Arab J Sci Eng (2014) 39:7145–7157
DOI 10.1007/s13369-014-1289-y

RESEARCH ARTICLE - ELECTRICAL ENGINEERING

A Unified Performance Analysis of the Family of Normalized Least Mean Algorithms

Muhammad Moinuddin · Azzedine Zerguine

Received: 13 May 2013 / Accepted: 31 October 2013 / Published online: 8 July 2014

Abstract This work presents a unified performance analysis of the family of normalized least mean (NLM) algorithms under very weak assumptions. The key feature of the analysis is based on a recently proposed performance measure called effective weight deviation vector. The name is so given as it is the only component that contributes to the excess estimation error of the adaptive filter. Using this novel concept, both the steady-state analysis and the tracking analysis are presented with a unified framework for the whole family of the NLM algorithms. Thus, the derived results are valid for any normalized stochastic gradient algorithm minimizing $2p$th power of the error where p is an integer value. The novelty of the analysis resides in the fact that it does not impose restrictions on the dependence between input successive regressors, the dependence among input regressor elements, the length of the adaptive filter, the distribution of noise, and filter's input. Moreover, this approach is not limited to only small step-size value, and therefore, the analysis is valid for all the values of the step size in the stable range of the NLM algorithms. Furthermore, in this analysis, both stationary and non-stationary input signal and plant scenarios are considered. Consequently, asymptotic time-averaged convergence for the mean-squared effective weight deviation, mean absolute excess estimation error, and the mean-square excess estimation error for the NLM algorithm are established for both constant and time-varying plants. Finally, a number of simulation results are carried out to corroborate the theoretical findings.

Keywords LMS algorithm · LMF algorithm · NLMS algorithm · NLMF algorithm

الخلاصة

يعرض هذا العمل تحليل أداء موحد لعائلة خوارزميات أقل متوسط معايَر ضمن افتراضات صغيرة جداً. وتعتمد الميزة الأساسية للتحليل على قياس أداء مقترح حديث يدعى متجه انحراف الوزن الفعال. وقد أعطي هذا الاسم لأنه العنصر الوحيد الذي يساهم في خطأ التقدير الزائد للمرشح التكيفي. وباستخدام هذا المفهوم الجديد تم عرض تحليل حالة الاستقرار وتحليل التتبع في إطار عمل موحد لكامل عائلة خوارزميات أقل متوسط معايَر. وبالتالي، فإن النتائج المستمدة صالحة لأي خوارزمية انحدار عشوائي معايَر يقلل طاقة $(2pth)$ للخطأ، حيث p هي عدد صحيح. وتكمن جدّة التحليل في حقيقة أنها لا تفرض قيوداً على الاعتمادية بين المُراجعات الداخلة المتعاقبة ، والاعتمادية عبر عناصر المُراجع الداخل ، وطول المرشح التكيفي ، وتوزيع الضجيج ، ومدخلات المرشح. علاوة على ذلك، إن هذا النهج غير مقيد لقيم الخطوات الصغيرة فقط، وبالتالي فإن التحليل صالح لكل القيم على طول الخطوة في المدى المستقر في خوارزميات أقل متوسط معايَر. وتم -إضافة إلى ذلك في هذا التحليل -الأخذ بعين الاعتبار إشارة الداخل المستقرة وغير المستقرة وسيناريوهات النظام. وتم بناء على ذلك تأسيس مقارب يتقارب بمتوسط الوقت مع متوسط مربع انحراف الوزن الفعال ، والمتوسط المطلق لخطأ التقدير الزائد ، ومتوسط مربع خطأ التقدير الزائد لخوارزمية أقل متوسط معايَر، وذلك لكل من الأنظمة الثابتة والمتغيرة مع الوقت. وفي النهاية ، تم تنفيذ عدد من النتائج التجريبية لتأكيد النتائج النظرية.

1 Introduction

The most widely used algorithm for adaptive filters is the family least mean algorithms (LM) [1]. Since the family of LM algorithms belongs to the gradient-type algorithms, it

M. Moinuddin
Electrical and Computer Engineering Department, King Abdul Aziz University, Jeddah 21859, Saudi Arabia
e-mail: mmsansari@kau.edu.sa

M. Moinuddin
Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdul Aziz University, Jeddah 21859, Saudi Arabia

A. Zerguine (✉)
Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
e-mail: azzedine@kfupm.edu.sa

inherits their low computational complexity and their slow convergence, especially on highly correlated signals like speech. This dependency is removed by the input normalization as in the case of normalized least mean-square (NLMS) algorithm [2] and the normalized least mean fourth (NLMF) algorithm [3], which results in a great improvement in the convergence behavior. However, this normalization complicates the performance analysis of these types of algorithms. Several works have attempted to study the performance of such algorithms [2,4–9]. Unfortunately, the results of these works are mostly (1) approximate by using strong assumptions, (2) lack transparency, and/or (3) do not result in closed-form expressions.

Before discussing the features of the approach proposed herein and its contributions, we provide, as a motivation, a summary of selected techniques that have been employed earlier in the literature for the study of adaptive algorithms.

(a) Independence assumption. It is common to assume that the successive regressors are independent in what is widely known as the independence assumptions [10]. Despite being unrealistic, the independence assumptions are among the most heavily used assumptions in adaptive filtering analysis.
(b) Restricted class of inputs. The input sequence is usually assumed to be white and/or has a Gaussian distribution (e.g., [11–18]).
(c) Gaussian noise. Noise is sometimes restricted to be iid Gaussian as in [11,12], and [19].
(d) Assumptions on the statistics of the weight-error vector. While it is common to impose statistical assumptions on the regression and noise sequences, similar conditions can also be imposed on weight-error vector. For example, in studying the sign-LMS algorithm, it was assumed in [19] that the elements of the weight-error vector are jointly Gaussian. This assumption was shown in [20] to be valid asymptotically.
(e) Long Filter Assumption. In [1], the steady-state and the transient performance of adaptive filters are analyzed by restricting the length of adaptive filter. This restriction allows to assume that the residual error is Gaussian [11,14], or that its conditional value is [12,13]. By central limit arguments, this assumption is justified for long adaptive filters [11,14].
(f) Small Step-size Assumption. It is very common to employ small step-size assumption in analyzing the performance of adaptive filters [1,10].

In contrast to the above-mentioned approaches, our method is not employing any of these assumptions. Our work is based on a recently introduced performance measure called *effective weight deviation vector* [21]. This was originally introduced for the convergence analysis of the NLMS algorithm [21]. This vector is the component of weight deviation vector in the direction of input regressor vector. It is shown that the effective weight deviation is the only component that contributes to the excess estimation error [21]. As a result, the analysis based on the study of this component gives more insight into the performance of the adaptive algorithm. Using this approach, we have presented the steady-state analysis for NLMF algorithm with constant channel and stationary input [22].

1.1 Contribution and Organization

In this work, we have developed a unified framework for the performance analysis of the family of NLM algorithms by employing the concept of effective weight deviation vector given in [21]. The main contribution of this work is a unified performance analysis of the family of NLM algorithms that has the following advantages:

1. It is a unified analysis, which is valid for the whole family of NLM algorithms.
2. It holds for arbitrary dependence among successive regressor vectors.
3. It holds for arbitrary dependence among the elements of regressor vector.
4. This analysis is not restricted to the class of long filters.
5. It holds for arbitrary distributions of the filter input and the noise.
6. It holds for all the values of the step size in the range that insures the stability of the NLM algorithms, and
7. It can be applied for both stationary as well as non-stationary input.
8. It can be applied for both stationary as well as non-stationary channels.

The work is organized as follows. After introducing our system model in the following subsection, a brief overview of the newly introduced performance measure is presented in Sect. 2.2. In Sect. 3, asymptotic time-averaged convergence analysis for the mean-squared effective weight deviation, mean absolute excess estimation error, and the mean-square excess estimation error of the NLM algorithm is carried out. The asymptotic time-averaged tracking analysis for a random walk channel is presented in Sect. 4, and consequently, the expressions for the mean-squared effective weight deviation, mean absolute excess estimation error, and the mean-square excess estimation error of the NLM algorithm are derived. Simulation results are presented to validate the theoretical findings in Sect. 6, and paper is ended with concluding remarks in Sect. 7.

