Signal Processing 91 (2011) 2091–2100

Contents lists available at ScienceDirect

Signal Processing

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

Stability analysis of adaptive filters with regression vector nonlinearities

Leonardo Rey Vega [a,*,1], Hernan Rey [a,1], Jacob Benesty [b]

[a] Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina
[b] INRS-EMT, Université du Québec, Montréal, Québec, Canada H5A 1K6

ARTICLE INFO

Article history:
Received 6 May 2010
Received in revised form 22 February 2011
Accepted 24 March 2011
Available online 30 March 2011

Keywords:
Adaptive filters
Mean stability
Mean-square stability
Step-size parameter

ABSTRACT

We present a unified framework to analyze the mean and mean-square stability of a large class of adaptive filters. We do this without obtaining a full transient model, allowing us to acquire sufficient conditions on the stability without assuming a given statistical distribution for the input regressors. We also apply the proposed framework to some popular adaptive filtering schemes, showing that in some cases the sufficient conditions derived are very tight and even necessary too.

1. Introduction

Adaptive filters are recursive estimators. This recursive nature raises the question of stability. As they are stochastic systems, several criteria can be used. Mean and mean-square stability are preferred, because with some appropriate assumptions they are basically easier to study.

A lot of effort has been put into analyzing the convergence properties of adaptive filters (see [1,2] and references therein). Besides the question of stability, it is also important to study the steady-state and transient behavior. However, the stability issue is the most critical, because it determines when an adaptive filter can be implemented and be useful for the application of interest.

The stability of several algorithms has been treated in the literature. However, in general, the stability analysis has been done for particular algorithms, and not from a general point of view or in a unified way. In [3,4] the question of stability (and also transient and steady-state behavior) is addressed for a large family of adaptive filters. Nevertheless, in general and for non-Gaussian input signals, the study of stability proposed in those works results in the analysis of an M^2-dimensional state-space equation, where M is the length of the adaptive filter. For adaptive filter lengths on the order of hundreds of taps, as several applications require [5], the numerical solution of the stability conditions derived in those works could be precluded.

In this paper, we will concentrate on the stability issue. In order to accomplish this we will not try to develop an exact transient model for the mean and mean-square error vector. Instead, we will look for necessary and sufficient conditions on quantities relevant to the mean and mean-square behavior. After that, we will try to constrain those quantities in order to guarantee stability of the adaptive filter. As this will be done in full generality, the results obtained can be applied with minor changes to a large family of adaptive filters and without assumptions on the statistical distribution of the input regressors. We will illustrate this through the application of the obtained results to several well-known adaptive filters.

* Corresponding author.
E-mail addresses: lrey@fi.uba.ar (L. Rey Vega), hrey@fi.uba.ar (H. Rey), benesty@emt.inrs.ca (J. Benesty).

[1] Work partially supported by CONICET.

0165-1684/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2011.03.018

First, we present certain definitions and notation used throughout the paper. For simplicity, we assume that all the signals involved are real; the more general complex case is a straightforward extension of the real case. Let $\mathbf{w}=(w_1,w_2,\ldots,w_M)^T$ be the unknown $M\times 1$ system. The $M\times 1$ input vector at time i, $\mathbf{x}_i$, is combined by the system giving an output $y_i=\mathbf{w}^T\mathbf{x}_i$. This output is observed, but is usually corrupted by noise, v_i, which will be considered additive. Thus, each input $\mathbf{x}_i$ gives an output $d_i=\mathbf{w}^T\mathbf{x}_i+v_i$. We want to find $\hat{\mathbf{w}}_i$ to estimate $\mathbf{w}$. This adaptive filter receives the same input, leading to an output error $e_i=d_i-\hat{\mathbf{w}}_i^T\mathbf{x}_i$. We also define the misalignment vector $\tilde{\mathbf{w}}_i=\mathbf{w}-\hat{\mathbf{w}}_i$. We denote the identity matrix and the zero matrix of appropriate dimensions by $\mathbf{I}$ and $\mathbf{0}$, respectively, while $\mathrm{tr}(\cdot)$, $E[\cdot]$, and $\mathrm{diag}(\cdot)$ denote the trace, expectation, and diagonal operators, respectively. For a matrix $\mathbf{A}$, $\lambda_m(\mathbf{A}),m=1,2,\ldots,M$, denote its eigenvalues, with $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ being the largest and smallest, respectively.

