Signal Processing 131 (2017) 20–26

Contents lists available at ScienceDirect

Signal Processing

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

Short communication

Mean square deviation analysis of LMS and NLMS algorithms with white reference inputs ☆

Sheng Zhang [a,*], Jiashu Zhang [a], Hing Cheung So [b,1]

[a] The Sichuan Province Key Lab of Signal and Information Processing at Southwest Jiaotong University, Chengdu 610031, China

[b] The Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 2 June 2016
Accepted 25 July 2016
Available online 27 July 2016

Keywords:
LMS
NLMS
Step-size bound
Steady-state
Kurtosis

ABSTRACT

This paper investigates the mean square performance of the least mean square (LMS) and normalized LMS (NLMS) algorithms with white reference inputs. Their closed-form mean square deviation (MSD) expressions for the transient and steady-state regimes are derived. Additionally, bounds on the step-size which guarantee mean square stability are given. It is found that the step-size bound and transient behavior of the LMS and the steady-state MSD of the NLMS depend on the kurtosis of the input signal. Convergence rates and steady-state MSDs of the two algorithms are then compared, which shows that the normalized variant with a large step-size would offer faster convergence rate than the LMS scheme. However, when small step-sizes are employed, the LMS achieves lower steady-state MSD than the NLMS at the same convergence rate.

1. Introduction

Adaptive filtering has attracted much research interest in both theoretical and applied aspects for a long time [1–3]. Due to the good performance and easy implementation, the least mean square (LMS) and normalized LMS (NLMS) algorithms have been widely used in various applications [4–7].

There have been numerous works [8–20] on analyzing the performance of the LMS and NLMS algorithms for Gaussian inputs. In [9–11], the mean and mean square behaviors of the LMS algorithm for Gaussian inputs were studied with the use of the fourth-order Gaussian moment. When the input is stationary and zero-mean Gaussian distributed, the convergence performance of the NLMS algorithm had been presented in [14–16], which shows that this method exhibits improved convergence rate in the mean, but at the expense of high steady-state error. In these analyses, the independence assumption was used. The comprehensive study of the independence theory results were obtained by Gardner in [17]. Assuming sufficiently small step-size conditions, the steady-state behavior of the LMS algorithm was presented in [18], where the independent assumption is not required. In [19–21], closed-form expressions for the transient behaviors and the steady-state excess mean square error of the LMS and NLMS were developed. When all signals are bounded, new upper bound for the step-size of LMS algorithm was proposed in [22]. In addition, their mean square performance for cyclostationary white Gaussian input signals had been studied in [23]. However, the characteristics of the LMS and NLMS algorithms for non-Gaussian inputs have not been well studied in the literature.

In this paper, we study the mean square behavior of the LMS and NLMS algorithms in the context of adaptive noise canceling [24,25] for different white reference input distributions. It points out the effect of the kurtosis of the reference input signal on their convergence performance. Specifically, the analysis is done in the context of tracking a first-order Markov plant, which is a random walk to model a non-stationary signal for the optimum weights [6,19,23]. Mathematical models are derived for the mean square behavior of the LMS and NLMS. Their step-size bounds and steady-state behaviors are also produced. Simulation results are performed to validate our theoretical development.

2. System model

Consider the adaptive noise cancellation application shown in Fig. 1, where $x(n)$ is the reference input and $s(n)$ is the desired signal. The primary input $d(n)$ to the canceller is $d(n) = \mathbf{w}^T(n)\mathbf{x}(n) + s(n)$, where $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-L+1)]^T$, $\mathbf{w}(n) = [w_1(n), w_2(n), \ldots, w_L(n)]^T$, and $(\cdot)^T$ is the transpose of $(\cdot)$. The canceller output is $e(n) = d(n) - \hat{\mathbf{w}}^T(n)\mathbf{x}(n)$, where $\hat{\mathbf{w}}(n) = [\hat{w}_1(n), \hat{w}_2(n), \ldots, \hat{w}_L(n)]^T$

☆This work was partially supported by National Science Foundation of China (Grant: 61271341).

* Corresponding author.
E-mail addresses: 604266280@qq.com (S. Zhang), jszhang@home.swjtu.edu.cn (J. Zhang), hcso@ee.cityu.edu.hk (H.C. So).

