Signal Processing 128 (2016) 18–27

Contents lists available at ScienceDirect

Signal Processing

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

Stochastic behavior of the nonnegative least mean fourth algorithm for stationary Gaussian inputs and slow learning

Jingen Ni [a], Jian Yang [a], Jie Chen [b,*], Cédric Richard [c], José Carlos M. Bermudez [d]

[a] School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
[b] Center of Intelligent Acoustics and Immersive Communications (CIAIC), School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China
[c] The Université de Nice Sophia-Antipolis, UMR CNRS 7293, Observatoire de la Côte d'azur, Laboratoire Lagrange, Parc Valrose, 06102 Nice, France
[d] The Department of Electrical Engineering, Federal University of Santa Catarina, 88040-900 Florianópolis, SC, Brazil

ARTICLE INFO

Article history:
Received 19 November 2015
Received in revised form
26 March 2016
Accepted 29 March 2016
Available online 30 March 2016

Keywords:
Adaptive filter
Least mean fourth (LMF)
Mean weight behavior
Nonnegativity constraint
Second-order moment
System identification

ABSTRACT

Some system identification problems impose nonnegativity constraints on the parameters to be estimated due to inherent physical characteristics of the unknown system. The nonnegative least-mean-square (NNLMS) algorithm and its variants allow one to address this problem in an online manner. A nonnegative least mean fourth (NNLMF) algorithm has been recently proposed to improve the performance of these algorithms in cases where the measurement noise is not Gaussian. This paper provides a first theoretical analysis of the stochastic behavior of the NNLMF algorithm for stationary Gaussian inputs and slow learning. Simulation results illustrate the accuracy of the proposed analysis.

1. Introduction

Adaptive filtering algorithms are widely used to address system identification problems in applications such as adaptive noise cancellation [1,2], echo cancellation [3,4], active noise control [5,6], and distributed learning [7,8]. Due to physical characteristics, some problems require the imposition of nonnegativity constraints on the parameters to be estimated in order to avoid uninterpretable results [9]. Over recent decades, nonnegativity as a physical constraint has been studied extensively (see, e.g., nonnegative least-squares [10–13] and nonnegative matrix factorization [14–18]).

The nonnegative least-mean-square (NNLMS) algorithm was derived in [19] to address online system identification problems subject to nonnegativity constraints. Its convergence behavior was analyzed in [19,20]. The NNLMS algorithm is a fixed-point iteration scheme based on the Karush–Kuhn–Tucker (KKT) optimality conditions. The NNLMS algorithm updates the parameter estimates from streaming data at each time instant and is suitable for online system identification. Variants of the NNLMS algorithm were proposed in [21,22] to address specific robustness and convergence issues.

In certain practical contexts, it has been shown that adaptive algorithms with weight updates based on higher-order moments of the estimation error may have better mean-square error (MSE) convergence properties than the LMS algorithm [23–27]. This is the case, for instance, of the least mean fourth (LMF) algorithm, whose weight update is proportional to the third power of the estimation error. The LMF algorithm was proposed in [28], where it was verified that it could outperform the LMS algorithm in the presence of non-Gaussian measurement noise. This desirable property has led to a series of studies about the convergence behavior of the LMF algorithm and some of its variants [29–40]. Recently, a nonnegative LMF (NNLMF) algorithm was proposed in [41] to improve the performance of the NNLMS algorithm under non-Gaussian measurement noise. It was shown in [41] that, when compared to the NNLMS algorithm, the NNLMF algorithm can lead to faster convergence speed for equivalent steady-state performance or improved steady-state performance for the same convergence speed. The results shown in [41] are exclusively based on Monte Carlo simulation. Nevertheless, they clearly show that there is a need to better understand the convergence properties of the NNLMF algorithm. Up to now, there has been no study of the stochastic behavior of the NNLMF algorithm.

This paper provides a first statistical analysis of the NNLMF algorithm behavior. We derive an analytical model of the

* Corresponding author.
E-mail addresses: jni@suda.edu.cn (J. Ni), 20124228029@suda.edu.cn (J. Yang), dr.jie.chen@ieee.org (J. Chen), cedric.richard@unice.fr (C. Richard), j.bermudez@ieee.org (J.C.M. Bermudez).

http://dx.doi.org/10.1016/j.sigpro.2016.03.020
0165-1684/

algorithm behavior for slow learning and Gaussian inputs. Based on statistical assumptions typical in the analysis of adaptive algorithms, we derive recursive analytical expressions for the mean-weight behavior and for the excess MSE. The high-order nonlinearities imposed on the weight update by the third error power and by the non-negativity constraints result in a difficult mathematical analysis, requiring novel approximations not usually employed in adaptive filter analyses. Monte Carlo simulation results illustrate the accuracy of the analysis. It is known that the study of the stability of the LMF algorithm is relatively complex [30,31]. This complexity increases for the NNLMF algorithm. A theoretical stability study could not be accommodated in this work, and is left for future studies.

The paper is organized as follows. In Section 2, we describe the system model and provide an overview of the NNLMS and NNLMF algorithms. In Section 3, we introduce statistical assumptions used in the analysis of the NNLMF algorithm. The mean and mean-square analyses are performed in Sections 4 and 5, respectively. Simulation results are used to validate the analysis in Section 6. Finally, Section 7 concludes the paper.

In the sequel, normal font letters are used for scalars, boldface lowercase letters for vectors, and boldface uppercase letters for matrices. Furthermore, $(\cdot)^\top$ denotes vector or matrix transposition, $\mathrm{E}\{\cdot\}$ denotes statistical expectation, $\|\cdot\|$ is the ℓ_2-norm of a vector, $\circ$ denotes the Hadamard product, $\mathrm{Tr}\{\cdot\}$ computes the trace of a matrix, $\mathbf{D}_\alpha$ represents a diagonal matrix whose main diagonal is the vector $\boldsymbol{\alpha}$, $\mathbf{1}$ is the all-one column vector, and $\mathbf{I}$ is the identity matrix.

2. Nonnegative system identification

Online system identification aims at estimating the system impulse response from observations of both the input signal $u(n)$ and the desired response $d(n)$, as shown in Fig. 1. The desired response is assumed to be modeled by

$$d(n) = \mathbf{w}^{*\top}\mathbf{u}(n) + z(n) \tag{1}$$

where $\mathbf{u}(n) = [u(n), u(n-1), \ldots, u(n-M+1)]^\top$ is the input vector consisting of the M most recent input samples, $\mathbf{w}^* = [w_0^*, w_1^*, \ldots, w_{M-1}^*]^\top$ denotes the true weight vector of the unknown system, and $z(n)$ represents the measurement noise. For nonnegative system identification, nonnegativity constraints are imposed on the estimated weights w_i, $i \in \{0, 1, \ldots, M-1\}$, leading to the constrained optimization problem [9]

$$\mathbf{w}^o = \arg\min_{\mathbf{w}} J(\mathbf{w})$$
$$\text{subject to} \quad w_i \geq 0 \tag{2}$$

where $\mathbf{w}$ is the free-variable weight vector with w_i being its ith entry, $J(\mathbf{w})$ is a differentiable and strictly convex objective function

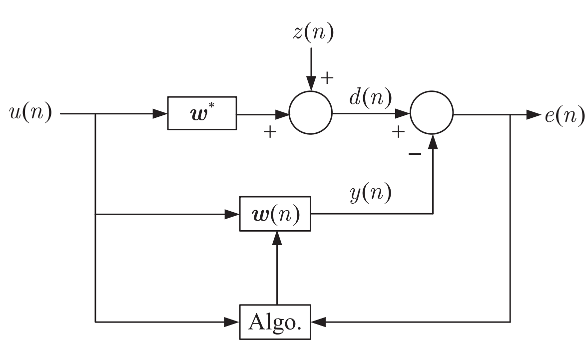


Fig. 1. Block diagram of system identification using an adaptive filter, which is widely used in many practical applications.

of $\mathbf{w}$, and $\mathbf{w}^o$ represents the solution to the above constrained optimization problem.

Based on the Karush–Kuhn–Tucker conditions, the authors in [19] derived a fixed-point iteration scheme to address the optimization problem (2). Using the mean square error (MSE) cost function

$$J[\mathbf{w}(n)] = \mathrm{E}\left\{[d(n) - \mathbf{w}^\top(n)\mathbf{u}(n)]^2\right\} \tag{3}$$

and a stochastic approximation yielded the NNLMS algorithm update equation

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{D}_{\mathbf{u}(n)}\mathbf{w}(n)e(n) \tag{4}$$

where $\mathbf{w}(n)$ denotes the weight vector of the adaptive filter at instant n, $e(n) = d(n) - \mathbf{w}^\top(n)\mathbf{u}(n)$ is the error signal, and μ is a positive step-size.

To improve the convergence performance of the adaptive filter for non-Gaussian measurement noise, the authors in [41] proposed to replace the MSE criterion with the mean fourth error (MFE) criterion

$$J[\mathbf{w}(n)] = \mathrm{E}\left\{[d(n) - \mathbf{w}^\top(n)\mathbf{u}(n)]^4\right\}. \tag{5}$$

This has led to the NNLMF algorithm update equation

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{D}_{\mathbf{u}(n)}\mathbf{w}(n)e^3(n) \tag{6}$$

The entry-wise form of (6) is

$$w_i(n+1) = w_i(n) + \mu u(n-i)w_i(n)e^3(n) \tag{7}$$

where $w_i(n)$ is the ith entry of the weight vector $\mathbf{w}(n)$. The update term of (7) is highly nonlinear in $\mathbf{w}(n)$, leading to a more complex behavior than that of the already studied LMF algorithm. In the following we study the stochastic behavior of (7).

3. Statistical assumptions

The analysis of any adaptive filtering algorithm requires the use of statistical assumptions for feasibility. The analysis is based on the study of the behavior of the weight-error vector, defined as

$$\tilde{\mathbf{w}}(n) = \mathbf{w}(n) - \mathbf{w}^* \tag{8}$$

and we employ the following frequently used statistical assumptions:

A1: The input signal $u(n)$ is stationary, zero-mean, and Gaussian.