2. Problem formulation

In order to obtain full generality in our analysis, we will assume the following form for the adaptive filter:

$$\hat{\mathbf{w}}_{i+1}=\alpha\hat{\mathbf{w}}_i+\boldsymbol{\mu}\mathbf{f}(\mathbf{x}_i)e_i, \tag{1}$$

where α is a positive number typically in (0,1], $\boldsymbol{\mu}=\mathrm{diag}(\mu_1,\mu_2,\ldots,\mu_M)$, $\mu_m>0$, and $\mathbf{f}:\mathbb{R}^M\to\mathbb{R}^M$ is in principle an arbitrary multidimensional function. The algorithm is initialized with an arbitrary vector $\hat{\mathbf{w}}_0$, which is typically equal to the null vector. The mathematical form assumed in (1) is sufficiently general to encompass several popular gradient-based adaptive filters. Important cases in this family are the least-mean-square (LMS), the normalized LMS (NLMS), the signed regressor (SR), the leaky LMS (LLMS), the multiple step-size LMS (MLMS), etc. [1,2]. Using the definition of the misalignment vector and the model assumed for the signal d_i, we can write

$$\tilde{\mathbf{w}}_{i+1}=[\alpha\mathbf{I}-\boldsymbol{\mu}\mathbf{f}(\mathbf{x}_i)\mathbf{x}_i^T]\tilde{\mathbf{w}}_i-\boldsymbol{\mu}\mathbf{f}(\mathbf{x}_i)v_i+(1-\alpha)\mathbf{w}. \tag{2}$$

Our interest is to analyze the mean and mean-square stability of $\tilde{\mathbf{w}}_i$. For this purpose and based on some statistical properties of the input regressor, we will look for conditions on $\boldsymbol{\mu}$ that guarantee stable behavior of the adaptive filter. Given the fact that we restrict ourselves to the case $\boldsymbol{\mu}=\mathrm{diag}(\mu_1,\mu_2,\ldots,\mu_M)$ with $\mu_m>0$, we can work with $\tilde{\mathbf{c}}_i=\boldsymbol{\mu}^{-1/2}\tilde{\mathbf{w}}_i$. Defining $\tilde{\mathbf{f}}(\mathbf{x}_i)=\boldsymbol{\mu}^{1/2}\mathbf{f}(\mathbf{x}_i)$, $\tilde{\mathbf{x}}_i=\boldsymbol{\mu}^{1/2}\mathbf{x}_i$, and $\mathbf{c}=\boldsymbol{\mu}^{-1/2}\mathbf{w}$, (2) can be rewritten as

$$\tilde{\mathbf{c}}_{i+1}=[\alpha\mathbf{I}-\tilde{\mathbf{f}}(\mathbf{x}_i)\tilde{\mathbf{x}}_i^T]\tilde{\mathbf{c}}_i-\tilde{\mathbf{f}}(\mathbf{x}_i)v_i+(1-\alpha)\mathbf{c}. \tag{3}$$

From (3) we can obtain

$$\tilde{\mathbf{c}}_{i+1}=\mathbf{A}(i,0)\tilde{\mathbf{c}}_0-\sum_{j=0}^{i}\mathbf{A}(i,j+1)[\tilde{\mathbf{f}}(\mathbf{x}_j)v_j-(1-\alpha)\mathbf{c}], \tag{4}$$

where the matrix $\mathbf{A}(i,j)$ is defined by

$$\mathbf{A}(i,j)=\begin{cases}[\alpha\mathbf{I}-\tilde{\mathbf{f}}(\mathbf{x}_i)\tilde{\mathbf{x}}_i^T] & \\ \quad\times[\alpha\mathbf{I}-\tilde{\mathbf{f}}(\mathbf{x}_{i-1})\tilde{\mathbf{x}}_{i-1}^T]\cdots[\alpha\mathbf{I}-\tilde{\mathbf{f}}(\mathbf{x}_j)\tilde{\mathbf{x}}_j^T], & j\le i,\\ \mathbf{I}, & j=i+1,\\ \mathbf{0}, & j>i+1,\end{cases} \tag{5}$$

In the following, we will assume that the noise signal v_i is a zero-mean i.i.d. sequence with variance σ_v^2 and independent from the input $\mathbf{x}_i$, which is also a zero-mean stationary signal with correlation matrix $\mathbf{R}_{xx}$. We will also make the standard independence assumption on the input regressors [1,3].