[1] EURASIP Member.

http://dx.doi.org/10.1016/j.sigpro.2016.07.027
0165-1684/

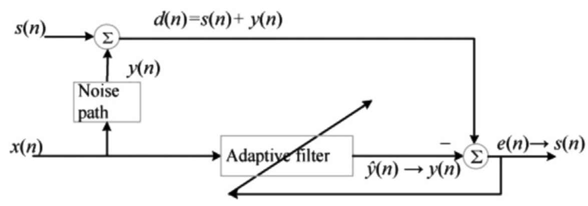

Fig. 1. Adaptive noise cancelling.

contains the adaptive filter tap weights at time n. The unknown coefficient vector in the context of a non-stationary system is a first-order Markov model $\mathbf{w}(n+1) = \mathbf{w}(n) + \mathbf{q}(n)$, where $\mathbf{q}(n)$ denotes the random perturbation.

The MSD is $\mathrm{MSD}(n) = \mathrm{Tr}(E\{(\mathbf{w}(n) - \hat{\mathbf{w}}(n))(\mathbf{w}(n) - \hat{\mathbf{w}}(n))^T\})$, where $\mathrm{Tr}(\cdot)$ is the trace of a matrix and $E\{\cdot\}$ is the mathematical expectation of a random variable.

To keep the calculations mathematically tractable, the following assumptions are adopted for the LMS and NLMS algorithms.

Assumption I. $\{x(n)\}$ is an $i.\,i.\,d.$ stationary sequence with even probability density function, finite variance $\sigma_x{}^2$, finite $E\{x^4(n)\}$, and finite $E\{1/(\mathbf{x}^T(n)\mathbf{x}(n))\}$. The kurtosis κ of $x(n)$ is $\kappa = \frac{E\{x^4(n)\}}{[E\{x^2(n)\}]^2}$. In addition, the existence of $E\{x^4(n)\}$ and $E\{1/(\mathbf{x}^T(n)\mathbf{x}(n))\}$ is needed for the LMS and NLMS algorithms, respectively. This is because the non-existence of $E\{x^4(n)\}$ and $E\{1/(\mathbf{x}^T(n)\mathbf{x}(n))\}$ will lead to divergence of the LMS and NLMS in the mean square sense [26].

Assumption II. $\{s(n)\}$ is *i.i.d.* with zero-mean and finite variance $\sigma_s{}^2$, and is independent of $\mathbf{x}(n)$ for all n.

Assumption III. The random perturbation vector $\mathbf{q}(n)$ is stationarily *i.i.d.* zero-mean vectors with $E\{\mathbf{q}(n)\mathbf{q}^T(n)\} = \sigma_q^2\mathbf{I}$, and is mutually independent of $\mathbf{x}(n)$ and $s(n)$.

Assumption IV. $\hat{\mathbf{w}}(n)$ is independent of $\mathbf{x}(n)$, which is valid when the step-size is sufficiently small [19–21].

3. Mean square analysis of LMS for white reference inputs

The LMS weight update recursion is

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu\mathbf{x}(n)e(n) \tag{1}$$

where μ is the step-size. Let $\mathbf{v}(n) = \mathbf{w}(n) - \hat{\mathbf{w}}(n)$ denote the weight error vector, the canceller output is

$$e(n) = s(n) + \mathbf{v}^T(n)\mathbf{x}(n) \tag{2}$$

and the weight error vector of the LMS can be expressed as

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \mu\mathbf{x}(n)\big(s(n) + \mathbf{x}^T(n)\mathbf{v}(n)\big) + \mathbf{q}(n) \tag{3}$$

Using Assumptions II–IV and post-multiplying (3) by its transpose, taking expectations yields

$$\begin{aligned} E\{\mathbf{v}(n+1)\mathbf{v}^T(n+1)\} &= E\{\mathbf{v}(n)\mathbf{v}^T(n)\} - 2\mu\sigma_x^2 E\{\mathbf{v}(n)\mathbf{v}^T(n)\} \\ &\quad + \mu^2 E\{\mathbf{x}(n)\mathbf{x}^T(n)E\{\mathbf{v}(n)\mathbf{v}^T(n)\}\mathbf{x}(n)\mathbf{x}^T(n)\} \\ &\quad + \mu^2\sigma_x^2 E\{s^2(n)\}\mathbf{I} + \sigma_q^2\mathbf{I} \end{aligned} \tag{4}$$