A2: The input vector $\mathbf{u}(n)$ and the weight vector $\mathbf{w}(n)$ are independent.

A3: The measurement noise $z(n)$ is zero-mean, i.i.d., and independent of any other signal. Moreover, it has an even probability density function so that all odd moments of $z(n)$ are equal to zero.

A4: The statistical dependence between $\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)$ and $\tilde{\mathbf{w}}(n)$ can be neglected.

A5: The weight-error vector $\tilde{\mathbf{w}}(n)$ and $[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2$ are statistically independent.

Assumption A2 is the well-known independence assumption, which has been successfully used in the analysis of many adaptive algorithms, including the LMF algorithm [29]. Assumption A3 is often used in the analysis of higher-order moments in adaptive algorithms, and is practically reasonable. Assumption A4 is accurate for slow learning, as in this case weight fluctuations tend to be much smaller than their mean. For faster learning, this assumption becomes less valid, but has been found to provide an approximation with acceptable impact on the model accuracy. Assumption A5 is reasonable for a large number of taps. Simulation results will show that the models obtained using these assumptions can accurately predict the behavior of the NNLMF algorithm.

4. Mean weight analysis

The ith entry of the weight-error vector (8) is given by

$$\tilde{w}_i(n) = w_i(n) - w_i^*. \tag{9}$$

Subtracting w_i^* from both sides of (7), we have

$$\tilde{w}_i(n+1) = \tilde{w}_i(n) + \mu u(n-i) w_i(n) e^3(n). \tag{10}$$

Substituting (9) into (10) yields

$$\tilde{w}_i(n+1) = \tilde{w}_i(n) + \mu u(n-i)\left\{[\tilde{w}_i(n) + w_i^*]e^2(n)\right\}e(n). \tag{11}$$

By employing (1) and (9), the estimation error $e(n)$ can be equivalently expressed in terms of $\tilde{\mathbf{w}}(n)$ as

$$e(n) = z(n) + \mathbf{w}^{*\top}\mathbf{u}(n) - \mathbf{w}^\top(n)\mathbf{u}(n) = z(n) - \tilde{\mathbf{w}}^\top(n)\mathbf{u}(n). \tag{12}$$

From (12) the term $[\tilde{w}_i(n) + w_i^*]e^3(n)$ in (11) is of fourth-order in $\tilde{w}_i(n)$. This makes the analysis significantly more difficult than that of the LMS or LMF algorithm.

To make the problem tractable, we linearize the nonlinear term within brackets in (11):

$$f[\tilde{w}_i(n)] = [w_i^* + \tilde{w}_i(n)]e^2(n) \tag{13}$$

via a first-order Taylor expansion as done in [21]. Taking the derivative of $f[\tilde{w}_i(n)]$ with respect to $\tilde{w}_i(n)$, we have

$$\frac{\partial f[\tilde{w}_i(n)]}{\partial \tilde{w}_i(n)} = e^2(n) - 2e(n)u(n-i)[w_i^* + \tilde{w}_i(n)]. \tag{14}$$

Considering that $\tilde{w}_i(n)$ fluctuates around $\mathrm{E}\{\tilde{w}_i(n)\}$, we approximate the high-order stochastic term $f[\tilde{w}_i(n)]$ at time instant n by its first-order Taylor expansion about $\mathrm{E}\{\tilde{w}_i(n)\}$ as

$$\begin{aligned} f[\tilde{w}_i(n)] &\simeq f[\mathrm{E}\{\tilde{w}_i(n)\}] + \left.\frac{\partial f[\tilde{w}_i(n)]}{\partial \tilde{w}_i(n)}\right|_{\mathrm{E}\{\tilde{w}_i(n)\}}[\tilde{w}_i(n) - \mathrm{E}\{\tilde{w}_i(n)\}] \\ &= e_{\mathrm{E},i}^2(n)w_i^* + 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\mathrm{E}\{\tilde{w}_i(n)\} \\ &\quad + \{e_{\mathrm{E},i}^2(n) - 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\}\tilde{w}_i(n) \end{aligned} \tag{15}$$

where $e_{\mathrm{E},i}(n) = e(n)|_{\tilde{\mathbf{w}}_{\mathrm{E},i}(n)}$ and is expressed as

$$e_{\mathrm{E},i}(n) = z(n) - \tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n) = e(n) - u(n-i)[\mathrm{E}\{\tilde{w}_i(n)\} - \tilde{w}_i(n)] \tag{16}$$

with

$$\tilde{\mathbf{w}}_{\mathrm{E},i}(n) = [\tilde{w}_0(n), \ldots, \tilde{w}_{i-1}(n), \mathrm{E}\{\tilde{w}_i(n)\}, \tilde{w}_{i+1}(n), \ldots, \tilde{w}_{M-1}(n)]^\top. \tag{17}$$

Combining (11), (13), and (15) yields

$$\tilde{w}_i(n+1) = \tilde{w}_i(n) + \mu u(n-i)[x_i(n) + s_i(n)\tilde{w}_i(n)]e(n) \tag{18}$$

where

$$x_i(n) = e_{\mathrm{E},i}^2(n)w_i^* + 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\mathrm{E}\{\tilde{w}_i(n)\} \tag{19}$$

$$s_i(n) = e_{\mathrm{E},i}^2(n) - 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]. \tag{20}$$

Expressions (19) and (20) can be easily written in vector form as follows:

$$\begin{aligned} \mathbf{x}(n) &= [x_0(n), x_1(n), \ldots, x_{M-1}(n)]^\top \\ &= \mathbf{D}_{\mathbf{e}_{\mathrm{E}}(n)}^2\mathbf{w}^* + 2\mathbf{D}_{\mathbf{e}_{\mathrm{E}}(n)}\mathbf{D}_{\mathbf{u}(n)}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}[\mathbf{w}^* + \mathrm{E}\{\tilde{\mathbf{w}}(n)\}] \end{aligned} \tag{21}$$

$$\begin{aligned} \mathbf{s}(n) &= [s_0(n), s_1(n), \ldots, s_{M-1}(n)]^\top \\ &= \mathbf{D}_{\mathbf{e}_{\mathrm{E}}(n)}^2\mathbf{1} - 2\mathbf{D}_{\mathbf{e}_{\mathrm{E}}(n)}\mathbf{D}_{\mathbf{u}(n)}[\mathbf{w}^* + \mathrm{E}\{\tilde{\mathbf{w}}(n)\}] \end{aligned} \tag{22}$$

where

$$\begin{aligned} \mathbf{e}_{\mathrm{E}}(n) &= [e_{\mathrm{E},0}(n), e_{\mathrm{E},1}(n), \ldots, e_{\mathrm{E},M-1}(n)]^\top = e(n) \\ &\quad \mathbf{1} - \mathbf{D}_{\mathbf{u}(n)}[\mathrm{E}\{\tilde{\mathbf{w}}(n)\} - \tilde{\mathbf{w}}(n)]. \end{aligned} \tag{23}$$

Thus, we can write (18) in matrix form as

$$\begin{aligned} \tilde{\mathbf{w}}(n+1) &= \tilde{\mathbf{w}}(n) + \mu\mathbf{D}_{\mathbf{u}(n)}[\mathbf{x}(n) + \mathbf{D}_{\mathbf{s}(n)}\tilde{\mathbf{w}}(n)][z(n) - \tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)] = \tilde{\mathbf{w}}(n) \\ &\quad - \mu\mathbf{D}_{\mathbf{u}(n)}[\mathbf{x}(n) + \mathbf{D}_{\mathbf{s}(n)}\tilde{\mathbf{w}}(n)]\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n) + \mu\mathbf{D}_{\mathbf{u}(n)}[\mathbf{x}(n) + \mathbf{D}_{\mathbf{s}(n)}\tilde{\mathbf{w}}(n)] \\ &\quad z(n) = \tilde{\mathbf{w}}(n) - \mu\mathbf{p}(n) + \mu\mathbf{q}(n) \end{aligned} \tag{24}$$

where

$$\mathbf{p}(n) = \mathbf{D}_{\mathbf{u}(n)}[\mathbf{x}(n) + \mathbf{D}_{s(n)}\tilde{\mathbf{w}}(n)]\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n) \tag{25}$$

$$\mathbf{q}(n) = \mathbf{D}_{\mathbf{u}(n)}[\mathbf{x}(n) + \mathbf{D}_{s(n)}\tilde{\mathbf{w}}(n)]z(n). \tag{26}$$

Taking expectations of both sides of (24) yields

$$\mathrm{E}\{\tilde{\mathbf{w}}(n+1)\} = \mathrm{E}\{\tilde{\mathbf{w}}(n)\} - \mu\mathrm{E}\{\mathbf{p}(n)\} + \mu\mathrm{E}\{\mathbf{q}(n)\}. \tag{27}$$

Next, we need to calculate $\mathrm{E}\{\mathbf{p}(n)\}$ and $\mathrm{E}\{\mathbf{q}(n)\}$ to express (27) in an explicit form. From (25), the ith entry of $\mathbf{p}(n)$ can be written as

$$\begin{aligned} p_i(n) &= u(n-i)\Big\{e_{\mathrm{E},i}^2(n)w_i^* + 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\mathrm{E}\{\tilde{w}_i(n)\} \\ &\quad + \left\{e_{\mathrm{E},i}^2(n) - 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\right\}\tilde{w}_i(n)\Big\}\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n). \end{aligned} \tag{28}$$

We rewrite (28) as

$$p_i(n) = p_{i,\mathrm{a}}(n) + p_{i,\mathrm{b}}(n) + p_{i,\mathrm{c}}(n) \tag{29}$$

where

$$p_{i,\mathrm{a}}(n) = u(n-i)e_{\mathrm{E},i}^2(n)w_i^*\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n) \tag{30a}$$

$$p_{i,\mathrm{b}}(n) = u(n-i)e_{\mathrm{E},i}^2(n)\tilde{w}_i(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n) \tag{30b}$$

$$p_{i,\mathrm{c}}(n) = 2[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]u^2(n-i)e_{\mathrm{E},i}(n)[\mathrm{E}\{\tilde{w}_i(n)\} - \tilde{w}_i(n)]\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n). \tag{30c}$$