3. Mean stability

Taking expectation on both sides of (4) and using the assumptions on the noise and the input signals, we obtain

$$\begin{aligned}E[\tilde{\mathbf{c}}_{i+1}]&=E[\mathbf{A}(i,0)]\tilde{\mathbf{c}}_0+(1-\alpha)\sum_{j=0}^{i}E[\mathbf{A}(i,j+1)]\mathbf{c}\\&=\mathbf{B}^{i+1}\tilde{\mathbf{c}}_0+(1-\alpha)\sum_{j=0}^{i}\mathbf{B}^{i-j}\mathbf{c},\end{aligned} \tag{6}$$

where

$$\mathbf{B}=\alpha\mathbf{I}-\boldsymbol{\mu}^{1/2}\mathbf{C}_{xx}\boldsymbol{\mu}^{1/2}, \tag{7}$$

and

$$\mathbf{C}_{xx}=E[\mathbf{f}(\mathbf{x}_i)\mathbf{x}_i^T]. \tag{8}$$

A necessary and sufficient condition for the stability of (6) is to choose $\boldsymbol{\mu}$ according to

$$|\alpha-\lambda_m(\boldsymbol{\mu}^{1/2}\mathbf{C}_{xx}\boldsymbol{\mu}^{1/2})|<1,\quad m=1,2,\ldots,M. \tag{9}$$

In several cases of interest, the matrix $\mathbf{C}_{xx}$ is positive definite, which implies that $\lambda_m(\boldsymbol{\mu}^{1/2}\mathbf{C}_{xx}\boldsymbol{\mu}^{1/2})$, $m=1,2,\ldots,M$ are positive, so a careful choice of $\boldsymbol{\mu}$ with $\mu_m>0$ can guarantee the mean stability. However, when $\mathbf{C}_{xx}$ is not positive definite (or some of its eigenvalues have negative real parts), it would be possible that no choice of $\boldsymbol{\mu}$ will guarantee stability of the algorithm. For example, with the SR algorithm, where $\mathbf{C}_{xx}=E[\mathrm{sign}(\mathbf{x}_i)\mathbf{x}_i^T]$, some class of input signals could lead to eigenvalues with negative real parts [6]. In the important case when $\mathbf{C}_{xx}$ is positive definite, we can write (9) as

$$0<\lambda_m(\boldsymbol{\mu}^{1/2}\mathbf{C}_{xx}\boldsymbol{\mu}^{1/2})<1+\alpha,\quad m=1,2,\ldots,M. \tag{10}$$

4. Mean-square stability

It is known that in order to obtain a better picture of the way in which an adaptive filter works, besides the mean behavior, we need to look into the mean-square performance and stability. In the following, we will derive a sufficient condition for the mean-square stability of (1). We will assume that $\mathbf{C}_{xx}$ is symmetric and positive definite. This will be true in most cases. The general case can also be analyzed, but the mathematics are more involved and will not be done here. This assumption implies that $\mathbf{B}$ defined in (7) is symmetric.

From (4) and using the assumptions on the noise and the input signals, we obtain

$$\begin{aligned}E[\|\tilde{\mathbf{c}}_{i+1}\|^2]&=\tilde{\mathbf{c}}_0^TE[\mathbf{A}^T(i,0)\mathbf{A}(i,0)]\tilde{\mathbf{c}}_0\\&\quad+2(1-\alpha)\sum_{j=0}^{i}\tilde{\mathbf{c}}_0^TE[\mathbf{A}^T(i,0)\mathbf{A}(i,j+1)]\mathbf{c}\end{aligned}$$

$$+\sigma_v^2\sum_{j=0}^{i}E[\tilde{\mathbf{f}}^T(\mathbf{x}_j)\mathbf{A}^T(i,j+1)\mathbf{A}(i,j+1)\tilde{\mathbf{f}}(\mathbf{x}_j)]$$
$$+(1-\alpha)^2\sum_{j=0}^{i}\sum_{k=0}^{i}\mathbf{c}^TE[\mathbf{A}^T(i,k+1)\mathbf{A}(i,j+1)]\mathbf{c}. \tag{11}$$

Eq. (11) is valid in the general case where the input regressors are not necessarily independent. In order to analyze the stability of (11) we will use the independence assumption. This is done only for maintaining simple mathematics. It should be emphasized that it would be possible to perform a stability analysis from (11) without the independence assumption. However, the mathematical difficulties would be greater, and sooner or later some proper mixing condition on the input would be needed. Mixing conditions [7] allow us to introduce dependence between successive input regressors. This dependency could be arbitrary, with the condition that it decreases sufficiently fast when the time lag between the two input regressors under consideration is large.