2 Our Approach

Our approach to analyze the NLM algorithm is based on a recently proposed performance measure called *effective weight deviation vector*, which was originally introduced

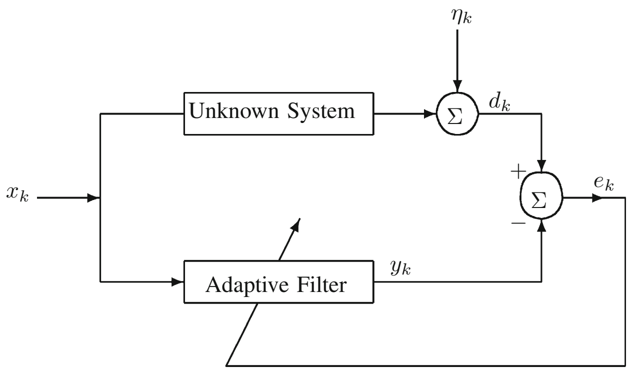

Fig. 1 Adaptive system identification

for the convergence analysis of the NLMS algorithm [21]. Before presenting the brief overview of the effective weight deviation vector, system model used for the analysis is described in the ensuing section.

2.1 System Model

Consider the case of adaptive plant identification problem [1,10] as shown in Fig. 1. The desired response d_k for the adaptive filter is obtained from output of the plant, i.e.,

$$d_k = \mathbf{c}^T \mathbf{x}_k + \eta_k, \tag{1}$$

where

$$\mathbf{c} = [c_1, c_2, \ldots, c_N]^T \tag{2}$$

is the vector composed of plant parameters that are constant, and

$$\mathbf{x}_k = [x_{1,k}, x_{2,k}, \ldots, x_{N,k}]^T \tag{3}$$

is the input data vector at time k, η_k is the plant noise, N is the number of plant parameters, and $[\cdot]^T$ is the transpose of $[\cdot]$. The inputs $x_{1,k}, x_{2,k}, \ldots,$ and $x_{N,k}$ may be successive samples of same signal, such as in the case of adaptive echo canceling and adaptive line enhancement [1,10]. They may also be the instantaneous output of N parallel sensors, such as in the case of adaptive beam-forming [1,10]. The identification of plant is made by an adaptive FIR filter whose weight vector $\mathbf{w}_k$, assumed of dimension N, is adapted on the basis of error e_k given by

$$e_k = d_k - \mathbf{w}_k^T \mathbf{x}_k. \tag{4}$$

the adaptation algorithm considered in this work is NLM algorithm, which is the normalized version of least mean family algorithms [3] and it can be described by

$$\mathbf{w}_{k+1} = \mathbf{w}_k + 2\mu p e_k^{2p-1} \frac{\mathbf{x}_k}{||\mathbf{x}_k||^2}, \tag{5}$$

where $\mu > 0$ is the algorithm step size, $2p$ is the error power in the cost function to be minimized where p is an integer value, and the norm of a vector $\mathbf{x}$ is defined as $||\mathbf{x}|| \equiv \sqrt{\mathbf{x}^T\mathbf{x}}$. The factor $2p$ in the coefficient of the second term in Eq. (5) can be incorporated within the step size μ. Thus, the NLM algorithm can be rewritten as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu e_k^{2p-1} \frac{\mathbf{x}_k}{||\mathbf{x}_k||^2}, \tag{6}$$

It can be seen that for $p = 1$, the algorithm in Eq. (6) will result in the well-known NLMS algorithm [2] while it will become the NLMF algorithm [3] for $p = 2$.

The error e_k in Eq. (6) can be decomposed to two terms: the plant noise η_k and the excess estimation error ε_k defined by

$$\varepsilon_k = e_k - \eta_k. \tag{7}$$

ε_k is also termed as adaptation noise since it represents the noise that appears at the filter output due to adaptation. The signal behavior of adaptive filter is described by the evolution of the moments of ε_k with the time. The weight deviation vector is defined by

$$\mathbf{v}_k = \mathbf{w}_k - \mathbf{c}. \tag{8}$$

due to Eqs. (1), (4), (7), and (8), and it can be shown that

$$\varepsilon_k = -\mathbf{v}_k^T \mathbf{x}_k. \tag{9}$$

2.2 Review of Effective Weight Deviation: A New Performance Measure

In this section, a brief overview of effective weight deviation vector [21] is presented. Let $\mathbf{u}_k$ denote a unit vector along the direction of the vector $\mathbf{x}_k$, i.e.,

$$\mathbf{u}_k = \begin{cases} \frac{\mathbf{x}_k}{||\mathbf{x}_k||} & \text{if } \mathbf{x}_k \neq 0 \\ \text{an arbitrary unit vector} & \text{if } \mathbf{x}_k = 0 \end{cases}$$

Consequently, the weight deviation vector $\mathbf{v}_k$ can be decomposed to two orthogonal components: the first component $\overline{\mathbf{v}}_k$ is the projection of $\mathbf{v}_k$ along the direction of vector $\mathbf{x}_k$ while the second component $\widetilde{\mathbf{v}}_k$ is orthogonal to $\mathbf{x}_k$. The vectors $\overline{\mathbf{v}}_k$ and $\widetilde{\mathbf{v}}_k$ are given by

$$\overline{\mathbf{v}}_k = (\mathbf{u}_k^T \mathbf{v}_k)\mathbf{u}_k, \tag{10}$$

$$\widetilde{\mathbf{v}}_k = \mathbf{v}_k - \overline{\mathbf{v}}_k. \tag{11}$$

Due to unit vector $\mathbf{u}_k$ and Eq. (10), the vector $\overline{\mathbf{v}}_k$ satisfies

$$\overline{\mathbf{v}}_k = \frac{\mathbf{v}_k^T \mathbf{x}_k}{||\mathbf{x}_k||^2} \mathbf{x}_k. \tag{12}$$

Equations (11), (12), and (9) imply that

$$\overline{\mathbf{v}}_k^T \mathbf{x}_k = \mathbf{v}_k^T \mathbf{x}_k = -\varepsilon_k, \tag{13}$$

Ultimately, it can be shown that

$$\widetilde{\mathbf{v}}_k^T \mathbf{x}_k = 0. \tag{14}$$

Thus, only the component $\overline{\mathbf{v}}_k$ contributes to the excess estimation error. The reminder, $\widetilde{\mathbf{v}}_k$, of the weight deviation vector $\mathbf{v}_k$ does not contribute to the excess estimation error. For this reason, $\overline{\mathbf{v}}_k$ is called as "the effective weight deviation vector" [21]. From Eqs. (10) and (13), it can be shown that

$$|\varepsilon_k| = ||\overline{\mathbf{v}}_k||||\mathbf{x}_k||. \tag{15}$$

Equation (15) shows that what matters in determining the magnitude of excess estimation error is the length of vector $\overline{\mathbf{v}}_k$ rather than the length of $\mathbf{v}_k$. Thus, studying the behavior of $||\overline{\mathbf{v}}_k||$ gives a generally brighter insight into the performance of the algorithm than studying the behavior of $||\mathbf{v}_k||$. The theoretical advantage of $\overline{\mathbf{v}}_k$ in the context of the NLM algorithm is that it can be analyzed without the need to calculate mathematical expectations of quantities normalized by $||\mathbf{x}_k||^2$. This is due to the fact that the normalization by $||\mathbf{x}_k||^2$ already included in the definition of $\overline{\mathbf{v}}_k$, as seen by Eq. (12). Therefore, $\overline{\mathbf{v}}_k$ enables a rigorous analysis of the NLM algorithm under weak assumptions. In this work, we derived an upper bound on the long-term average of mean-squared effective weight deviation ($E\left[||\overline{\mathbf{v}}_k||^2\right]$), mean-square excess estimation error ($E\left[\varepsilon_k^2\right]$), and mean absolute excess estimation error ($E\left[|\varepsilon_k|\right]$). These long-term averages are defined as follows:

$$L_1 \triangleq \operatorname*{Limsup}_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} E\left[||\overline{\mathbf{v}}_j||^2\right], \tag{16}$$

$$L_2 \triangleq \operatorname*{Limsup}_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} E\left[\varepsilon_j^2\right], \tag{17}$$

$$L_3 \triangleq \operatorname*{Limsup}_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} E\left[|\varepsilon_j|\right], \tag{18}$$

where the notation "Limsup" is defined by

$$\operatorname*{Limsup}_{k\to\infty} s_k \equiv \operatorname*{Lim}_{k\to\infty}\left(\sup_{i\geq k} s_i\right), \tag{19}$$

where "sup" means supremum that refers to least upper bound. The smaller is the value of long-term averages Eqs. (16), (17), and (18), the finer is the steady-state performance of algorithm and vice versa. The upper bound of the long-term average Eq. (16) is used along with Eq. (15) to derive bounds for mean-square excess estimation error and mean absolute excess estimation error.