Define $\mathbf{r}_i$ as the ith column vector of the input vector correlation matrix $\mathbf{R} = \mathrm{E}\left\{\mathbf{u}(n)\mathbf{u}^\top(n)\right\}$. Using Assumptions A1–A5, it is shown in Appendices A and B that the expected values of (30a) and (30b) can be approximated by

$$\begin{aligned} \mathrm{E}\{p_{i,\mathrm{a}}(n)\} &= \mathrm{E}\left\{u(n-i)e_{\mathrm{E},i}^2(n)w_i^*\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\right\} \simeq 3\,\mathrm{Tr}\left\{\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\right\} \\ &\quad \mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\mathbf{r}_i w_i^* + \sigma_z^2\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\mathbf{r}_i w_i^* \end{aligned} \tag{31}$$

$$\begin{aligned} \mathrm{E}\{p_{i,\mathrm{b}}(n)\} &= \mathrm{E}\left\{u(n-i)e_{\mathrm{E},i}^2(n)\tilde{w}_i(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\right\} \simeq 3\,\mathrm{Tr}\left\{\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\right\} \\ &\quad \mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\mathbf{r}_i\mathrm{E}\{\tilde{w}_i(n)\} + \sigma_z^2\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\mathbf{r}_i\mathrm{E}\{\tilde{w}_i(n)\} \end{aligned} \tag{32}$$

where $\sigma_z^2 = \mathrm{E}\left\{z^2(n)\right\}$ denotes the variance of the measurement noise. It is also shown in Appendix A that $\tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n)$ can be approximated by $\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)$. Using this approximation in (16) yields

$$e_{\mathrm{E},i}(n) \simeq z(n) - \tilde{\mathbf{w}}^\top(n)\mathbf{u}(n). \tag{33}$$

Substituting (33) into (30c), we have

$$p_{i,\mathrm{c}}(n) = 2[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]u^2(n-i)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)[\mathrm{E}\{\tilde{w}_i(n)\} - \tilde{w}_i(n)]z(n) - 2[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]u^2(n-i)[\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)]^2[\mathrm{E}\{\tilde{w}_i(n)\} - \tilde{w}_i(n)]. \tag{34}$$

Then, using Assumptions A2, A3, and A5, the expected value of (34) becomes

$$\mathrm{E}\{p_{i,\mathrm{c}}(n)\} = -2[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\mathrm{E}\left\{u^2(n-i)[\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)]^2\right\} \mathrm{E}\left\{[\mathrm{E}\{\tilde{w}_i(n)\} - \tilde{w}_i(n)]\right\} = 0 \tag{35}$$

due to the last expectation. Therefore, we have

$$\mathrm{E}\{p_i(n)\} = \mathrm{E}\{p_{i,\mathrm{a}}(n)\} + \mathrm{E}\{p_{i,\mathrm{b}}(n)\}. \tag{36}$$

Its vector form is consequently given by

$$\mathrm{E}\{\mathbf{p}(n)\} = \mathrm{E}\{\mathbf{p}_\mathrm{a}(n)\} + \mathrm{E}\{\mathbf{p}_\mathrm{b}(n)\} \tag{37}$$

where

$$\mathrm{E}\{\mathbf{p}_\mathrm{a}(n)\} = 3\,\mathrm{Tr}\left\{\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\right\}\mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\} + \sigma_z^2\mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\} \tag{38}$$

$$\mathrm{E}\{\mathbf{p}_\mathrm{b}(n)\} = 3\,\mathrm{Tr}\left\{\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\right\}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\} + \sigma_z^2\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}. \tag{39}$$

Therefore, we have

$$\mathrm{E}\{\mathbf{p}(n)\} = 3\,\mathrm{Tr}\left\{\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\right\}\mathbf{D}_{[\mathbf{w}^*+\mathrm{E}\{\tilde{\mathbf{w}}(n)\}]}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\} + \sigma_z^2\mathbf{D}_{[\mathbf{w}^*+\mathrm{E}\{\tilde{\mathbf{w}}(n)\}]}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}. \tag{40}$$

From (26), we directly find that the ith entry of $\mathbf{q}(n)$ can be written as

$$q_i(n) = u(n-i)\Big\{e_{\mathrm{E},i}^2(n)w_i^* + 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\mathrm{E}\{\tilde{w}_i(n)\} + \left\{e_{\mathrm{E},i}^2(n) - 2e_{\mathrm{E},i}(n)u(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\right\}\tilde{w}_i(n)\Big\}z(n). \tag{41}$$

Using (16), we rewrite the above equation as

$$q_i(n) = q_{i,\mathrm{a}}(n) + q_{i,\mathrm{b}}(n) - q_{i,\mathrm{c}}(n) + q_{i,\mathrm{d}}(n) \tag{42}$$

where

$$q_{i,\mathrm{a}}(n) = u(n-i)[z(n) - \tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n)]^2 w_i^* z(n) \tag{43a}$$

$$q_{i,\mathrm{b}}(n) = u(n-i)[z(n) - \tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n)]^2 \tilde{w}_i(n)z(n) \tag{43b}$$

$$q_{i,\mathrm{c}}(n) = 2[z(n) - \tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n)]u^2(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\tilde{w}_i(n)z(n) \tag{43c}$$

$$q_{i,\mathrm{d}}(n) = 2[z(n) - \tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n)] u^2(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\mathrm{E}\{\tilde{w}_i(n)\}z(n). \tag{43d}$$

Since the mean weight behavior is not usually sensitive in the approximation $\mathrm{E}\{\tilde{w}_i(n)\tilde{w}_j(n)\} \simeq \mathrm{E}\{\tilde{w}_i(n)\}\mathrm{E}\{\tilde{w}_j(n)\}$, $\forall\, i, j$ (see [19,21]), using this approximation as well as Assumptions A1–A4, we have

$$\mathrm{E}\{q_{i,\mathrm{a}}(n)\} \simeq -2\mathrm{E}\left\{z^2(n)u(n-i)\tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n)\right\}w_i^* = -2\sigma_z^2 w_i^*\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\mathbf{r}_i \tag{44a}$$

$$\mathrm{E}\{q_{i,\mathrm{b}}(n)\} \simeq -2\mathrm{E}\left\{z^2(n)u(n-i)\tilde{\mathbf{w}}_{\mathrm{E},i}^\top(n)\mathbf{u}(n)\mathrm{E}\{\tilde{w}_i(n)\}\right\} = -2\sigma_z^2 w_i^*\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\mathbf{r}_i \tag{44b}$$

$$\mathrm{E}\{q_{i,\mathrm{c}}(n)\} \simeq 2\sigma_z^2\mathrm{E}\left\{u^2(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\right\}\mathrm{E}\{\tilde{w}_i(n)\} \tag{44c}$$

$$\mathrm{E}\{q_{i,\mathrm{d}}(n)\} = 2\sigma_z^2\mathrm{E}\left\{u^2(n-i)[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\right\}\mathrm{E}\{\tilde{w}_i(n)\}. \tag{44d}$$

Consequently, (42) leads to

$$\mathrm{E}\{q_i(n)\} = \mathrm{E}\{q_{i,\mathrm{a}}(n)\} + \mathrm{E}\{q_{i,\mathrm{b}}(n)\} - \mathrm{E}\{q_{i,\mathrm{c}}(n)\} + \mathrm{E}\{q_{i,\mathrm{d}}(n)\} \simeq -2\sigma_z^2[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\mathbf{r}_i \tag{45}$$

which can be written in matrix form as

$$\mathrm{E}\{\mathbf{q}(n)\} = -2\sigma_z^2\mathbf{D}_{[\mathbf{w}^*+\mathrm{E}\{\tilde{\mathbf{w}}(n)\}]}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}. \tag{46}$$

Finally, using (40) and (46) in (27), we obtain

$$\mathrm{E}\{\tilde{\mathbf{w}}(n+1)\} = \mathrm{E}\{\tilde{\mathbf{w}}(n)\} - 3\mu\sigma_z^2\mathbf{D}_{[\mathbf{w}^*+\mathrm{E}\{\tilde{\mathbf{w}}(n)\}]}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\} - 3\mu\mathrm{Tr}\left\{\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\right\}\mathbf{D}_{[\mathbf{w}^*+\mathrm{E}\{\tilde{\mathbf{w}}(n)\}]}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}. \tag{47}$$

Expression (47) predicts the mean weight behavior of the NNLMF algorithm, and will be used in the next section to compute the second-order moment.

In the case that the input to the system is zero-mean white noise, i.e., $\mathbf{R} = \sigma_u^2\mathbf{I}$ with $\sigma_u^2 = \mathrm{E}\left\{u^2(n)\right\}$, (47) reduces to

$$\begin{aligned}\mathrm{E}\{\tilde{\mathbf{w}}(n+1)\} &= \mathrm{E}\{\tilde{\mathbf{w}}(n)\} - 3\mu\sigma_z^2\sigma_u^2\mathbf{D}_{[\mathbf{w}^*+\mathrm{E}\{\tilde{\mathbf{w}}(n)\}]}\mathrm{E}\{\tilde{\mathbf{w}}(n)\} - 3\mu\sigma_u^4\mathrm{Tr}\left\{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}\right\} \mathbf{D}_{[\mathbf{w}^*+\mathrm{E}\{\tilde{\mathbf{w}}(n)\}]}\mathrm{E}\{\tilde{\mathbf{w}}(n)\} \\ &= \mathrm{E}\{\tilde{\mathbf{w}}(n)\} - 3\mu\sigma_z^2\sigma_u^2\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}[\mathbf{w}^* + \mathrm{E}\{\tilde{\mathbf{w}}(n)\}] - 3\mu\sigma_u^4 \parallel \mathrm{E}\left\{\tilde{\mathbf{w}}(n)\right\}\parallel^2 \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}[\mathbf{w}^* + \mathrm{E}\{\tilde{\mathbf{w}}(n)\}]\end{aligned} \tag{48}$$

which can be expressed in entry-wise form as

$$\mathrm{E}\{\tilde{w}_i(n+1)\} = \mathrm{E}\{\tilde{w}_i(n)\} - 3\mu\sigma_z^2\sigma_u^2\mathrm{E}\{\tilde{w}_i(n)\}[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}] - 3\mu\sigma_u^4\sum_{m=0}^{M-1}\mathrm{E}^2\{\tilde{w}_m(n)\}\mathrm{E}\{\tilde{w}_i(n)\}[w_i^* + \mathrm{E}\{\tilde{w}_i(n)\}]. \tag{49}$$

