating the proposed analysis. In the next subsections, we use theoretical curves to analyze algorithm behavior.
4.3. Transient analysis

Figure 4 shows the evolution of the theoretical curves of $MSE(k)$ using DSA and PDSA, we show that effectively, the transient state is improved when we use the PDSA. When $\mu_f = 0.002$, the convergence time is twice the one obtained using the DSA algorithm when the input is white ($\rho = 0$). In fact, this was expected since the PDSA is updated one time each two iterations. We notice therefore, that steady state becomes worse, if the step size is greater ($\mu_f = 0.003$). The adaptation step size should result from a good tradeoff between good convergence rate and acceptable steady state.

It is important to note that the same kind of results are obtained with the other pre-whitened sign algorithms (for example Pre-whitened Sign Algorithm). Due to lack of space, results are not presented in this paper.

4.4. Steady state behavior

Figure 5 shows the evolution of the Excess Mean Square Error ($EMSE = MSE - P_n$) versus the normalized step size ($\nu = \mu LP_n$ for DSA and $\nu_f = \mu_f LP_{n_f}$ for PDSA), using the same simulation conditions as previous, with $\rho = 0.8$. This figure shows that for both algorithms, steady state degrades when the step size increases. For small step size, the DSA outperforms, it is due to additive noise amplification in the term $\epsilon^T(k)$ ($P_{n_f} > P_n$). We therefore remind that PDSA convergence rate is better. One advantage of PDSA is the increasing slope, it is slower than that of DSA. This means that, if we increase the step size, the PDSA diverges slower than DSA, this is explained by the fact that power of the pre-whitened signal is smaller than the original signal.

5. CONCLUSION

In this paper, we have applied the pre-whitening concept for sign algorithms. We derived an algorithm using both input and error pre-whitening in the adaptation process, which is carried each two iterations. We emphasis our analysis on Pre-whitened Dual Sign Algorithm. The Transient behavior and steady state are then derived. We justify that the proposed concept has low complexity and provides good convergence rate, with acceptable steady state performances. As perspectives, we think about using this algorithm in acoustic echo cancellation, in order to improve algorithm robustness during double talk. In fact, classical sign algorithms fail during crossing from single talk to double talk.

6. REFERENCES