3 Convergence Analysis of the NLM Algorithm

In carrying out the convergence analysis, we have considered the two scenarios of stationary and non-stationary input signals separately.

3.1 Scenario 1: Stationary Input Signal

In this case, before presenting the convergence analysis, some necessary assumptions required for our analysis are listed as follows:

- **CA1**: The sequences $\{\mathbf{x}_k\}$ and $\{\eta_k\}$ are mutually independent.
- **CA2**: The sequence $\{\mathbf{x}_k\}$ is stationary with finite $E\left[1/||\mathbf{x}_k||^2\right]$.
- **CA3**: The sequence $\{\eta_k\}$ is a stationary sequence of independent zero mean random variables with finite even moments (i.e., with finite $E[\eta_k^2] = \phi_\eta^2 = \sigma_\eta^2$, $E[\eta_k^4] = \phi_\eta^4$, $E[\eta_k^6] = \phi_\eta^6$, etc.).

Assumptions **CA1** and **CA3** are well-known independence assumption while assumption **CA2** can be well justified as in the case of NLMS algorithm [21]. It is worthwhile to note that these independence assumptions are weak as they are not employing independence between input successive regressors and the independence among input regressor elements. Assumption **CA2** is valid for the case of the NLM algorithm to ensure its stability. This is because of the fact that the nonexistence of $E\left[1/||\mathbf{x}_k||^2\right]$ will diverge the algorithm due to its high effective step-size $\mu/||\mathbf{x}_k||^2$.

3.1.1 Analysis of Effective Weight Deviation Vector

The update recursion for the weight deviation vector ($\mathbf{v}_k$) is obtained using Eqs. (1), (4), (6), and (8) and can be shown to be

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \mu(\eta_k - \mathbf{v}_k^T\mathbf{x}_k)^{2p-1}\frac{\mathbf{x}_k}{||\mathbf{x}_k||^2}. \tag{20}$$

As we are going to derive the upper bound for steady-state scenario, the higher-order terms of $\mathbf{v}_k^T\mathbf{x}_k$ can be ignored (since excess estimation error is very small at steady state). Thus, after expanding the term $(\eta_k - \mathbf{v}_k^T\mathbf{x}_k)^{2p-1}$ with binomial expansion, we can use the following approximation:

$$\mathbf{v}_{k+1} \approx \mathbf{v}_k + \mu(\eta_k^{2p-1} - (2p-1)\eta_k^{2p-2}\mathbf{v}_k^T\mathbf{x}_k)\frac{\mathbf{x}_k}{||\mathbf{x}_k||^2}. \tag{21}$$

Now, taking expectation after squaring both sides of the above equation and using assumptions **CA1** and **CA3**, it is found that

$$E[||\mathbf{v}_{k+1}||^2] = E[||\mathbf{v}_k||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} - (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big) E\left[\frac{(\mathbf{v}_k^T\mathbf{x}_k)^2}{||\mathbf{x}_k||^2}\right] + \mu^2\phi_\eta^{4p-2} E\left[\frac{1}{||\mathbf{x}_k||^2}\right], \quad (22)$$

where ϕ_η^{2p-2}, ϕ_η^{4p-2}, and ϕ_η^{4p-4} are the $(2p-2)$th, $(4p-2)$th, and $(4p-4)$th moments of the noise sequence η, respectively. Using the definition of weight deviation vector given in Eq. (12), we can rewrite the above equation as follows:

$$E[||\mathbf{v}_{k+1}||^2] = E[||\mathbf{v}_k||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} - (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big) E\left[||\overline{\mathbf{v}}_k||^2\right] + \mu^2\phi_\eta^{4p-2} E\left[\frac{1}{||\mathbf{x}_k||^2}\right], \quad (23)$$

Iterating the above equation backward $(k-1)$ iterations and using the stationarity assumption **CA2** along with assumptions **CA1** and **CA3**, the above equation can be set up as follows:

$$E[||\mathbf{v}_{k+1}||^2] = E[||\mathbf{v}_1||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} - (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big) \sum_{j=1}^{k} E\left[||\overline{\mathbf{v}}_j||^2\right] + k\mu^2\phi_\eta^{4p-2} E\left[\frac{1}{||\mathbf{x}_1||^2}\right]. \quad (24)$$

Since $E[||\mathbf{v}_{k+1}||^2]$ is a positive quantity and it converges provided that $0 < \mu < \frac{2}{(2p-1)\phi_\eta^{2p-2}}$, on dividing the above equation by k, one obtains

$$0 \leq \frac{1}{k} E[||\mathbf{v}_1||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} - (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big) \frac{1}{k}\sum_{j=1}^{k} E\left[||\overline{\mathbf{v}}_j||^2\right] + \mu^2\phi_\eta^{4p-2} E\left[\frac{1}{||\mathbf{x}_1||^2}\right] \quad (25)$$

Finally, by taking the limit as $k \to \infty$ on both sides of the above equation and using the definition Eq. (16), it can be shown that the following bound exists:

$$L_1 \leq \frac{\mu\phi_\eta^{4p-2}}{\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)} \times E\left[\frac{1}{||\mathbf{x}_1||^2}\right]. \quad (26)$$

This relation gives us an upper bound on the long-term average of the mean-squared norm of $\overline{\mathbf{v}}_k$. The above result is obtained under very weak assumptions, and it has all the points of strength mentioned in Sect. 1. It can be observed that by substituting $p = 1$ in Eq. (26), the bound will be reduced to the result for the NLMS algorithm reported in [21]. Moreover, it can be noticed from this bound that the numerator term $\mu\phi_\eta^{4p-2}$ is more effective than the denominator term $\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)$. As a result, it can be inferred that L_1 is a monotonically increasing function of μ. This fact can be observed in the simulation results (see Fig. 2).

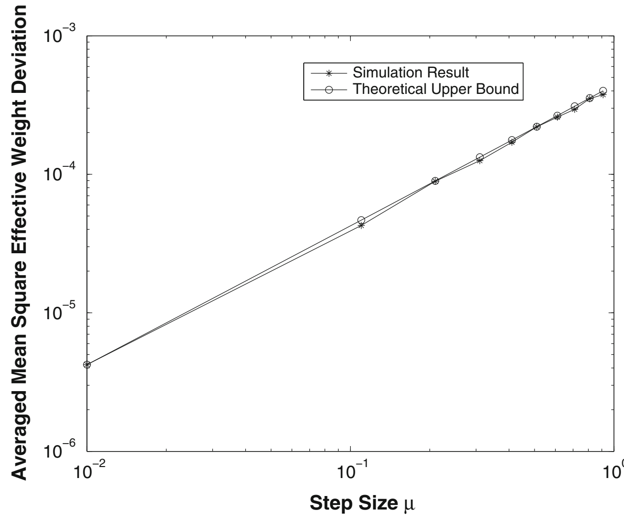

Fig. 2 Long-term average of the mean-squared effective weight deviation for constant channel identification, $N = 4$, Gaussian noise

3.1.2 Analysis of Excess Estimation Error

In analyzing the convergence of excess estimation error, we have considered the following two cases:

Case 1: Bounded Plant Input

To proceed the analysis with bounded plant input, we use the following assumption:

CA4: There exists a positive number B such that $||\mathbf{x}_k|| < B$ for all k.