Using the equality $\mathrm{E}\{\tilde{w}_i(n+1)\} = \mathrm{E}\{\tilde{w}_i(n)\}$ as $n \to \infty$, (49) becomes

$$\mathrm{E}\{\tilde{w}_i(\infty)\}[w_i^* + \mathrm{E}\{\tilde{w}_i(\infty)\}]\left\{3\mu\sigma_z^2\sigma_u^2 + 3\mu\sigma_u^4\sum_{m=0}^{M-1}\mathrm{E}^2\{\tilde{w}_m(\infty)\}\right\} = 0. \tag{50}$$

Solving (50), we obtain $\mathrm{E}\{\tilde{w}_i(\infty)\} = 0$ or $\mathrm{E}\{\tilde{w}_i(\infty)\} = -w_i^*$, which means that $w_i^\mathrm{o} = w_i^*$ and $w_i^\mathrm{o} = 0$ are the two fixed points of the mean weight behavior of the NNLMF algorithm, where w_i^o is the ith entry of $\mathbf{w}^\mathrm{o}$. This result is consistent with that of the NNLMS algorithm.

5. Second-order moment analysis

Let $\mathbf{K}(n) = \mathrm{E}\left\{\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\right\}$ be the covariance matrix of the weight-error vector. Under certain simplifying assumptions [2], the excess mean-square error (EMSE) is given by

$$\xi(n) = \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}. \tag{51}$$

In the previous section, we used the approximation $\mathbf{K}(n) \simeq \mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\left\{\tilde{\mathbf{w}}^\top(n)\right\}$. This approximation is accurate enough for the analysis of the mean weight behavior. However, the effect of the second-order moments of the weight-error vector on the EMSE behavior is more significant [19]. Therefore, we need a more accurate expression for $\mathbf{K}(n)$ in order to characterize the EMSE.

Subtracting $\mathbf{w}_*$ from both sides of (6) yields

$$\tilde{\mathbf{w}}(n+1) = \tilde{\mathbf{w}}(n) + \mu \mathbf{D}_{\mathbf{u}(n)}\mathbf{w}(n)e^3(n). \tag{52}$$

Post-multiplying (52) by its transpose, considering the equality $\mathbf{D}_{\mathbf{u}(n)}\mathbf{w}(n) = \mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)$, and taking the expected value, we have

$$\mathbf{K}(n+1) = \mathbf{K}(n) + \mu\boldsymbol{\Phi}_1(n) + \mu^2\boldsymbol{\Phi}_2(n) \tag{53}$$

where

$$\boldsymbol{\Phi}_1(n) = \mathrm{E}\left\{e^3(n)[\tilde{\mathbf{w}}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)} + \mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\tilde{\mathbf{w}}^\top(n)]\right\} \tag{54}$$

$$\boldsymbol{\Phi}_2(n) = \mathrm{E}\left\{e^6(n)\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\}. \tag{55}$$

By using (12), $e^3(n)$ and $e^6(n)$ can be expanded, respectively, as follows:

$$e^3(n) = z^3(n) - 3z^2(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n) + 3z(n)[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2 - [\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^3 \tag{56}$$

$$e^6(n) = z^6(n) - 6z^5(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n) + 15z^4(n)[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2 - 20z^3(n)[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^3 + 15z^2(n)[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^4 - 6z(n)[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^5 + [\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^6. \tag{57}$$

Using (56) in (54) with Assumption A3, we have

$$\begin{aligned}\boldsymbol{\Phi}_1(n) &= E\left\{\left\{-3z^2(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n) - [\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^3\right\} \times [\tilde{\mathbf{w}}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)} + \mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\tilde{\mathbf{w}}^\top(n)]\right\} \\ &= -3\sigma_z^2\boldsymbol{\Theta}_1(n) - 3\sigma_z^2\boldsymbol{\Theta}_1^\top(n) + \boldsymbol{\Theta}_2(n) + \boldsymbol{\Theta}_2^\top(n)\end{aligned} \tag{58}$$

where

$$\boldsymbol{\Theta}_1(n) = \mathrm{E}\left\{\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\} \tag{59}$$

$$\boldsymbol{\Theta}_2(n) = -\,\mathrm{E}\left\{\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\right\}. \tag{60}$$

Similarly, using (57) in (55) with Assumption A3, we have

$$\boldsymbol{\Phi}_2(n) = \mathrm{E}\left\{z^6(n)\right\}\boldsymbol{\Theta}_3(n) + 15\mathrm{E}\left\{z^4(n)\right\}\boldsymbol{\Theta}_4(n) + 15\sigma_z^2\boldsymbol{\Theta}_5(n) + \boldsymbol{\Theta}_6(n) \tag{61}$$

where

$$\boldsymbol{\Theta}_3(n) = \mathrm{E}\left\{\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\} \tag{62}$$

$$\boldsymbol{\Theta}_4(n) = \mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\} \tag{63}$$

$$\boldsymbol{\Theta}_5(n) = \mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^4\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\} \tag{64}$$

$$\boldsymbol{\Theta}_6(n) = \mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^6\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\}. \tag{65}$$

In the following, we express $\boldsymbol{\Theta}_i(n)$, $i = 1, 2, \ldots, 6$, in terms of $\mathbf{R}$ and $\mathbf{K}(n)$.

5.1. $\boldsymbol{\Theta}_1(n)$

Using (8) in (59), and considering Assumptions A2 and A4, we can approximate (59) as

$$\begin{aligned}\boldsymbol{\Theta}_1(n) &= \mathrm{E}\left\{\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\tilde{\mathbf{w}}(n)}\right\} + \mathrm{E}\left\{\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}_*}\right\} \simeq \mathbf{K}(n) \\ &\quad \mathbf{R}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{K}(n)\mathbf{R}\mathbf{D}_{\mathbf{w}_*} = \mathbf{K}(n)\mathbf{R}[\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}_*}].\end{aligned} \tag{66}$$

5.2. $\boldsymbol{\Theta}_2(n)$

Likewise, using (8) in (60), we have

$$\begin{aligned}\boldsymbol{\Theta}_2(n) &= -\,\mathrm{E}\{\mathbf{D}_{\tilde{\mathbf{w}}(n)}\boldsymbol{\Xi}(n)\} - \mathrm{E}\{\mathbf{D}_{\mathbf{w}_*}\boldsymbol{\Xi}(n)\} \\ &\simeq -\,\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\mathrm{E}\{\boldsymbol{\Xi}(n)\} - \mathbf{D}_{\mathbf{w}_*}\mathrm{E}\{\boldsymbol{\Xi}(n)\}\end{aligned} \tag{67}$$

where

$$\boldsymbol{\Xi}(n) = \mathbf{u}(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n). \tag{68}$$

Notice that in the second line of (67) we neglect the correlation between $\mathbf{D}_{\tilde{\mathbf{w}}(n)}$ and $\boldsymbol{\Xi}(n)$. This approximation is reasonable as $\boldsymbol{\Xi}(n)$ is a function of fourth-order products of elements of $\tilde{\mathbf{w}}(n)$, whose values can be obtained from infinitely many different vectors $\tilde{\mathbf{w}}(n)$. In [29], the expected value of $\boldsymbol{\Xi}(n)$ for zero-mean Gaussian inputs has been approximated using Assumptions A4 and A5 as

$$\mathrm{E}\{\boldsymbol{\Xi}(n)\} \simeq 3\,\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{R}\mathbf{K}(n). \tag{69}$$

Using (69) in (67) yields

$$\boldsymbol{\Theta}_2(n) \simeq -\,3\,\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}[\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}_*}]\mathbf{R}\mathbf{K}(n). \tag{70}$$

5.3. $\boldsymbol{\Theta}_3(n)$

Substituting (8) into (62) and using Assumption A3, we have

$$\begin{aligned}\boldsymbol{\Theta}_3(n) &= \mathrm{E}\left\{\mathbf{D}_{\tilde{\mathbf{w}}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\tilde{\mathbf{w}}(n)}\right\} \\ &\quad + \mathrm{E}\left\{\mathbf{D}_{\tilde{\mathbf{w}}}(n)\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}_*}\right\} + \mathrm{E}\left\{\mathbf{D}_{\mathbf{w}_*}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\tilde{\mathbf{w}}}(n)\right\} \\ &\quad + \mathrm{E}\left\{\mathbf{D}_{\mathbf{w}_*}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}_*}\right\}.\end{aligned} \tag{71}$$

It was shown in [19] that $\mathrm{E}\left\{\mathbf{D}_{\mathbf{u}(n)}\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{D}_{\mathbf{u}(n)}\right\} \simeq \mathbf{R}\bigcirc\mathbf{K}(n)$. Since

$$\mathbf{D}_{\tilde{\mathbf{w}}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\tilde{\mathbf{w}}(n)} = \mathbf{D}_{\mathbf{u}(n)}\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{D}_{\mathbf{u}(n)} \tag{72}$$

we can approximate (71) as

$$\boldsymbol{\Theta}_3(n) \simeq \mathbf{R}\bigcirc\mathbf{K}(n) + \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\mathbf{R}\mathbf{D}_{\mathbf{w}_*} + \mathbf{D}_{\mathbf{w}_*}\mathbf{R}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}_*}\mathbf{R}\mathbf{D}_{\mathbf{w}_*}. \tag{73}$$