Now, we define

$$\mathbf{D}(i,j)=E[\mathbf{A}^T(i,j)\mathbf{A}(i,j)]. \tag{12}$$

Using the definition of $\mathbf{A}(i,j)$ and the independence assumption, we have

$$E[\mathbf{A}^T(i,k)\mathbf{A}(i,j)]=\begin{cases}\mathbf{D}(i,k)\mathbf{B}^{k-j}, & j\le k,\\ \mathbf{B}^{j-k}\mathbf{D}(i,j), & j>k.\end{cases} \tag{13}$$

With these definitions, we can write the four terms on the RHS of (11) as

$$T_1=\tilde{\mathbf{c}}_0^T\mathbf{D}(i,0)\tilde{\mathbf{c}}_0, \tag{14}$$

$$T_2=2(1-\alpha)\sum_{j=0}^{i}\tilde{\mathbf{c}}_{-1}^T\mathbf{B}^{j+1}\mathbf{D}(i,j+1)\mathbf{c}, \tag{15}$$

$$T_3=\sigma_v^2\sum_{j=0}^{i}\mathrm{tr}[\tilde{\mathbf{F}}_{xx}\mathbf{D}(i,j+1)], \tag{16}$$

$$T_4=(1-\alpha)^2\sum_{j=0}^{i}\left[\sum_{k=j}^{i}\mathbf{c}^T\mathbf{D}(i,k+1)\mathbf{B}^{k-j}\mathbf{c}+\sum_{k=0}^{j-1}\mathbf{c}^T\mathbf{B}^{j-k}\mathbf{D}(i,j+1)\mathbf{c}\right], \tag{17}$$

respectively, and also define

$$\tilde{\mathbf{F}}_{xx}=E[\tilde{\mathbf{f}}(\mathbf{x}_j)\tilde{\mathbf{f}}^T(\mathbf{x}_j)], \tag{18}$$

which will be assumed to be positive definite. The mean-square stability of $\tilde{\mathbf{w}}_i$ is equivalent to $\lim_{i\to\infty}E[\|\tilde{\mathbf{w}}_i\|^2]<\infty$, which is also equivalent (given that $\mu_m>0$, $m=1,2,\ldots,M$) to $\lim_{i\to\infty}E[\|\tilde{\mathbf{c}}_i\|^2]<\infty$. In this way, restricting us to the study of $E[\|\tilde{\mathbf{c}}_i\|^2]$, we have the following theorem:

Theorem 1. *A necessary and sufficient condition for* $\lim_{i\to\infty}E[\|\tilde{\mathbf{c}}_{i+1}\|^2]<\infty$, *for every* $\tilde{\mathbf{c}}_0$ *and* $\mathbf{c}$ *is given by the satisfaction of* (10) *and, in addition,*

$$\exists N(\boldsymbol{\mu}),\gamma(\boldsymbol{\mu})\ \text{with}\ N(\boldsymbol{\mu})>0\ \text{and}\ 0<\gamma(\boldsymbol{\mu})<1\ \text{such that}$$
$$\mathrm{tr}[\mathbf{D}(i,j+1)]\le N(\boldsymbol{\mu})\gamma^{i-j}(\boldsymbol{\mu}),\quad \forall i\ge j. \tag{19}$$

The proof of this theorem can be found in the appendix. The fact that the condition given in (19) is necessary and sufficient allows us, without loss of generality, to restrict $\mathrm{tr}[\mathbf{D}(i,k+1)]$ to exponential behavior. Although there are multiple ways to accomplish this, we will focus on a particular one, for which we will obtain only sufficient conditions for the mean-square stability, which from a practical point of view are more useful than necessary conditions.

From (A.24) and the independence assumption, we can write

$$\mathrm{tr}[\mathbf{D}(i,j)]=\mathrm{tr}[\mathbf{A}^T(i-1,j)\mathbf{G}_{xx}\mathbf{A}(i-1,j)]$$
$$=\mathrm{tr}[\mathbf{G}_{xx}\mathbf{A}(i-1,j)\mathbf{A}^T(i-1,j)]\le\lambda_{\max}(\mathbf{G}_{xx})\mathrm{tr}[\mathbf{D}(i-1,j)], \tag{20}$$

where $\mathbf{G}_{xx}$ is defined as

$$\mathbf{G}_{xx}=E[\mathbf{E}^T(i)\mathbf{E}(i)]. \tag{21}$$

This procedure can be repeated several times to obtain

$$\mathrm{tr}[\mathbf{D}(i,j)]\le M[\lambda_{\max}(\mathbf{G}_{xx})]^{i-j+1}. \tag{22}$$