This assumption is valid in many practical cases as naturally input data are bounded. Now, using the relation (15) and assumption **CA4**, the bound given in Eq. (26) is modified to the following:

$$L_2 \leq \frac{\mu\phi_\eta^{4p-2}B^2}{\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)} E\left[\frac{1}{||\mathbf{x}_1||^2}\right]. \quad (27)$$

This bound shows that the long-term average of the mean-squared excess estimation error can be reduced to an arbitrary small value by using very small value of the step size provided that $0 < \mu < \frac{2}{(2p-1)\phi_\eta^{2p-2}}$. Moreover, in achieving the above bound, we have used very weak assumptions and it has the same advantages as in the case of bound Eq. (26). Furthermore, this bound emphasizes the fact mentioned in Sect. 2.2 that a good behavior of the effective weight deviation vector implies a good behavior of the excess estimation error.

Case 2: Unbounded Plant Input
In this case too, we need the following assumption to simplify the analysis:
CA5: The sequence $\{\mathbf{x}_k\}$ is stationary with finite $E[||\mathbf{x}_k||^2]$.

This is a weak assumption as the second-order moment of input regressor generally exists. Now, using the relation (15) and assumption **CA5**, the bound given in Eq. (26) can be set up as follows:

$$L_3 \leq \sqrt{\frac{\mu\phi_\eta^{4p-2}}{\left(2(2p-1)\phi_\eta^{2p-2}-(2p-1)^2\mu\phi_\eta^{4p-4}\right)} E\left[||\mathbf{x}_1||^2\right] E\left[\frac{1}{||\mathbf{x}_1||^2}\right]}. \tag{28}$$

This bound implies that the long-term average of the absolute excess estimation error can be reduced to an arbitrary small value by using very small value of the step size provided that $0 < \mu < \frac{2}{(2p-1)\phi_\eta^{2p-2}}$. Moreover, it can be noticed that the upper bound on the right-hand side of Eq. (28) will remain unchanged even if the sequence $\mathbf{x}_k$ is multiplied by a constant (as in case of amplification or attenuation). This indicates that the average behavior of the steady-state excess estimation error is not sensitive to the input power of the adaptive filter.

3.2 Scenario 2: Non-stationary Input Signal

In this section, we consider a non-stationary input signal. Consequently, assumption **A2** cannot be employed here. In this case, instead of assumption **CA2**, we will use the following assumption,
CA6: For a non-stationary sequence $\{\mathbf{x}_k\}$, the following long-term average exists:

$$F \triangleq \operatorname*{Limsup}_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} E\left[1/||\mathbf{x}_j||^2\right]. \tag{29}$$

It can be shown that the above assumption remains valid if the time-varying moment $E\left[1/||\mathbf{x}_j||^2\right]$ is convergent as in the case of the NLMS algorithm [21].

3.2.1 Analysis of Effective Weight Deviation Vector

Starting from Eq. (23), after iterating backward $(k-1)$ iterations and using assumption **CA6**, it can be shown that

$$E[||\mathbf{v}_{k+1}||^2] = E[||\mathbf{v}_1||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} - (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big)\sum_{j=1}^{k} E\left[||\overline{\mathbf{v}}_j||^2\right] + \mu^2\phi_\eta^{4p-2}\sum_{j=1}^{k} E\left[1/||\mathbf{x}_j||^2\right]. \tag{30}$$

Working along the same line of action as for the stationary input case, we will get

$$0 \leq \frac{1}{k}E[||\mathbf{v}_1||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} - (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big)\frac{1}{k}\sum_{j=1}^{k} E\left[||\overline{\mathbf{v}}_j||^2\right] + \mu^2\phi_\eta^{4p-2}\frac{1}{k}\sum_{j=1}^{k} E\left[1/||\mathbf{x}_j||^2\right]. \tag{31}$$

Finally, by taking the limit as $k \to \infty$ on both sides of the above equation and using the definition Eq. (29), it can be shown that the following bound exists:

$$L_1 \leq \frac{\mu\phi_\eta^{4p-2}F}{\left(2(2p-1)\phi_\eta^{2p-2}-(2p-1)^2\mu\phi_\eta^{4p-4}\right)}. \tag{32}$$

Comparing the above result with that of the stationary input case given by Eq. (26), it can be observed that the long-term average of the mean-squared effective weight deviation for non-stationary input is depending on time-varying moments of $E\left[1/||\mathbf{x}_j||^2\right]$ instead of a constant moment $E\left[1/||\mathbf{x}_1||^2\right]$.

3.2.2 Analysis of Excess Estimation Error

In this case too, we have considered both the bounded and unbounded input scenarios to study the convergence of excess estimation error.

Case 1: Bounded Plant Input
For non-stationary bounded plant input, the assumption **CA4** is employed as in the case of stationary bounded input. Ultimately, by using the assumptions **CA4** and **CA6**, bound on the long-term average of mean-squared excess estimation error can be obtained from Eq. (32) as follows:

$$L_2 \leq \frac{\mu\phi_\eta^{4p-2}B^2F}{\left(2(2p-1)\phi_\eta^{2p-2}-(2p-1)^2\mu\phi_\eta^{4p-4}\right)}. \tag{33}$$

This bound shows that the long-term average of the mean-squared excess estimation error can be reduced to an arbitrary small value by using very small value of the step size. Moreover, in achieving the above bound, we have used very weak assumptions, and it has the same advantages as in the case of bound Eq. (26). Furthermore, this bound emphasizes the fact mentioned in Sect. 2.2 that a good behavior of the effective weight deviation vector implies a good behavior of the excess estimation error.

Case 2: Unbounded Plant Input

For non-stationary unbounded plant input, we will make use of the following assumption to simplify the analysis:

CA7: The following long-term average of $E[||\mathbf{x}_k||^2]$ exists:

$$P \triangleq \operatorname*{Limsup}_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} E\left[||\mathbf{x}_j||^2\right]. \tag{34}$$

Similar to the case of the NLMS algorithm, the above assumption is justified if the time-varying moment $E\left[||\mathbf{x}_j||^2\right]$ is convergent. Consequently, using the relation (15) and assumption **CA7**, the bound given in Eq. (32) can be set up as follows:

$$L_3 \le \sqrt{\frac{\mu\phi_\eta^{4p-2} P F}{\left(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\right)}}. \tag{35}$$

This bound implies that the long-term average of the absolute excess estimation error can be reduced to an arbitrary small value by using very small value of the step size. Moreover, it can be noticed that the upper bound on the right-hand side of Eq. (28) will remain unchanged even if the sequence $\mathbf{x}_k$ is multiplied by a constant (as in case of amplification or attenuation). This indicates that the average behavior of the steady-state excess estimation error is not sensitive to the input power of the adaptive filter.

4 Tracking Analysis of the NLM Algorithm

In the case of tracking analysis, adaptive plant identification of a time-varying plant is considered. More specifically, the plant parameters vary according to the random walk channel. Consequently, the desired response for the adaptive filter will be

$$d_k = \mathbf{c}_k^T\mathbf{x}_k + \eta_k, \tag{36}$$

where $\mathbf{c}_k$ is a time-varying plant varies according to random walk model, i.e.,

$$c_{k+1} = \mathbf{c}_k + \mathbf{q}_k, \tag{37}$$

where $\mathbf{q}_k$ is a random vector of same dimension as that of plant $\mathbf{c}_k$. In case of tracking, the effective weight deviation vector defined in Eq. (8) is modified to

$$\mathbf{v}_k = \mathbf{w}_k - \mathbf{c}_k. \tag{38}$$

4.1 Scenario 1: Stationary Input Signal and Plant

The assumptions employed in the tracking analysis of the NLM algorithm with stationary input signal and plant are listed as follows (note that some of the assumptions are same as employed in the case of convergence analysis):

TA1: The sequences $\{\mathbf{x}_k\}$, $\{\mathbf{q}_k\}$, and $\{\eta_k\}$ are mutually independent.