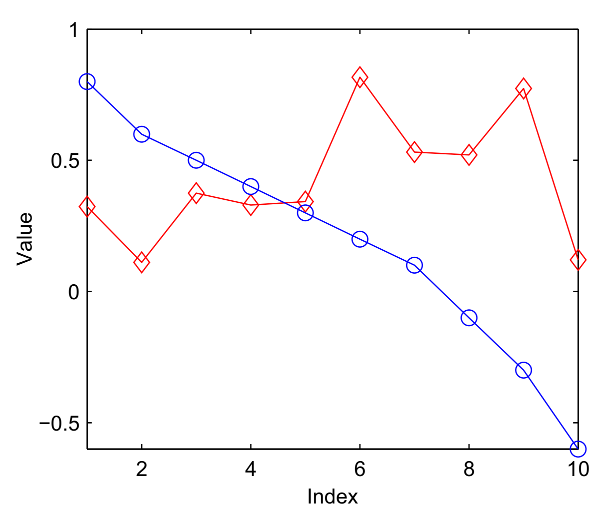
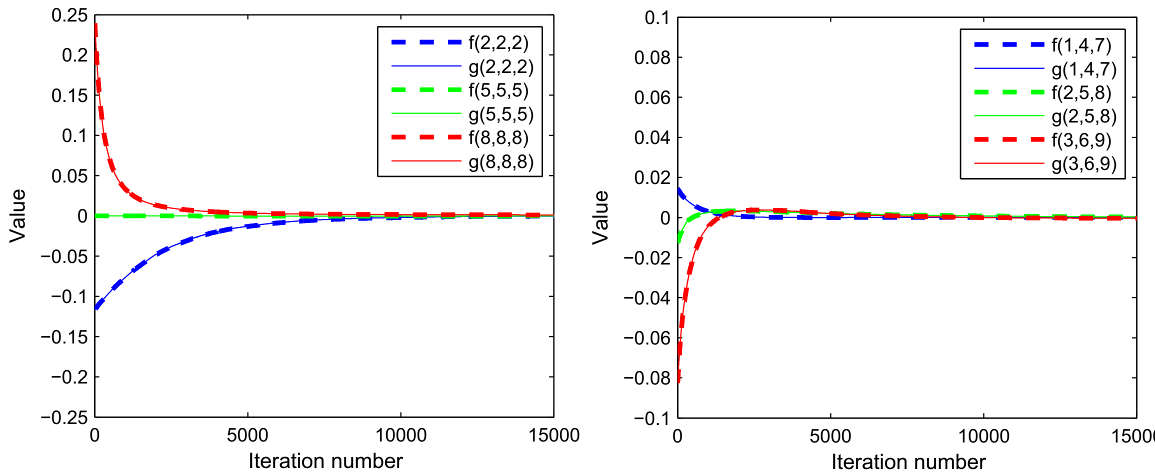


Fig. 2. Unknown system impulse response $\mathbf{w}_*$ (○) and initial weight vector $\mathbf{w}(0)$ (◇) drawn from the uniform distribution $U([0;\,1])$.



(a) (b)

Fig. 3. Plots of $f(i,j,k) = \mathrm{E}\{\tilde{w}_i(n)\tilde{w}_j(n)\tilde{w}_k(n)\}$ and $g(i,j,k) = \mathrm{E}\{\tilde{w}_i(n)\tilde{w}_j(n)\}\mathrm{E}\{\tilde{w}_k(n)\}$ in the case where the input is a correlated signal and the noise is a uniformly distributed sequence with $\mu = 2 \times 10^{-5}$, showing that Assumption A4 is valid for slow learning. (a) Weight indices $i = j = k$ and (b) weight indices $i \neq j \neq k$.

5.4. $\boldsymbol{\Theta}_4(n)$

Substituting (8) into (63) and using Assumption A3, (63) can be written as

$$\begin{aligned}\boldsymbol{\Theta}_4(n) &= 15\mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\}\\ &= 15\mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\mathbf{D}_{\mathbf{u}(n)}\mathbf{w}(n)\mathbf{w}^\top(n)\mathbf{D}_{\mathbf{u}(n)}\right\}\\ &= 15\mathrm{E}\left\{\mathbf{D}_{\mathbf{u}(n)}[\tilde{\mathbf{w}}(n) + \mathbf{w}^*]\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\right.\\ &\qquad \left.[\tilde{\mathbf{w}}(n) + \mathbf{w}^*]^\top\mathbf{D}_{\mathbf{u}(n)}\right\}\\ &= 15\{\boldsymbol{\Theta}_{4.1}(n) + \boldsymbol{\Theta}_{4.2}(n) + \boldsymbol{\Theta}_{4.3}(n) + \boldsymbol{\Theta}_{4.4}(n)\}\end{aligned} \tag{74}$$

where

$$\boldsymbol{\Theta}_{4.1}(n) = \mathrm{E}\left\{\mathbf{D}_{\mathbf{u}(n)}\mathbf{w}^*\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{w}^{*\top}\mathbf{D}_{\mathbf{u}(n)}\right\} \tag{75a}$$

$$\boldsymbol{\Theta}_{4.2}(n) = \mathrm{E}\left\{\mathbf{D}_{\mathbf{u}(n)}\mathbf{w}^*\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{D}_{\mathbf{u}(n)}\right\} \tag{75b}$$

$$\boldsymbol{\Theta}_{4.3}(n) = \mathrm{E}\left\{\mathbf{D}_{\mathbf{u}(n)}\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{w}^{*\top}\mathbf{D}_{\mathbf{u}(n)}\right\} \tag{75c}$$

$$\boldsymbol{\Theta}_{4.4}(n) = \mathrm{E}\left\{\mathbf{D}_{\mathbf{u}(n)}\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{u}(n)\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^\top(n)\mathbf{D}_{\mathbf{u}(n)}\right\}. \tag{75d}$$

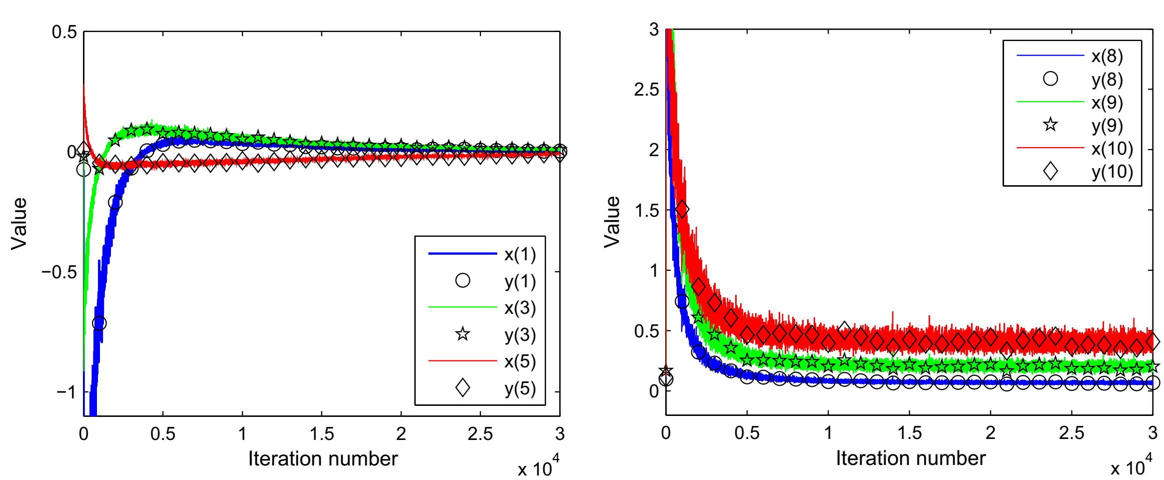
We find that the above quantities, $\boldsymbol{\Theta}_{4.1}(n)$–$\boldsymbol{\Theta}_{4.4}(n)$, correspond to [19, Eqs. (45)–(48)], respectively, which have been computed under Assumptions A1–A5. Therefore, the results obtained in [19] can be used directly here, yielding

$$\boldsymbol{\Theta}_{4.1}(n) \simeq \mathbf{D}_{\mathbf{w}^*}\{2\mathbf{R}\mathbf{K}(n)\mathbf{R} + \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{R}\}\mathbf{D}_{\mathbf{w}^*} \tag{76a}$$

$$\boldsymbol{\Theta}_{4.2}(n) \simeq \mathbf{D}_{\mathbf{w}^*}\left\{2\mathbf{R}\mathbf{K}(n)\mathbf{R} + \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{R}\right\}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} \tag{76b}$$

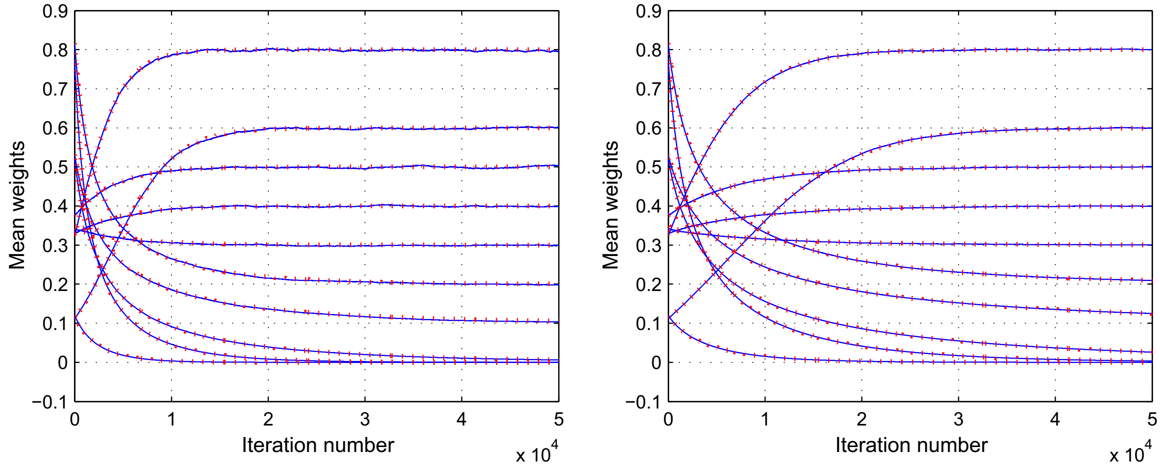
$$\boldsymbol{\Theta}_{4.3}(n) \simeq \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\left\{2\mathbf{R}\mathbf{K}(n)\mathbf{R} + \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{R}\right\}\mathbf{D}_{\mathbf{w}^*} \tag{76c}$$

$$\boldsymbol{\Theta}_{4.4}(n) \simeq \{2\mathbf{R}\mathbf{K}(n)\mathbf{R} + \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{R}\}\bigcirc\mathbf{K}(n). \tag{76d}$$