From this and in view of the result of Theorem 1, we can guarantee the mean-square stability by choosing $\boldsymbol{\mu}$ in such a way that

$$\lambda_{\max}(\mathbf{G}_{xx})<1. \tag{23}$$

The matrix $\mathbf{G}_{xx}$ can be expressed as

$$\mathbf{G}_{xx}=\alpha^2\mathbf{I}-\alpha\boldsymbol{\mu}^{1/2}\mathbf{C}_{xx}^T\boldsymbol{\mu}^{1/2}-\alpha\boldsymbol{\mu}^{1/2}\mathbf{C}_{xx}\boldsymbol{\mu}^{1/2}+\boldsymbol{\mu}^{1/2}\mathbf{H}_{xx}\boldsymbol{\mu}^{1/2}, \tag{24}$$

where

$$\mathbf{H}_{xx}=E[\mathbf{f}^T(\mathbf{x}_i)\boldsymbol{\mu}\mathbf{f}(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T]. \tag{25}$$

As (23) is equivalent to asking for

$$\mathbf{G}_{xx}<\mathbf{I}, \tag{26}$$

where the inequality sign refers to the usual partial ordering defined for positive definite matrices [8], we will need to search for $\boldsymbol{\mu}$ in order to guarantee this.

We will see that, although the exponential bound (22) might not be the tightest one, this approach can lead to some well-known results on the stability of classical adaptive filters that are actually very tight. At this point, we only keep the sufficiency and could lose the necessity.

5. Particular cases

In this section, we will use the results of the previous sections to analyze the stability of several well-known adaptive filters. For some algorithms, the results depend not only on the correlation matrix $\mathbf{R}_{xx}$ but also on some other input moments. If these moments have no well-known closed-form expression for a particular input distribution, they can be estimated by simulation. The computational load of such a task is negligible in comparison with that required by other stability analyses proposed in the literature [3,4].

5.1. Least-mean-square (LMS)

In this case $\mathbf{f}(\mathbf{x}_i)=\mathbf{x}_i$, $\alpha=1$, and $\boldsymbol{\mu}=\mu\mathbf{I}$. For mean stability, (10) reduces to the well-known relation

$$0<\mu<\frac{2}{\lambda_{\max}(\mathbf{R}_{xx})}. \quad (27)$$

The mean-square stability condition (26) reduces to

$$2\mathbf{R}_{xx}>\mu E[\|\mathbf{x}_i\|^2\mathbf{x}_i\mathbf{x}_i^T], \quad (28)$$

which is equivalent to

$$0<\mu<2\lambda_{\min}(\mathbf{R}_{xx}\mathbf{J}_{xx}^{-1}), \quad (29)$$

where

$$\mathbf{J}_{xx}=E[\|\mathbf{x}_i\|^2\mathbf{x}_i\mathbf{x}_i^T]. \quad (30)$$

Eq. (29) can also be rewritten as

$$0<\mu<\frac{2}{\lambda_{\max}(\mathbf{R}_{xx}^{-1}\mathbf{J}_{xx})}. \quad (31)$$

This expression, in combination with (27), guarantees the mean-square stability of the LMS algorithm. We note that this condition is valid, in principle, for all input statistical distributions. However, in order to obtain a more explicit characterization, we need an expression for the fourth-order term $\mathbf{J}_{xx}$. If we add the assumption that the input regressor is Gaussian, we can obtain an explicit expression for that term [2]:

$$\mathbf{J}_{xx}=2\mathbf{R}_{xx}^2+\mathbf{R}_{xx}\mathrm{tr}(\mathbf{R}_{xx}), \quad (32)$$

so condition (31) can be written as

$$0<\mu<\frac{2}{2\lambda_{\max}(\mathbf{R}_{xx})+\mathrm{tr}(\mathbf{R}_{xx})}. \quad (33)$$

This is only a sufficient condition for the stability of the LMS with Gaussian signals. However, it is close enough to the necessary and sufficient condition that can be obtained from an elaborated model of the transient behavior of the LMS algorithm for Gaussian signals [1,3]. In fact, for white Gaussian signals, condition (33) is also necessary. A simplified (and more restrictive) bound for μ can be obtained using the fact that $\lambda_{\max}(\mathbf{R}_{xx})\le\mathrm{tr}(\mathbf{R}_{xx})$, so (33) becomes

$$0<\mu<\frac{2}{3\mathrm{tr}(\mathbf{R}_{xx})}, \quad (34)$$

which is the same result obtained in [9].