TA2 $\equiv$ **CA2**.

TA3 $\equiv$ **CA3**.

TA4: The sequence $\{\mathbf{q}_k\}$ is a stationary sequence of independent zero mean random vectors with finite second moments.

Assumptions **TA2** and **TA3** have same justification as mentioned in Sect. 3 while assumptions **TA1** and **TA4** are commonly used in analyzing the tracking of a random walk channel [10].

4.1.1 Analysis of Effective Weight Deviation Vector

The update recursion for the weight deviation vector ($\mathbf{v}_k$) is obtained using Eqs. (4), (5), (36), and (37) and found to be

$$\mathbf{v}_{k+1} \approx \mathbf{v}_k + \mu\left(\eta_k^{2p-1} - (2p-1)\eta_k^{2p-2}\mathbf{v}_k^T\mathbf{x}_k\right)\frac{\mathbf{x}_k}{||\mathbf{x}_k||^2} - \mathbf{q}_k, \tag{39}$$

where we have used the same approximation as was used in Sect. 3.1.1. Upon squaring both sides of the above equation and using assumptions **TA1**−**TA4** and Eq. (38), it is found that

$$\begin{aligned} E[||\mathbf{v}_{k+1}||^2] = E[||\mathbf{v}_k||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} \\ -(2p-1)^2\mu^2\phi_\eta^{4p-4}\Big) E\left[||\bar{\mathbf{v}}_k||^2\right] \\ +\mu^2\phi_\eta^{4p-2}E\left[\frac{1}{||\mathbf{x}_k||^2}\right] + E\left[||\mathbf{q}_1||^2\right]. \end{aligned} \tag{40}$$

Iterating the above equation backward $(k-1)$ iterations gives the following recursion:

$$\begin{aligned} E[||\mathbf{v}_{k+1}||^2] = E[||\mathbf{v}_1||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} \\ -\ (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big)\sum_{j=1}^{k} E\left[||\bar{\mathbf{v}}_j||^2\right] \\ +k\mu^2\phi_\eta^{4p-2}E\left[\frac{1}{||\mathbf{x}_1||^2}\right] + kE\left[||\mathbf{q}_1||^2\right]. \end{aligned} \tag{41}$$

Since $E[||\mathbf{v}_{k+1}||^2]$ is a positive quantity, therefore upon dividing the above recursion by k, one obtains

$$0 \le \frac{1}{k}E[||\mathbf{v}_1||^2] - \Big(2(2p-1)\mu\phi_\eta^{2p-2} - (2p-1)^2\mu^2\phi_\eta^{4p-4}\Big)\frac{1}{k}\sum_{j=1}^{k}E\left[||\overline{\mathbf{v}}_j||^2\right] + \mu^2\phi_\eta^{4p-2}E\left[\frac{1}{||\mathbf{x}_1||^2}\right] + E\left[||\mathbf{q}_1||^2\right]. \quad (42)$$

Finally, by taking the limit as $k \to \infty$ and using the definition Eq. (16), the following upper bound is provided

$$L_1 \le \frac{1}{\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)} \times \left[\mu\phi_\eta^{4p-2}E\left[\frac{1}{||\mathbf{x}_1||^2}\right] + \mu^{-1}E\left[||\mathbf{q}_1||^2\right]\right]. \quad (43)$$

It is worthwhile to note that we are not restricting our tracking analysis for slow plant variation, which is commonly employed in the analysis of adaptive filtering. Moreover, the bound given by Eq. (43) has all the points of strength which are mentioned in Sect. 1. Furthermore, unlike the constant plant case, the bound for tracking random walk plant Eq. (43) is not a monotonically increasing function of step size. In contrast, the first term in the bracket (i.e., $\mu\phi_\eta^{4p-2}E\left[\frac{1}{||\mathbf{x}_1||^2}\right]$) is increasing with step size while the second term (i.e., $\mu^{-1}E\left[||\mathbf{q}_1||^2\right]$) is decreasing with step size.

4.1.2 Analysis of Excess Estimation Error

The cases of bounded and unbounded plant input are analyzed separately.

Case 1: Bounded Plant Input
Here too, assumption **CA4** can be employed as was used in the convergence analysis of constant plant in Sect. 3. Consequently, using **TA1**−**TA4** and **CA4**, it can be shown that bound on the long-term average of mean-square excess estimation error is as follows:

$$L_2 \le \frac{B^2}{\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)} \times \left[\mu\phi_\eta^{4p-2}E\left[\frac{1}{||\mathbf{x}_1||^2}\right] + \mu^{-1}E\left[||\mathbf{q}_1||^2\right]\right]. \quad (44)$$

Case 2: Unbounded Plant Input
For tracking with unbounded plant input, the assumption **CA5** is still valid. As a result, by employing **TA1** − **TA4** and **CA5**, bound on the long-term average of absolute excess estimation error can be shown to be

$$L_3 \le \frac{E\left[||\mathbf{x}_1||^2\right]}{\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)} \times \left[\mu\phi_\eta^{4p-2}E\left[\frac{1}{||\mathbf{x}_1||^2}\right] + \mu^{-1}E\left[||\mathbf{q}_1||^2\right]\right]. \quad (45)$$

4.2 Scenario 2: Non-stationary Input Signal and Plant

In this section, we consider non-stationary input signal and plant. Here, in addition to assumptions **CA6** and **CA7**, we will also make use of the following assumption instead of the assumption **TA4**,

TA5: The sequence $\{\mathbf{q}_k\}$ is a sequence of independent zero mean random vectors with a finite value of the following long-term average

$$Q \triangleq \underset{k\to\infty}{Limsup}\,\frac{1}{k}\sum_{j=1}^{k}E\left[||\mathbf{q}_j||^2\right]. \quad (46)$$

This assumption can also be justified similar to the justification of **CA6** and **CA7**, that is, the long-term average Q will exist if its time-varying moment ($E\left[||\mathbf{q}_j||^2\right]$) is convergent.

4.2.1 Analysis of Effective Weight Deviation Vector

Starting from Eq. (40), after some simplification steps by employing assumptions **TA1**, **TA3**, **TA5**, **CA6**, and **CA7**, the upper bound on the long-term average of mean-squared effective weight deviation is found to be

$$L_1 \le \frac{1}{\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)} \times \left[\mu\phi_\eta^{4p-2}F + \mu^{-1}Q\right]. \quad (47)$$

4.2.2 Analysis of Excess Estimation Error

Here again, the cases of bounded and unbounded plant input are investigated separately.

Case 1: Bounded Plant Input
By employing **TA1**, **TA3**, **TA5**, **CA4**, and **CA6**, bound on the long-term average of mean-square excess estimation error is found to be

$$L_2 \le \frac{B^2}{\Big(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\Big)} \times \left[\mu\phi_\eta^{4p-2}F + \mu^{-1}Q\right]. \quad (48)$$

Case 2: Unbounded Plant Input
Here, by employing **TA1**, **TA3**, **TA5**, **CA5**, **CA6**, and **CA7**, bound on the long-term average of absolute excess estimation

Table 1 Summary of the results for convergence analysis

Bound on	Stationary input	Non-stationary input
L_1	$\frac{\mu\phi_\eta^{4p-2}}{D} E\left[\frac{1}{\|\|\mathbf{x}_1\|\|^2}\right]$	$\frac{\mu\phi_\eta^{4p-2} F}{D}$
L_2	$\frac{\mu\phi_\eta^{4p-2} B^2}{D} E\left[\frac{1}{\|\|\mathbf{x}_1\|\|^2}\right]$	$\frac{\mu\phi_\eta^{4p-2} B^2 F}{D}$
L_3	$\sqrt{\frac{\mu\phi_\eta^{4p-2}}{D} E\left[\|\|\mathbf{x}_1\|\|^2\right] E\left[\frac{1}{\|\|\mathbf{x}_1\|\|^2}\right]}$	$\sqrt{\frac{\mu\phi_\eta^{4p-2} P F}{D}}$

error can be shown to be

$$L_3 \leq \frac{P}{\left(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\right)} \times \left[\mu\phi_\eta^{4p-2} F + \mu^{-1} Q\right]. \quad (49)$$