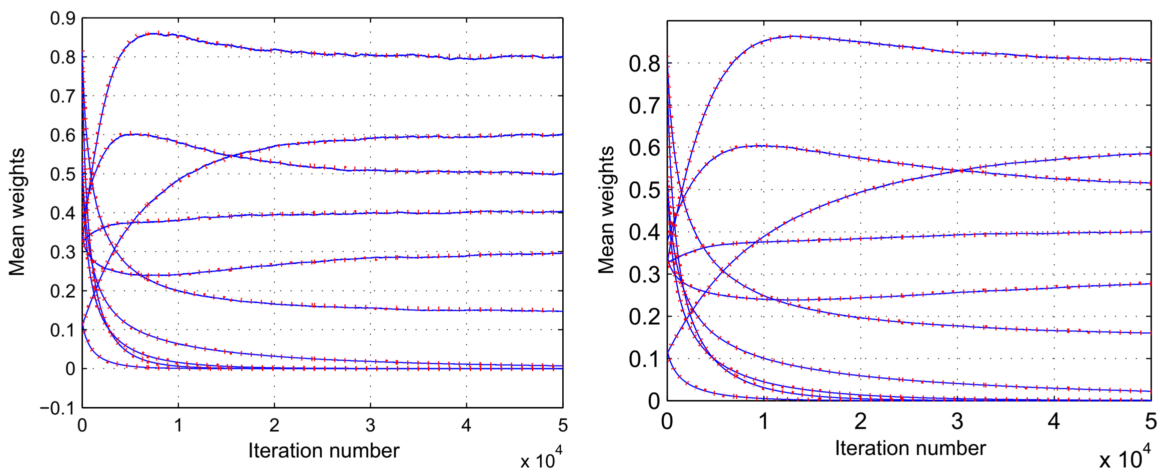
(a) (b)

Fig. 4. Plots of $x(i) = \mathrm{E}\{\tilde{w}_i(n)[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\}$ and $y(i) = \mathrm{E}\{\tilde{w}_i(n)\}\mathrm{E}\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\}$ in the case where the input is a correlated signal and the noise is a uniformly distributed sequence with $\mu = 2 \times 10^{-5}$, showing that Assumption A5 is valid for slow learning. (a) Positive unknown weights and (b) negative unknown weights.



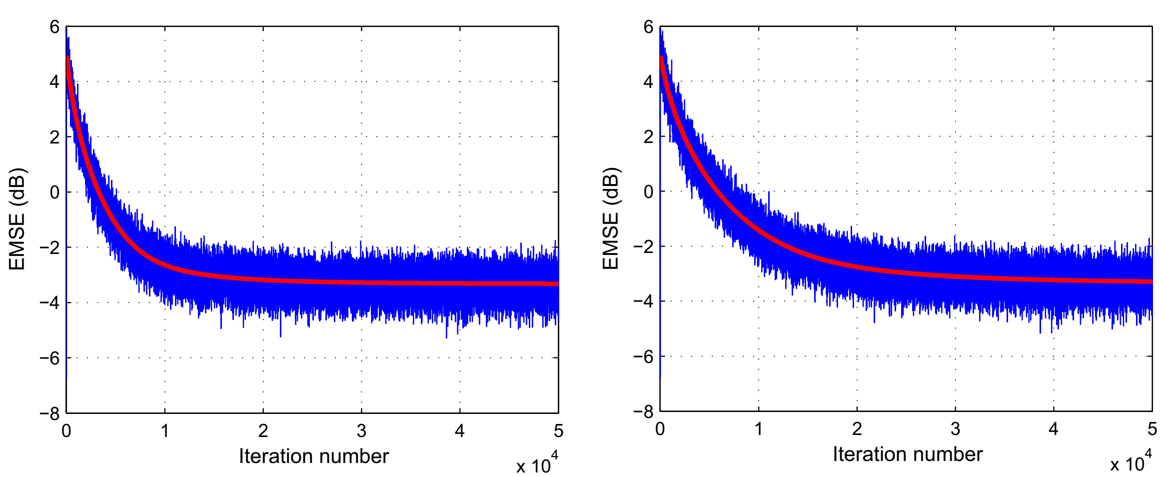
(a) (b)

Fig. 5. Convergence of the NNLMF mean weights in the case where the input is a white Gaussian signal with $\mu = 2 \times 10^{-5}$, showing that the theoretical (dotted) and simulation (solid) curves are perfectly superimposed. (a) Uniform noise and (b) binary noise.



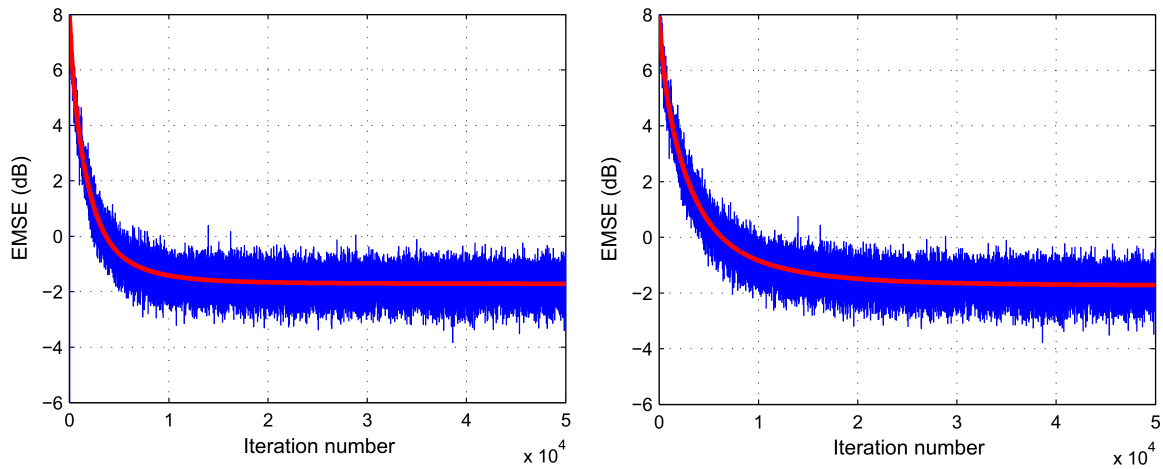
(a) (b)

Fig. 6. Convergence of the NNLMF mean weights in the case where the input is a correlated signal with $\mu = 2 \times 10^{-5}$, showing that the theoretical (dotted) and simulation (solid) curves are perfectly superimposed. (a) Uniform noise and (b) binary noise.



(a) (b)

Fig. 7. Convergence of the NNLMF second-order moment in the case where the input is a white Gaussian signal with $\mu = 2 \times 10^{-5}$, showing that the theoretical (solid) curves match well with the simulation results (dashed). (a) Uniform noise and (b) binary noise.



(a) (b)

Fig. 8. Convergence of the NNLMF second-order moment in the case where the input is a correlated signal with $\mu = 2 \times 10^{-5}$, showing that the theoretical (solid) curves match well with the simulation results (dashed). (a) Uniform noise and (b) binary noise.

Defining

$$\boldsymbol{\Upsilon}(n) = 2\mathbf{R}\mathbf{K}(n)\mathbf{R} + \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{R} \tag{77}$$

and substituting (76) into (74) leads to

$$\boldsymbol{\Theta}_4(n) \simeq [\mathbf{D}_{\mathbf{w}^*}\boldsymbol{\Upsilon}(n)\mathbf{D}_{\mathbf{w}^*} + \mathbf{D}_{\mathbf{w}^*}\boldsymbol{\Upsilon}(n)\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\boldsymbol{\Upsilon}(n)\mathbf{D}_{\mathbf{w}^*} + \boldsymbol{\Upsilon}(n)\bigcirc\mathbf{K}(n)]. \tag{78}$$

5.5. $\boldsymbol{\Theta}_5(n)$

The term $\boldsymbol{\Theta}_5(n)$ contains higher-order moments of $\tilde{\mathbf{w}}(n)$ and $\mathbf{u}(n)$. Computing this term also requires approximations. One approximation that preserves the second-order moments is to split the expectation as $\mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^4\right\}$ $\mathrm{E}\left\{\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\}$. The intuition behind this approximation is that each element of the matrix $\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}$ corresponds to only one of the M^2 terms of the sum $[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2$, which tends to reduce their correlation for reasonably large M. Moreover, we shall assume that $\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)$ is zero-mean Gaussian to simplify the evaluation of the above expectations. This assumption becomes more valid as M increases (by the Central Limit theorem) and tends to be reasonable for practical values of M. Under these assumptions, we have

$$\begin{aligned}\boldsymbol{\Theta}_5(n) &\simeq \mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^4\right\}\mathrm{E}\left\{\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\} = 3\left(\mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\right\}\right)^2 \boldsymbol{\Theta}_3(n) = 3[\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}]^2[\mathbf{R}\bigcirc\mathbf{K}(n) \\ &+ \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\mathbf{R}\mathbf{D}_{\mathbf{w}^*} + \mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathbf{D}_{\mathbf{w}^*}].\end{aligned} \tag{79}$$

5.6. $\boldsymbol{\Theta}_6(n)$

The same assumptions used to calculate $\boldsymbol{\Theta}_5(n)$ can be applied to approximate (65) as

$$\begin{aligned}\boldsymbol{\Theta}_6(n) &\simeq \mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^6\right\}\mathrm{E}\left\{\mathbf{D}_{\mathbf{w}(n)}\mathbf{u}(n)\mathbf{u}^\top(n)\mathbf{D}_{\mathbf{w}(n)}\right\} = 15\left(\mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\right\}\right)^3 \boldsymbol{\Theta}_3(n) = 15[\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}]^3[\mathbf{R}\bigcirc\mathbf{K}(n) \\ &+ \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\mathbf{R}\mathbf{D}_{\mathbf{w}^*} + \mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathbf{D}_{\mathbf{w}^*}].\end{aligned} \tag{80}$$