Employing Assumption I, the third expected value of the right side of (4) is

$$\begin{aligned} &E\{\mathbf{x}(n)\mathbf{x}^T(n)E\{\mathbf{v}(n)\mathbf{v}^T(n)\}\mathbf{x}(n)\mathbf{x}^T(n)\} \\ &\quad = \mathrm{Tr}(E\{\mathbf{v}(n)\mathbf{v}^T(n)\})\sigma_x^4\mathbf{I} + (\kappa-1)\sigma_x^4 E\{\mathbf{v}(n)\mathbf{v}^T(n)\} + \Delta \end{aligned} \tag{5}$$

where Δ is defined in *Appendix* and its main diagonal elements are zero. The kurtosis κ can never be less than 1, because $E\{x^4\} \geq [E\{x^2\}]^2$ for the random variable x, due to Schwartz inequality [28]. Substituting (5) into (4), we obtain

$$\begin{aligned} E\{\mathbf{v}(n+1)\mathbf{v}^T(n+1)\} &= E\{\mathbf{v}(n)\mathbf{v}^T(n)\} - 2\mu\sigma_x^2 E\{\mathbf{v}(n)\mathbf{v}^T(n)\} \\ &\quad + \mu^2\sigma_x^4\Big(\mathrm{Tr}(E\{\mathbf{v}(n)\mathbf{v}^T(n)\})\mathbf{I} \\ &\quad + (\kappa-1)E\{\mathbf{v}(n)\mathbf{v}^T(n)\}\Big) \\ &\quad + \mu^2\Delta + \mu^2\sigma_x^2 E\{s^2(n)\}\mathbf{I} + \sigma_q^2\mathbf{I} \end{aligned} \tag{6}$$

Taking the trace of both sides of (6) yields

$$\begin{aligned} \mathrm{MSD}(n+1) &= (1 - 2\mu\sigma_x^2 + \mu^2\sigma_x^4(L+\kappa-1))\mathrm{MSD}(n) \\ &\quad + \mu^2\sigma_x^2\sigma_s^2 L + \sigma_q^2 L \end{aligned} \tag{7}$$

It is shown that for different reference inputs with the same variance, the performance of the LMS is associated with the kurtosis of the reference input. In the following, it will be discussed in more detail.

3.1. Stability

The mean square stability requirement based on (7) is $|1 - 2\mu\sigma_x^2 + \mu^2\sigma_x^4(L+\kappa-1)| < 1$, i.e., μ should satisfy:

$$\mu \in \left(0, \frac{2}{(L+\kappa-1)\sigma_x^2}\right) \tag{8}$$

It is known that the selection range of the step-size of the LMS algorithm is strongly dependent on the power of the reference input signal [3]. However, (8) also shows that the step-size bound is also dependent on the kurtosis of the reference input signal.

Remark 1. (i) If there are different inputs with $\kappa_1 < \kappa_2$, we have

$$\frac{2}{(L+\kappa_2-1)\sigma_x^2} < \frac{2}{(L+\kappa_1-1)\sigma_x^2} \tag{9}$$

(ii) From (7), the transient MSD behavior of the LMS is affected by the factor $1 - 2\mu\sigma_x^2 + \mu^2\sigma_x^4(L+\kappa-1)$, which shows that the reference input with large kurtosis leads to slow convergence rate.

3.2. Steady-state performance

Assuming that (7) is operating in steady-state, and (8) is satisfied, we have

$$\mathrm{MSD}(\propto) = \frac{L\mu\sigma_s^2}{2 - \mu(L+\kappa-1)\sigma_x^2} + \frac{L\sigma_q^2}{2\mu\sigma_x^2 - \mu^2(L+\kappa-1)\sigma_x^4} \tag{10}$$

which is a monotonically increasing function of $\kappa > 0$.

Remark 2. (i) According to (10), the steady-state MSD of the LMS for the white reference input signal is not only controlled by the step-size, the filter length, and the powers of desired and reference input signals, but also dependent on the kurtosis of reference input. For small kurtosis κ with $\kappa \lll L$, the steady-state MSD is weakly correlated with κ. On the contrary, if there is a reference input signal with $\kappa \ggg L$, the correlation between steady-state MSD and the kurtosis κ cannot be ignored. When the step-size is sufficiently small such that $\mu \lll \frac{2}{(L+\kappa-1)\sigma_x^2}$, the steady-state MSD has little correlation with κ.

(ii) If the step-size satisfies the condition $\mu \lll \frac{2}{(L+\kappa-1)\sigma_x^2}$, we have from (10) for the time-varying system that

$$\mathrm{MSD}(\propto) = \frac{L\mu\sigma_s^2}{2} + \frac{L\sigma_q^2}{2\mu\sigma_x^2} \tag{11}$$

From (11), we have the minimum MSD and the corresponding optimum step-size [2]