5 Summary of Analysis

Here, the analytical results of Sects. 3 and 4 are summarized in the following two tables. Tables 1 and 2 demonstrate different bounds (L_1, L_2, and$L3$) with both stationary and non-stationary scenarios for the convergence and the tracking analysis of the NLM algorithm, respectively. In these tables, D stands for

$$D \triangleq \left(2(2p-1)\phi_\eta^{2p-2} - (2p-1)^2\mu\phi_\eta^{4p-4}\right) \quad (50)$$

6 Simulation Results

In this section, the performance analysis of the NLM algorithm is investigated in an unknown system identification problem with $\mathbf{c} = [1, 1, \ldots, 1]^T$. Specifically, we have investigated the performance of the NLM algorithm with $p = 2$ (as the result for $p = 1$ is already available in [21]), which corresponds to the well-known NLMF algorithm. System noise η is a zero mean i.i.d. sequence with variance 0.01. The plant input regressor vector $\mathbf{x}_k = [x_k, x_{k-1}, \ldots, x_{k-N+1}]^T$ with x_k being stationary zero mean unity variance. The objective of our simulations is to validate the derived analytical results without restrictions on

1. The dependence between successive regressors,
2. The dependence between the components of regressor,
3. The value of step size in the range $0 < \mu < \frac{2}{(2p-1)\phi_\eta^{2p-2}}$,
4. The length of adaptive filter,
5. The distribution of the filter input and the noise,
6. Stationarity of input and plant

The first two objectives are achieved via generating a highly correlated input sequence as follows:

$$x_k = \beta x_{k-1} + \sqrt{1-\beta^2} w_k \quad (51)$$

where β is a correlation factor and w_k is a zero mean unity variance i.i.d. sequence. In our simulations, we have used $\beta = 0.95$ showing a highly correlated input sequence. In order to show that our analytical results analysis holds for all the values of the step size in the range of stable NLMF algorithm, all simulation experiments are carried out for a wide range of step size [0.01, 1].

To meet the above-mentioned objectives, simulation experiments are carried out for both constant channel identification and tracking random walk channel scenarios.

6.1 Constant Channel Identification

Figure 2 compares the long-term average of the mean-squared effective weight deviation obtained by simulation and the upper bound given in Eq. (26) for Gaussian noise with filter length equal to 4 showing a good match between theory and simulation. It can be seen that the simulation is carried out over a wide range of step size showing that the analysis is not restricted to small value of step size only. To verify the non-restriction of analytical results by filter length, the same experiment is repeated filter length equal to 25 in Fig. 3. It can be depicted from these figures that the analytical result is valid for both long and short adaptive filters. Next, to check the non-restriction on distribution of the plant input and the noise, the same experiment is performed with uniform input and noise in Fig. 4. It can be depicted from the figure that the derived analytical result is not limited to a particular distribution of input and noise sequences.

Next, the analytical upper bound on the long-term average of mean-squared excess estimation error given by Eq. (27) is investigated in Fig. 5 with binary input (which is a bounded input), uniform noise, and filter length equal to 4. Here too, analytical result has modeled the simulation results very well.

Table 2 Summary of the results for tracking analysis

Bound on	Stationary input	Non-stationary input and plant
L_1	$\frac{1}{D}\left[\mu\phi_\eta^{4p-2} E\left[\frac{1}{\|\|\mathbf{x}_1\|\|^2}\right] + \mu^{-1} E\left[\|\|\mathbf{q}_1\|\|^2\right]\right]$	$\frac{1}{D}\left[\mu\phi_\eta^{4p-2} F + \mu^{-1} Q\right]$
L_2	$\frac{B^2}{D}\left[\mu\phi_\eta^{4p-2} E\left[\frac{1}{\|\|\mathbf{x}_1\|\|^2}\right] + \mu^{-1} E\left[\|\|\mathbf{q}_1\|\|^2\right]\right]$	$\frac{B^2}{D}\left[\mu\phi_\eta^{4p-2} F + \mu^{-1} Q\right]$
L_3	$\frac{E\left[\|\|\mathbf{x}_1\|\|^2\right]}{D}\left[\mu\phi_\eta^{4p-2} E\left[\frac{1}{\|\|\mathbf{x}_1\|\|^2}\right] + \mu^{-1} E\left[\|\|\mathbf{q}_1\|\|^2\right]\right]$	$\frac{P}{D}\left[\mu\phi_\eta^{4p-2} F + \mu^{-1} Q\right]$

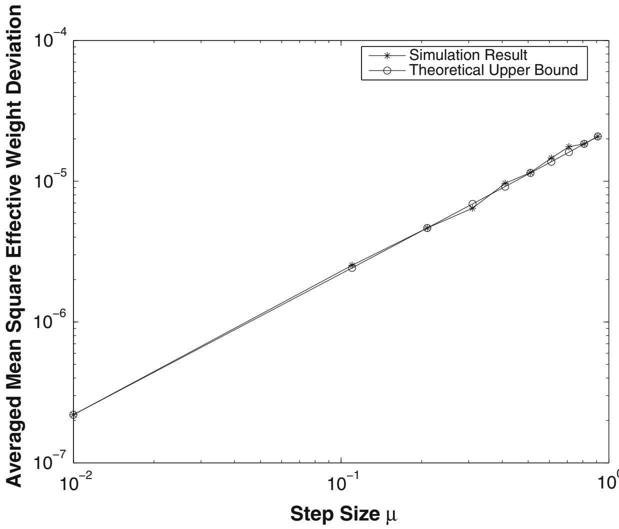

Fig. 3 Long-term average of the mean-squared effective weight deviation for constant channel identification, $N = 25$, Gaussian noise

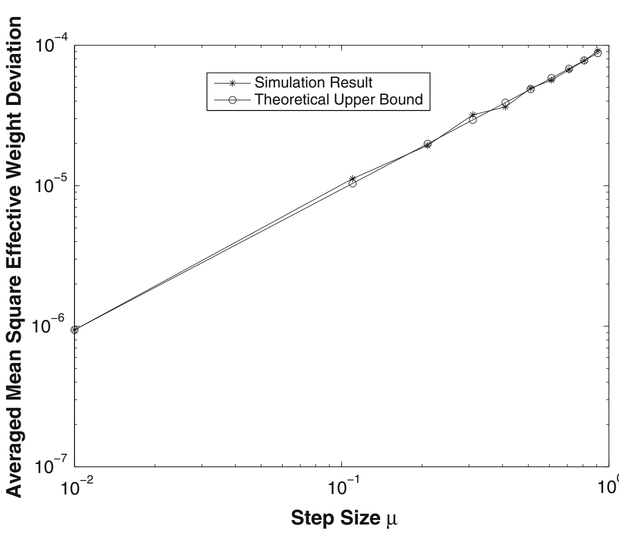

Fig. 4 Long-term average of the mean-squared effective weight deviation for constant channel identification, $N = 4$, uniform input and noise

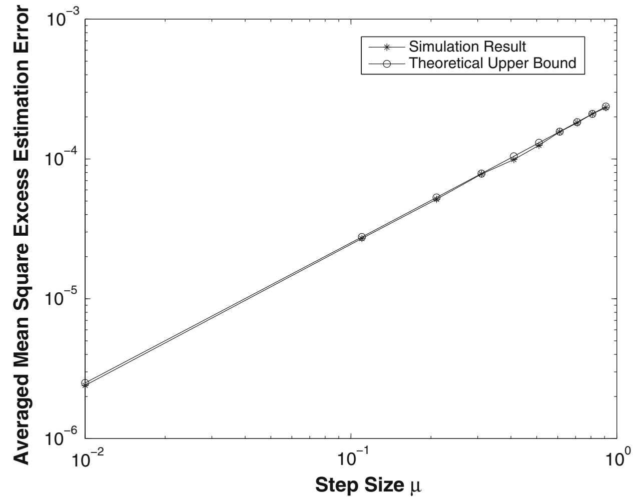

Fig. 5 Long-term average of mean-squared excess estimation error, $N = 4$, binary input, and Gaussian noise