5.7. $\mathbf{K}(n)$

Using $\boldsymbol{\Theta}_1(n)$ through $\boldsymbol{\Theta}_6(n)$ in (58) and (61), we obtain

$$\begin{aligned}\boldsymbol{\Phi}_1(n) &\simeq -3[\sigma_z^2 + \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}] \\ &\times \left\{\mathbf{K}(n)\mathbf{R}[\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}^*}] + [\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}^*}]\mathbf{R}\mathbf{K}(n)\right\}\end{aligned} \tag{81}$$

and

$$\begin{aligned}\boldsymbol{\Phi}_2(n) &\simeq \left\{\mathrm{E}\left\{z^6(n)\right\} + 45\sigma_z^2[\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}]^2 + 15[\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}]^3\right\} \\ &\times \left\{\mathbf{R}\bigcirc\mathbf{K}(n) + \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\mathbf{R}\mathbf{D}_{\mathbf{w}^*} + \mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathbf{w}^*}\mathbf{R}\mathbf{D}_{\mathbf{w}^*}\right\} + 15\mathrm{E}\left\{z^4(n)\right\} \\ &[\mathbf{D}_{\mathbf{w}^*}\boldsymbol{\Upsilon}(n)\mathbf{D}_{\mathbf{w}^*} + \mathbf{D}_{\mathbf{w}^*}\boldsymbol{\Upsilon}(n)\mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}} + \mathbf{D}_{\mathrm{E}\{\tilde{\mathbf{w}}(n)\}}\boldsymbol{\Upsilon}(n)\mathbf{D}_{\mathbf{w}^*} + \boldsymbol{\Upsilon}(n)\bigcirc\mathbf{K}(n)].\end{aligned} \tag{82}$$

Substituting (81) and (82) into (53), we obtain a recursive analytical model for the behavior of $\mathbf{K}(n)$, which can then be used in (51) to predict the EMSE behavior for the NNLMF algorithm. Note that $\mathrm{E}\left\{z^4(n)\right\}$ and $\mathrm{E}\left\{z^6(n)\right\}$ depend on the statistical distribution of the noise $z(n)$. For instance, if $z(n)$ is zero-mean Gaussian noise, then $\mathrm{E}\left\{z^4(n)\right\} = 3\sigma_z^4$ and $\mathrm{E}\left\{z^6(n)\right\} = 15\sigma_z^6$; if $z(n)$ is zero-mean uniform noise, then $\mathrm{E}\left\{z^4(n)\right\} = 9/5\sigma_z^4$ and $\mathrm{E}\left\{z^6(n)\right\} = 27/7\sigma_z^6$; if $z(n)$ is zero-mean binary noise, then $\mathrm{E}\left\{z^4(n)\right\} = \sigma_z^4$ and $\mathrm{E}\left\{z^6(n)\right\} = \sigma_z^6$.

6. Simulation results

This section presents simulations in the context of system identification with nonnegativity constraints to illustrate the accuracy of the models derived in Sections 4 and 5. The impulse response $\mathbf{w}^*$ of the unknown system is given by $[0.8, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, -0.1, -0.3, -0.6]^\top$. The initial weight $\mathbf{w}(0)$ is a vector drawn from the uniform distribution $U([0;1])$ and kept the same for all realizations. Both $\mathbf{w}^*$ and $\mathbf{w}(0)$ are shown in Fig. 2. The input is taken as either a zero-mean white Gaussian signal of unit power or a correlated signal obtained by filtering zero-mean white Gaussian noise with variance 3/4 through a first-order system $H(z) = 1/(1 - 0.5z^{-1})$, which yields a correlated input with unit variance. The above simulation setups are the same as those in [19]. The measurement noise is either uniformly distributed sequence in $[-5, 5]$ (SNR = −9.2 dB) or a binary sequence with samples randomly drawn from the set $\{2, -2\}$ (SNR = −6 dB). The step-size is chosen as $\mu = 2 \times 10^{-5}$ for slow learning. All simulated curves are obtained by averaging over 200 realizations.

We first evaluate Assumptions A4 and A5 by simulation. Assumption A4 can be described as $\mathrm{E}\{\tilde{w}_i(n)\tilde{w}_j(n)\tilde{w}_k(n)\} \simeq \mathrm{E}\{\tilde{w}_i(n)\tilde{w}_j(n)\}\mathrm{E}\{\tilde{w}_k(n)\}$. From Fig. 3, one can see that this approximation is valid. Likewise, Assumption A5 can be described as $\mathrm{E}\left\{\tilde{w}_i(n)[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\right\} \simeq \mathrm{E}\{\tilde{w}_i(n)\}\mathrm{E}\left\{[\mathbf{u}^\top(n)\tilde{\mathbf{w}}(n)]^2\right\}$. Fig. 4 shows that Assumption A5 is also valid. Figs. 5 and 6 show the mean weight behavior for white and correlated inputs, respectively. An excellent match can be verified between the behavior predicted by the proposed model and that obtained from Monte Carlo simulation. Figs. 7 and 8 show the EMSE behavior (in dB) for the same example. Here again one can verify an excellent match between theory and simulation results.

7. Conclusions

The NNLMF algorithm can outperform the NNLMS algorithm when the measurement noise is non-Gaussian. This paper studied the mean and second-moment behavior of the NNLMF algorithm for stationary Gaussian input signals and slow learning. The analysis was based on typical statistical assumptions and has led to a recursive model for predicting the algorithm behavior. Simulation results have shown an excellent match between the simulation results and the predicted behavior from theoretical models. Since stability conditions for convergence of the NNLMF algorithm are difficult to determine analytically, a theoretical stability analysis will be a topic for future work.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 61471251 and 61101217 and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20131164.

Appendix A. Detailed calculation of (31)

Substituting (16) into the first line of (31) yields

$$\begin{aligned}\mathrm{E}\{p_{i,a}(n)\} &= \mathrm{E}\Big\{u(n-i)[z(n)-\tilde{\mathbf{w}}_{\mathrm{E},i}^{\top}(n)\mathbf{u}(n)]^2 w_i^* \tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)\Big\}\\ &= \mathrm{E}\Big\{u(n-i)[\tilde{\mathbf{w}}_{\mathrm{E},i}^{\top}(n)\mathbf{u}(n)]^2 w_i^* \tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)\Big\}\\ &\quad - 2\mathrm{E}\Big\{u(n-i)z(n)\tilde{\mathbf{w}}_{\mathrm{E},i}^{\top}(n)\mathbf{u}(n)w_i^*\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)\Big\}\\ &\quad + \mathrm{E}\Big\{u(n-i)z^2(n)w_i^*\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)\Big\}. \end{aligned} \tag{83}$$

Using Assumptions A2 and A3, and noting that w_i^* is deterministic, we can simplify (83) to

$$\begin{aligned}\mathrm{E}\{p_{i,a}(n)\} &= \mathrm{E}\Big\{u(n-i)[\tilde{\mathbf{w}}_{\mathrm{E},i}^{\top}(n)\mathbf{u}(n)]^2 \tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)\Big\}\\ &\quad w_i^* + \sigma_z^2 \mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i w_i^*. \end{aligned} \tag{84}$$

Using (17) and the definition of $\mathbf{u}(n)$, we obtain

$$\tilde{\mathbf{w}}_{\mathrm{E},i}^{\top}(n)\mathbf{u}(n) = \sum_{j=0}^{i-1}\tilde{w}_j(n)u(n-j) + \mathrm{E}\{\tilde{w}(n-i)\}u(n-i) + \sum_{j=i+1}^{M-1}\tilde{w}_j(n)u(n-j). \tag{85}$$

Also,

$$\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n) = \sum_{j=0}^{i-1}\tilde{w}_j(n)u(n-j) + \tilde{w}(n-i)u(n-i) + \sum_{j=i+1}^{M-1}\tilde{w}_j(n)u(n-j). \tag{86}$$

We note that (85) has the same expression as (86) except for the ith term, $\mathrm{E}\{\tilde{w}(n-i)\}u(n-i)$. Therefore, we can approximate $\tilde{\mathbf{w}}_{\mathrm{E},i}^{\top}(n)\mathbf{u}(n)$ by $\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)$ for large values of M. With this approximation, (84) can be written as

$$\mathrm{E}\{p_{i,a}(n)\} \simeq \mathrm{E}\Big\{u(n-i)[\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)]^3\Big\}w_i^* + \sigma_z^2\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i w_i^*. \tag{87}$$

The following approximation has been derived in [29]:

$$\mathrm{E}\Big\{\mathbf{u}(n)[\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)]^3\Big\} \simeq 3\,\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}. \tag{88}$$

The ith entry of (88) satisfies

$$\begin{aligned}\mathrm{E}\Big\{u(n-i)[\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)]^3\Big\} &\simeq 3\,\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathbf{r}_i^{\top}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\\ &= 3\,\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i. \end{aligned} \tag{89}$$

Substituting (89) into (87) yields

$$\mathrm{E}\{p_{i,a}(n)\} \simeq 3\,\mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i w_i^* + \sigma_z^2\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i w_i^*. \tag{90}$$

In order to simplify the model and avoid higher-order statistics, the approximation $\mathrm{E}\{\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^{\top}(n)\} \simeq \mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\{\tilde{\mathbf{w}}^{\top}(n)\}$ was used in the mean weight behavior analysis of the NNLMS, for which the detailed explanation was given in [19]. Using this approximation in (90), we obtain (31). Note that the recursive model derived for $\mathbf{K}(n)$ in Section 5 can also be employed to predict the mean weight behavior of the NNLMF algorithm. Nevertheless, a sufficiently accurate mean weight behavior model can be obtained by using this first-order approximation.