5.2. Leaky LMS (LLMS)

In this case $\mathbf{f}(\mathbf{x}_i)=\mathbf{x}_i$, $\boldsymbol{\mu}=\mu\mathbf{I}$, and $\alpha=1-\beta\mu$ [1], where $\beta>0$ and $\mu<1/\beta$. Eq. (10) simplifies to

$$0<\mu<\frac{2}{\beta+\lambda_{\max}(\mathbf{R}_{xx})}. \quad (35)$$

From (26), we can write

$$2(\beta\mathbf{I}+\mathbf{R}_{xx})>\mu(\beta^2\mathbf{I}+2\beta\mathbf{R}_{xx}+\mathbf{J}_{xx}), \quad (36)$$

from which we can obtain

$$0<\mu<\frac{2}{\lambda_{\max}[(\beta\mathbf{I}+\mathbf{R}_{xx})^{-1}(\beta^2\mathbf{I}+2\beta\mathbf{R}_{xx}+\mathbf{J}_{xx})]}. \quad (37)$$

This is a sufficient stability bound for the LLMS algorithm under a general input distribution, and is consistent with (31) for $\beta=0$. If the input is Gaussian, we can use (32) to obtain

$$0<\mu<\max_{m=1,2,\ldots,M}\left[\frac{2(\beta+\lambda_m)}{(\beta+\lambda_m)^2+\lambda_m^2+\lambda_m\mathrm{tr}(\mathbf{R}_{xx})}\right], \quad (38)$$

where λ_m, $m=1,2,\ldots,M$ denote the eigenvalues of $\mathbf{R}_{xx}$, being consistent again with (33) for $\beta=0$. This stability bound is sufficiently tight with respect to the sufficient and necessary condition derived in [10] for Gaussian signals (again, for white Gaussian signals condition (38) is also necessary), and is a better sufficient condition than the approximation given in [10, Eq. (43)]. Useful analyses for the Leaky LMS algorithm are also performed in [11,12].

5.3. Normalized LMS (NLMS)

In this case, we have $\mathbf{f}(\mathbf{x}_i)=\mathbf{x}_i/\|\mathbf{x}_i\|^2$, $\boldsymbol{\mu}=\mu\mathbf{I}$, and $\alpha=1$. Defining

$$\mathbf{K}_{xx}=E\left[\frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{x}_i\|^2}\right], \quad (39)$$

we can show that (10) can be put as

$$0<\mu<\frac{2}{\lambda_{\max}(\mathbf{K}_{xx})}. \quad (40)$$

In the same manner, (26) reduces to

$$2\mathbf{K}_{xx}>\mu\mathbf{K}_{xx}, \quad (41)$$

from which we have

$$0<\mu<\frac{2}{\lambda_{\max}(\mathbf{K}_{xx}^{-1}\mathbf{K}_{xx})}=2. \quad (42)$$

In this way, we obtain the well-known stability bound $0<\mu<2$ for the NLMS algorithm.

We emphasize the fact that the above bound does not depend on the statistical distribution of the input signal and is not restricted to the Gaussian case. We note that when the NLMS has regularization [1], i.e., $\mathbf{f}(\mathbf{x}_i)=\mathbf{x}_i/(\|\mathbf{x}_i\|^2+\delta)$ with $\delta>0$, we see that the upper limit of the stability region is greater than 2, but the exact value can be very difficult to compute even in the Gaussian case, because we need to compute $\mathbf{K}_{xx}$ and

$$E\left[\frac{\|\mathbf{x}_i\|^2}{(\|\mathbf{x}_i\|^2+\delta)^2}\mathbf{x}_i\mathbf{x}_i^T\right]. \quad (43)$$

Recently, in [13], it was shown how to calculate these moments for the general circular complex and correlated Gaussian case.