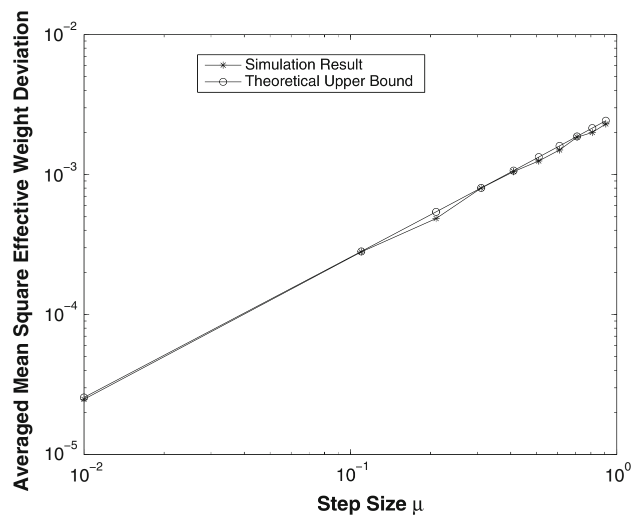

Fig. 6 Long-term average of mean-squared effective weight deviation for constant channel identification, $N = 4$, non-stationary input

In order to verify the analytical findings for non-stationary input, a non-stationary input sequence is generated via the following model:

$$x_k = x_{k-1} + w_k \tag{52}$$

where w_k is a zero mean unity variance i.i.d. sequence with variance σ_w^2. The value of σ_w^2 used in our experiments is equal to 0.01. With this non-stationary input, uniform noise, and filter length equal to 4, the analytical results given in Eqs. (32), (33), and (35) derived for non-stationary input scenario are investigated in Figs. 6, 7, and 8, respectively. It can be seen that simulation results very well substantiated the analytical findings showing validation of the analysis for non-stationary input also.

6.2 Tracking Random Walk Channel

In this section, simulation results are presented to validate the analytical results derived in Sect. 4. A random walk channel is generated using model Eq. (37). Figure 9 depicts the comparison between both the theoretical and simulation results of long-term average of mean-squared effective weight deviation for two different values of σ_q^2, i.e., $1e^{-6}$, and $1e^{-7}$. The filter length used is 4 while both input and noise are uniformly distributed. Analytical curve is obtained using Eq. (43). As can be seen in Fig. 9, close agreement between theory and simulation is obtained. It is observed from this figure that degradation in performance is obtained by increasing the value of σ_q^2. Also, unlike in the constant channel case, the long-term average of mean-squared effective weight devia-

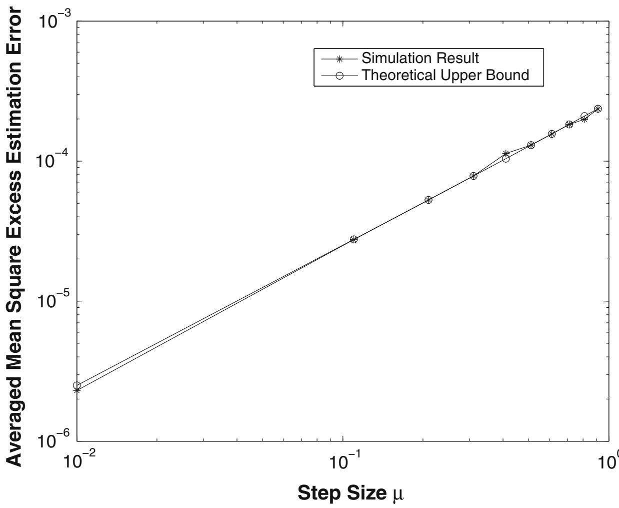

Fig. 7 Long-term average of mean-squared excess estimation error, $N = 4$, non-stationary input

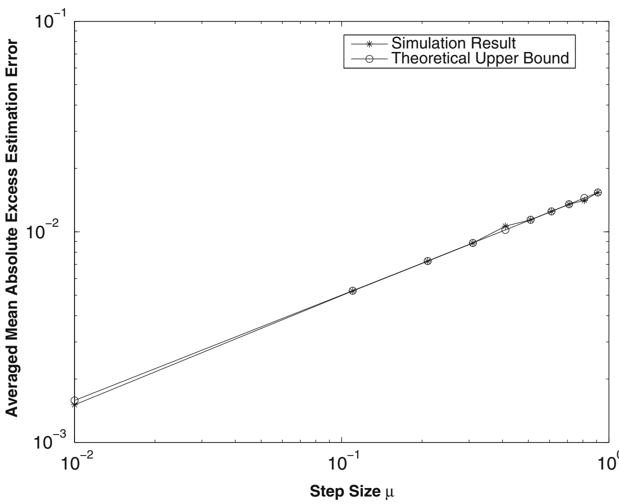

Fig. 8 Long-term average of mean absolute excess estimation error, $N = 4$, non-stationary input

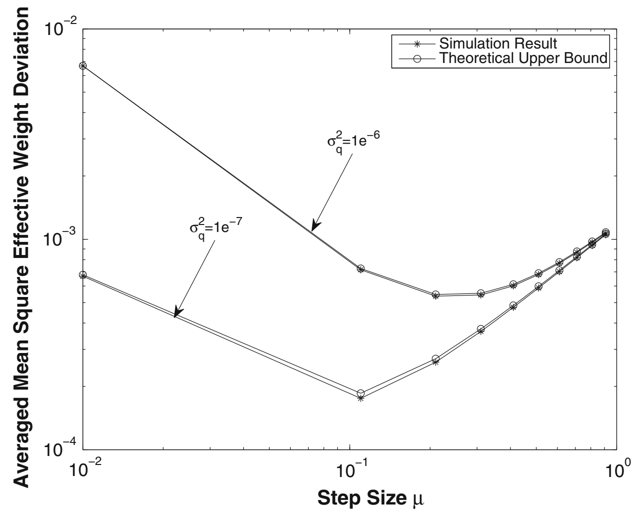

Fig. 9 Long-term average of mean-squared effective weight deviation for tracking random walk channel, $N = 4$, stationary uniform input

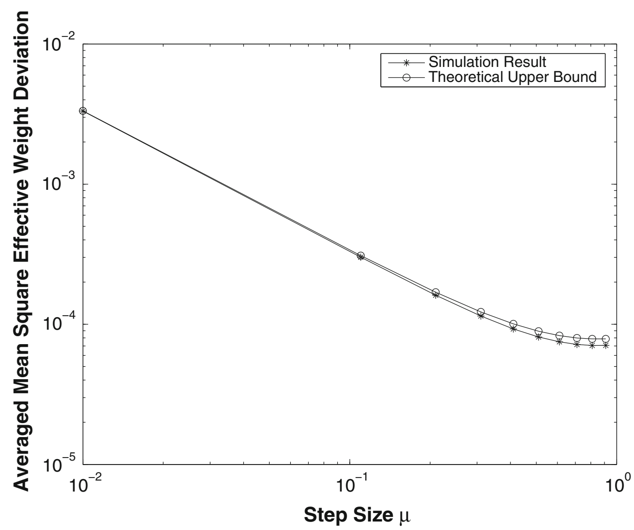

Fig. 10 Long-term average of mean-squared effective weight deviation for tracking random walk channel, $\sigma_q^2 = 1e^{-7}$, $N = 20$, stationary Gaussian input and noise

tion is not a monotonically decreasing function of the step size.

In Fig. 10, similar experiment is performed with a longer adaptive filter, that is, with a filter length equal to 20 while both input and noise are Gaussian distributed and $\sigma_q^2 = 1e^{-7}$. Here too, similar behavior is observed showing that our analytical results are not dependent on the adaptive filter's length and on the distribution of the filter input and the noise.

Next, the long-term average of mean-squared excess estimation error for tracking random walk channel with stationary uniform input and $\sigma_q^2 = 1e^{-7}$ is investigated in Fig. 11. Here, the analytical results are obtained by plotting Eq. (44). It can be easily seen from the result that there is a good match between simulation and the derived analytical results.