Appendix B. Detailed calculation of (32)

Using the approximation $\tilde{\mathbf{w}}_{\mathrm{E},i}^{\top}(n)\mathbf{u}(n) \simeq \tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)$ shown in Appendix A, $\mathrm{E}\{p_{i,b}(n)\}$ can be approximated as

$$\mathrm{E}\{p_{i,b}(n)\} \simeq \mathrm{E}\Big\{u(n-i)[\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)]^3\tilde{w}_i(n)\Big\} + \sigma_z^2\mathrm{E}\Big\{u(n-i)\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)\tilde{w}_i(n)\Big\}. \tag{91}$$

The term $\tilde{w}_i(n)$ can be considered weakly correlated with $u(n-i)[\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)]^3$ according to Assumptions A2, A4, and A5. Therefore, the first term on the right-hand side of (91) can be approximated as

$$\begin{aligned}\mathrm{E}\Big\{u(n-i)[\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)]^3\tilde{w}_i(n)\Big\} &\simeq \mathrm{E}\Big\{u(n-i)[\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)]^3\Big\}\mathrm{E}\{\tilde{w}_i(n)\} \simeq 3\\ &\quad \mathrm{Tr}\{\mathbf{R}\mathbf{K}(n)\}\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i\mathrm{E}\{\tilde{w}_i(n)\} \simeq 3\\ &\quad \mathrm{Tr}\Big\{\mathbf{R}\mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\Big\}\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i\mathrm{E}\{\tilde{w}_i(n)\}\\ &\quad . \end{aligned} \tag{92}$$

The approximation $\mathrm{E}\Big\{\tilde{\mathbf{w}}(n)\tilde{\mathbf{w}}^{\top}(n)\Big\} \simeq \mathrm{E}\{\tilde{\mathbf{w}}(n)\}\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}$ used in Appendix A implies that $\mathrm{E}\{\tilde{w}_i(n)\tilde{w}_j(n)\} \simeq \mathrm{E}\{\tilde{w}_i(n)\}\mathrm{E}\{\tilde{w}_j(n)\}, \forall\, i, j$. Thus, with Assumptions A2 and A4, the second term on the right-hand side of (91) can be approximated as

$$\sigma_z^2\mathrm{E}\Big\{u(n-i)\tilde{\mathbf{w}}^{\top}(n)\mathbf{u}(n)\tilde{w}_i(n)\Big\} \simeq \sigma_z^2\mathrm{E}\Big\{\tilde{\mathbf{w}}^{\top}(n)\Big\}\mathbf{r}_i\mathrm{E}\{\tilde{w}_i(n)\}. \tag{93}$$

Finally, using (92) and (93) in (91), one obtains (32).

References

[1] S. Haykin, Adaptive Filter Theory, 4th edition, Prentice-Hall, Upper Saddle River, NJ, 2002.
[2] A.H. Sayed, Adaptive Filters, Wiley, Hoboken, NJ, 2008.
[3] J. Ni, X. Chen, J. Yang, Two variants of the sign subband adaptive filter with improved convergence rate, Signal Process. 96 (2014) 325–331.
[4] Z. Zheng, H. Zhao, Affine projection M-estimate subband adaptive filters for robust adaptive filtering in impulsive noise, Signal Process. 120 (2016) 64–70.
[5] V. Patel, N.V. George, Compensating acoustic feedback in feed-forward active noise control systems using spline adaptive filters, Signal Process. 120 (2016) 448–455.
[6] L. Xiao, M. Wu, J. Yang, A new efficient filtered-x affine projection sign algorithm for active control of impulsive noise, Signal Process. 120 (2016) 456–461.
[7] J. Chen, C. Richard, A.H. Sayed, Multitask diffusion adaptation over networks, IEEE Trans. Signal Process. 62 (16) (2014) 4129–4144.
[8] J. Chen, C. Richard, A.H. Sayed, Diffusion LMS over multitask networks, IEEE Trans. Signal Process. 63 (11) (2015) 2733–2748.
[9] J. Chen, C. Richard, P. Honeine, H. Lantéri, C. Theys, System identification under non-negativity constraints, in: Proceedings of the European Signal Processing Conference (EUSIPCO), Aalborg, Denmark, 2010, pp. 1728–1732.
[10] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1995.
[11] R. Bro, S.D. Jong, A fast non-negativity-constrained least squares algorithm, J.

Chemom. 11 (5) (1997) 393–401.
[12] M.H.V. Benthem, M.R. Keenan, Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems, J. Chemom. 18 (10) (2004) 441–450.
[13] F. Benvenuto, R. Zanella, L. Zanni, M. Bertero, Nonnegative least-squares image deblurring: improved gradient projection approaches, Inverse Probl. 26 (2) (2010) 25004–25021.
[14] R. Badeau, N. Bertin, E. Vincent, Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization, IEEE Trans. Neural Netw. 21 (12) (2010) 1869–1881.
[15] K. Huang, N.S.A. Swami, Non-negative matrix factorization revisited: uniqueness and algorithm for symmetric decomposition, IEEE Trans. Signal Process. 62 (1) (2014) 211–224.
[16] G. Zhou, A. Cichocki, Q. Zhao, S. Xie, Nonnegative matrix and tensor factorizations: an algorithmic perspective, IEEE Signal Process. Mag. 31 (3) (2014) 54–65.
[17] Q. Sun, J. Lu, Y. Wu, H. Qiao, X. Huang, F. Hu, Non-informative hierarchical Bayesian inference for non-negative matrix factorization, Signal Process. 108 (2015) 309–321.
[18] S. Mirzaei, H.V. Hamme, Y. Norouzi, Blind audio source counting and separation of anechoic mixtures using the multichannel complex NMF framework, Signal Process. 115 (2015) 27–37.
[19] J. Chen, C. Richard, J.C.M. Bermudez, P. Honeine, Nonnegative least-mean-square algorithm, IEEE Trans. Signal Process. 59 (11) (2011) 5225–5235.
[20] J. Chen, J.C.M. Bermudez, C. Richard, P. Honeine, Steady-state performance of non-negative least-mean-square algorithm and its variants, IEEE Signal Process. Lett. 21 (8) (2014) 928–932.
[21] J. Chen, C. Richard, J.C.M. Bermudez, P. Honeine, Variants of non-negative least-mean-square algorithm and convergence analysis, IEEE Trans. Signal Process. 62 (15) (2014) 3990–4005.
[22] J. Chen, C. Richard, J.C.M. Bermudez, Reweighted Nonnegative Least-Mean-Square Algorithm and its Convergence Analysis, ⟨http://www.cedric-richard.fr/Articles/chen2015reweighted.pdf⟩.
[23] S.C. Pei, C.C. Tseng, Least mean p-power error criterion for adaptive FIR filter, IEEE J. Sel. Areas Commun. 12 (9) (1994) 1540–1547.
[24] O. Tanrikulu, A.G. Constantinides, Least-mean kurtosis: a novel higher order statistics based adaptive filtering algorithm, Electron. Lett. 30 (13) (1994) 189–190.
[25] J. Chambers, O. Tanrikulu, A. Constantinides, Least mean mixed norm adaptive filtering, Electron. Lett. 30 (19) (1994) 1574–1575.
[26] Y. Zhou, S.C. Chan, T.S. Ng, Least mean M-estimate algorithms for robust adaptive filtering in impulse noise, IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process. 47 (12) (2000) 1564–1568.
[27] E. Eweda, N.J. Bershad, J.C.M. Bermudez, Stochastic analysis of the least mean fourth algorithm for non-stationary white Gaussian inputs, Signal Image Video Process. 18 (1) (2014) 133–142.
[28] E. Walach, B. Widrow, The least mean fourth (LMF) adaptive algorithm and its family, IEEE Trans. Inf. Theory 30 (2) (1984) 275–283.
[29] P.I. Hubscher, J.C.M. Bermudez, An improved statistical analysis of the least mean fourth (LMF) adaptive algorithm, IEEE Trans. Signal Process. 51 (3) (2003) 664–671.
[30] V.H. Nascimento, J.C.M. Bermudez, Probability of divergence for the least-mean fourth algorithm, IEEE Trans. Signal Process. 54 (4) (2006) 1376–1385.
[31] P.I. Hubscher, J.C.M. Bermudez, V.H. Nascimento, A mean-square stability analysis of the least mean fourth adaptive algorithm, IEEE Trans. Signal Process. 55 (8) (2007) 4018–4028.
[32] A. Zerguine, M.K. Chan, T.Y. Al-Naffouri, M. Moinuddin, C.F.N. Cowan, Convergence and tracking analysis of a variable normalised LMF (XE-NLMF) algorithm, Signal Process. 89 (5) (2009) 778–790.
[33] N.J. Bershad, J.C.M. Bermudez, Mean-square stability of the normalized least-mean fourth algorithm for white Gaussian inputs, Digit. Signal Process. 21 (6) (2011) 694–700.
[34] E. Eweda, N.J. Bershad, Stochastic analysis of a stable normalized least mean fourth algorithm for adaptive noise canceling with a white Gaussian reference, IEEE Trans. Signal Process. 60 (12) (2012) 6235–6244.
[35] E. Eweda, A. Zerguine, New insights into the normalization of the least mean fourth algorithm, Signal Image Video Process. 7 (2) (2013) 255–262.
[36] O.R. Khattak, A. Zerguine, Leaky least mean fourth adaptive algorithm, IET Signal Process. 7 (2) (2013) 134–145.
[37] E. Eweda, Dependence of the stability of the least mean fourth algorithm on target weights non-stationarity, IEEE Trans. Signal Process. 62 (7) (2014) 1634–1643.
[38] E. Eweda, Mean-square stability analysis of a normalized least mean fourth algorithm for a Markov plant, IEEE Trans. Signal Process. 62 (24) (2014) 6545–6553.
[39] M.O. Sayina, Y. Yilmazb, A. Demirc, S.S. Kozat, The Krylov-proportionate normalized least mean fourth approach: formulation and performance analysis, Signal Process. 109 (2015) 1–13.
[40] J. Ni, J. Yang, Variable step-size diffusion least mean fourth algorithm for distributed estimation, Signal Process. 112 (2016) 93–97.
[41] J. Yang, J. Ni, Nonnegative least-mean-fourth algorithm, in: Proceedings of the International Conference on Wireless, Mobile and Multimedia Networks, 2013, pp. 185–188.