5.4. Signed regressor (SR)

For the SR algorithm, we have $\mathbf{f}(\mathbf{x}_i)=\mathrm{sign}(\mathbf{x}_i)$, $\boldsymbol{\mu}=\mu\mathbf{I}$, and $\alpha=1$. If we define

$$\mathbf{L}_{xx}=E[\mathrm{sign}(\mathbf{x}_i)\mathbf{x}_i^T] \quad (44)$$

and assume that this matrix is positive definite, (10) can be put as

$$0<\mu<\frac{2}{\lambda_{\max}(\mathbf{L}_{xx})}. \tag{45}$$

Also, (26) reduces to

$$2\mathbf{L}_{xx}>\mu M\mathbf{R}_{xx}, \tag{46}$$

which gives us

$$0<\mu<\frac{2}{M\lambda_{\max}(\mathbf{L}_{xx}^{-1}\mathbf{R}_{xx})}, \tag{47}$$

where we used the fact that $E[\|\text{sign}(\mathbf{x}_i)\|^2\mathbf{x}_i\mathbf{x}_i^T]=M\mathbf{R}_{xx}$. If the input is Gaussian, it is easy to prove (using Price's theorem [14]) that $\mathbf{L}_{xx}=\sqrt{(2/\pi\sigma_x^2)}\mathbf{R}_{xx}$, so (45) can be rewritten as

$$0<\mu<\sqrt{\frac{\pi\sigma_x^2}{2}}\cdot\frac{1}{\lambda_{\max}(\mathbf{R}_{xx})} \tag{48}$$

and (47) can be put as

$$0<\mu<\frac{1}{M}\sqrt{\frac{8}{\pi\sigma_x^2}}, \tag{49}$$

which is the same result obtained in [3,15]. From the results in [15], where this stability bound is calculated through a full transient analysis, we know that this bound is also a necessary condition for convergence.

5.5. Multiple step-size LMS (MLMS)

In this case, we have $\mathbf{f}(\mathbf{x}_i)=\mathbf{x}_i$, $\boldsymbol{\mu}=\text{diag}(\mu_m)$, $\mu_m>0$, $m=1,2,\ldots,M$, and $\alpha=1$. Eq. (10) gives

$$|1-\lambda_m(\boldsymbol{\mu}\mathbf{R}_{xx})|<1, \quad m=1,2,\ldots,M. \tag{50}$$

On the other hand, from (24) and (26) we obtain

$$2\mathbf{R}_{xx}>\mathbf{M}_{xx}(\boldsymbol{\mu}), \tag{51}$$

where

$$\mathbf{M}_{xx}(\boldsymbol{\mu})=E[\mathbf{x}_i^T\boldsymbol{\mu}\mathbf{x}_i\mathbf{x}_i\mathbf{x}_i^T]. \tag{52}$$

It is clear that

$$\mathbf{M}_{xx}(\boldsymbol{\mu})=\sum_{k=1}^{M}\mu_k\mathbf{N}_{xx}^{(k)}, \tag{53}$$

where $\mathbf{N}_{xx}^{(k)}$ are positive definite matrices given by

$$\mathbf{N}_{xx}^{(k)}=E[x_{i-1+k}^2\mathbf{x}_i\mathbf{x}_i^T], \quad \sum_{k=1}^{M}\mathbf{N}_{xx}^{(k)}=\mathbf{J}_{xx}. \tag{54}$$

The condition for the mean-square stability (51) can be written as

$$\sum_{k=1}^{M}\mu_k\mathbf{N}_{xx}^{(k)}<2\mathbf{R}_{xx}. \tag{55}$$

This means that the stability region for μ_m $m=1,2,\ldots,M$ is the intersection of the positive orthant with the feasibility region of the linear matrix inequality [16] given by (55). Without any assumption on the input distribution, it can be easily proved that the stability region is convex. However, not much more can be said without a statistical assumption on the input, which is necessary to determine the feasibility region of (55), and furthermore is, in general, impossible without a numerical procedure. In the Gaussian case, (55) can be written as

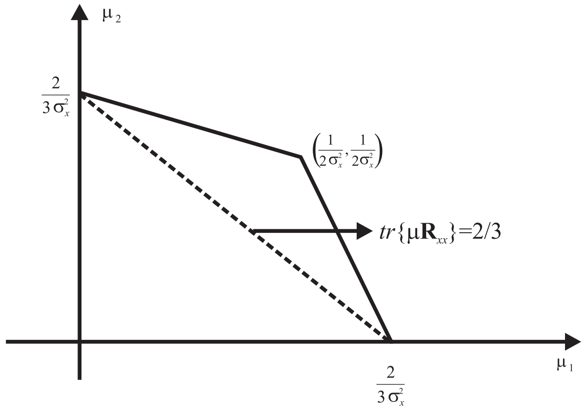

Fig. 1. Stability region of the MLMS algorithm for white Gaussian input ($M=2$).

$$2\mathbf{R}_{xx}>2\mathbf{R}_{xx}\boldsymbol{\mu}\mathbf{R}_{xx}+\mathbf{R}_{xx}\text{tr}(\boldsymbol{\mu}\mathbf{R}_{xx}). \tag{56}$$