Finally, in order to test the validation of our analysis for non-stationary input, the long-term average of mean-squared effective weight deviation for a non-stationary uniform input [generated using Eq. (52)] and uniform noise with $\sigma_q^2 = 1e^{-7}$ is plotted in Fig. 12. Analytical result is obtained from the expression Eq. (47). Again, consistency in the results is obtained, which further validates our analytical findings.

7 Conclusion

In this work, a unified rigorous convergence analysis of the NLM algorithm is carried out using a newly proposed per-

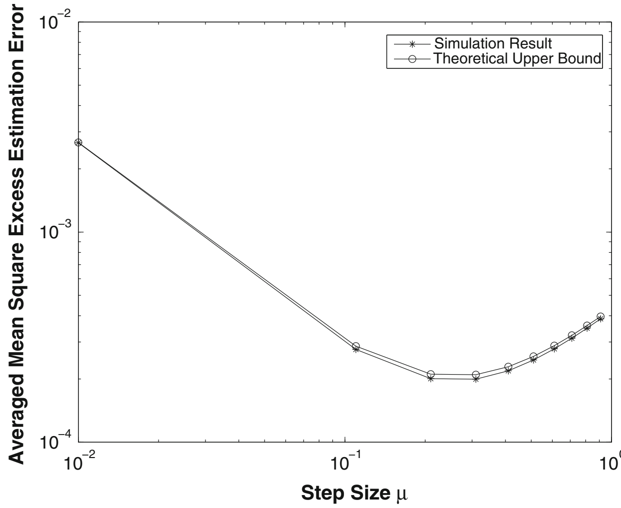

Fig. 11 Long-term average of mean-squared excess estimation error for tracking random walk channel, $N = 4$, $\sigma_q^2 = 1e^{-7}$, stationary uniform input

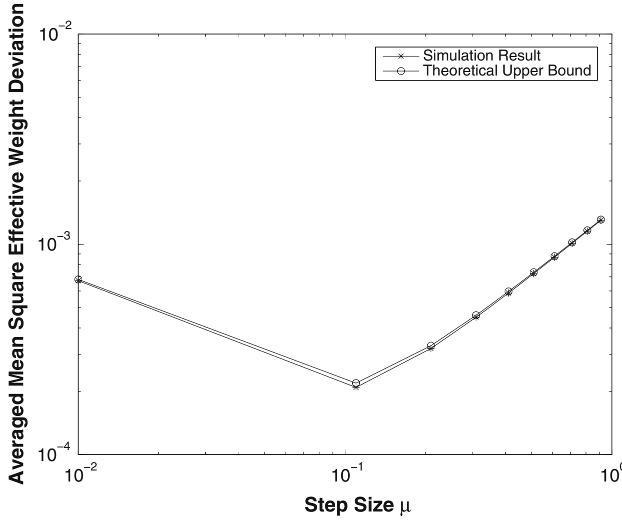

Fig. 12 Long-term average of mean-squared effective weight deviation for tracking random walk channel, $\sigma_q^2 = 1e^{-7}$, $N = 4$, non-stationary uniform input and noise

formance measure called effective weight deviation vector. Thus, the derived results are valid for any normalized stochastic gradient algorithm minimizing $2p$ power of the error. The analysis is rigorous in a sense that it is not restricted to any limitations on the dependence between input successive regressors, the dependence among input regressor elements, the length of the adaptive filter, the distribution of noise and filter's input, and the value of step size. Asymptotic time-averaged convergence for the mean-squared effective weight deviation, mean absolute excess estimation error, and the mean-square excess estimation error for the NLM algorithm is performed and consequently new explicit upper bounds for the long-term average of mean-squared effective weight deviation, mean-squared excess estimation error, and mean absolute excess estimation error are derived. Simulation results presented without imposing above-mentioned restrictions verified our theoretical findings.

Acknowledgments The first author acknowledges the support provided by King Abdulaziz University in carrying out this work. The second author would also like to acknowledge the support provided by KFUPM for funding this work through Project No. SB111012 and Project No. RG1216.

References

1. Sayed, A.H.: Fundamentals of Adaptive Filtering. Wiley-Interscience, New York (2003)
2. Nagumo, J.I.; Noda, A.: A learning method for system identification. IEEE Trans. Autom. Control **12**, 282–287 (1967)
3. Zerguine, A.: Convergence and steady-state analysis of the normalized Least Mean Fourth algorithm. Digit. Signal Process. **17**(1), 17–31 (2007)
4. Bershad, N.J.: Behavior of the ϵ-normalized LMS algorithm with Gaussian inputs. IEEE Trans. Acoust. Speech Signal Process. ASSP **35**(5), 636–644 (1987)
5. Kolodziej, J.E.; Tobias, O.J.; Seara, R.: An improved stochastic model of the NLMS algorithm for correlated input data. In: EUSIPCO 2007 (2007)
6. Barrault, G.; Costa, M.H.; Bermudez, J.C.M.; Lenzi, A.: A new analytical model for the NLMS algorithm. Proc. IEEE ICASSP **4**, 41–44 (2005)
7. Al-Naffouri, T.Y.; Sayed, A.H.: Transient analysis of data normalized adaptive filters. IEEE Trans. Signal Process. **51**(3), 639–652 (2003)
8. Douglas, S.C.; Meng, T.H.Y.: Normalized data nonlinearities for LMS Adaptation. IEEE Trans. Signal Process. **42**(6), 1352–1365 (1994)
9. Bermudez, J.C.M.; Costa, M.H.: A statistical analysis of the ϵ-NLMS and NLMS algorithms for correlated Gaussian signals. Revista da Sociedade Brasileira de Telecomunicaes **20**(2), 7–13 (2005)
10. Haykin, S.: Adaptive Filter Theory, 3rd edn. Prentice-Hall, Upper-Saddle River, NJ (1996)
11. Duttweiler, D.L.: Adaptive filter performance with nonlinearities in the correlation multiplier. IEEE Trans. Acoust. Speech Signal Process. **30**(4), 578–586 (1982)
12. Mathews, V.; Cho, S.: Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm. IEEE Trans. Acoust. Speech Signal Process. **35**(4), 450–454 (1987)
13. Bershad, N.J.: On error-saturation nonlinearities in LMS adaptation. IEEE Trans. Acoust. Speech Signal Process. **36**(4), 440–452 (1988)
14. Bershad, N.J.; Bonnet, M.: Saturation effects in LMS adaptive echo cancellation for binary data. IEEE Trans. Acoust. Speech Signal Process. **38**, 1687–1696 (1990)
15. Gibson, J.; Gray, S.: MVSE adaptive filtering subject to a constraint on MSE. IEEE Trans. Circuits Syst. **35**(5), 603–608 (1988)
16. Claasen, T.; Mecklenbräuker, W.: Comparison of the convergence of two algorithms for adaptive FIR digital filters. IEEE Trans. Circuits Syst. **28**(6), 510–518 (1981)
17. Gardner, W.A.: Learning characteristic of stochastic-descent algorithms: A general study, analysis, and critique. Signal Process. **6**(2), 113–133 (1984)

18. Rupp, M.: The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes. IEEE Trans. Signal Process. **41**(3), 1149–1160 (1993)
19. Koike, S.: Convergence analysis of a data echo canceler with a stochastic gradient adaptive FIR filter using the sign algorithm. IEEE Trans. Signal Process. **43**(12), 2852–2861 (1995)
20. Sharma, R.; Sethares, W.; Bucklew, J.: Asymptotic analysis of stochastic gradient-based adaptive filtering algorithms with general cost functions. IEEE Trans. Signal Process. **44**(9), 2186–2194 (1996)
21. Eweda, E.: A new approach for analyzing the limiting behavior of the normalized LMS algorithm under weak assumptions. Signal Process. **89**(11), 2143–2151 (2009)
22. Moinuddin, M.; Zerguine, A.: Steady-state analysis of the normalized least mean fourth algorithm without the independence and small step size assumptions. In: ICASSP 2009 (2009)