This is equivalent to

$$2\mathbf{I}>2\mathbf{R}_{xx}^{1/2}\boldsymbol{\mu}\mathbf{R}_{xx}^{1/2}+\text{tr}(\boldsymbol{\mu}\mathbf{R}_{xx})\mathbf{I}, \tag{57}$$

which implies that

$$2>2\lambda_{\max}(\boldsymbol{\mu}\mathbf{R}_{xx})+\text{tr}(\boldsymbol{\mu}\mathbf{R}_{xx}). \tag{58}$$

Using the fact that $\lambda_{\max}(\boldsymbol{\mu}\mathbf{R}_{xx})\le\text{tr}(\boldsymbol{\mu}\mathbf{R}_{xx})$, we can obtain an explicit and easily computable region on $\boldsymbol{\mu}$ for the stability of the algorithm. More precisely,

$$\text{tr}(\boldsymbol{\mu}\mathbf{R}_{xx})<\tfrac{2}{3}, \quad \mu_m>0, \ m=1,2,\ldots,M, \tag{59}$$

which is equivalent to the condition obtained in [2].

If the input is Gaussian and $\mathbf{R}_{xx}=\sigma_x^2\mathbf{I}$, the stability region can be obtained exactly. In that case, it can be proved that (56) can be written as the intersection of the following half-spaces:

$$\frac{2}{\sigma_x^2}>3\mu_m+\sum_{m'\neq m}\mu_{m'}, \quad \mu_m>0, \ m=1,2,\ldots,M. \tag{60}$$

In Fig. 1 we see the stability region (solid line) for a white Gaussian input with $M=2$. We also see the approximate stability region (dashed line) given by (59), which is valid even if the Gaussian input is colored, and which is strictly smaller than the region given by (60).

We also mention that in [3], it is claimed that the approach presented there can be used to analyze the MLMS, although the particular analysis is not carried out.

6. Simulation results

In order to test the suitability of the stability bounds derived in the previous section, we performed some numerical simulations. We will test the results only for the LMS and SR algorithms. As we are interested in identifying when the value of μ leads to instability, we need to analyze

$$\lim_{i\to\infty} E\left[\frac{\|\tilde{\mathbf{w}}_i\|^2}{\|\mathbf{w}\|^2}\right], \tag{61}$$

which is the normalized version of the "final" mean-square behavior. From a practical point of view, it is impossible to compute this quantity. For this reason, and in order to analyze the steady-state mismatch (SSM), we propose to work with

$$\text{SSM} = \left\langle E\left[\frac{\|\tilde{\mathbf{w}}_i\|^2}{\|\mathbf{w}\|^2}\right]\right\rangle = E[\text{SSM}_{\text{TA}}], \tag{62}$$

where $\langle \cdot \rangle$ denotes time averaging from iteration 39 500 to 40 000, and

$$\text{SSM}_{\text{TA}} = \left\langle \frac{\|\tilde{\mathbf{w}}_i\|^2}{\|\mathbf{w}\|^2}\right\rangle. \tag{63}$$

It is assumed that during this time interval the adaptive algorithm, if stable, will be close to its steady-state and, if unstable, it will present a high value of $E[\|\tilde{\mathbf{w}}_i\|^2/\|\mathbf{w}\|^2]$. The time averaging provides robustness against statistical fluctuations.

Clearly, SSM_{TA} is a random variable, with its mean being equal to the SSM. In order to obtain an estimate of this mean, 100 independent realizations of the algorithm would lead to the sequence $\{\text{SSM}_{\text{TA}}^{(n)}\}$, $n = 1,\ldots,100$. Then, the estimate of the mean can be computed as

$$\overline{\text{SSM}}_{\text{TA}} = \frac{1}{100}\sum_{n=1}^{100} \text{SSM}_{\text{TA}}^{(n)}, \tag{64}$$

This estimator is a random variable itself, whose mean is equal to the mean of SSM_{TA}, i.e., the SSM. Based on 100 realizations of the algorithm we will compute a single realization of the estimator (64). If this estimate turns out to be "small", should we believe that the algorithm is stable for that choice of μ? Or is it actually unstable but the "small" estimate was a product of statistical fluctuations? In order to tackle this issue, we can estimate the population standard deviation of the estimator $\overline{\text{SSM}}_{\text{TA}}$ (also known as *standard error of the mean*) by computing

$$\hat{\sigma}_{\overline{\text{SSM}}_{\text{TA}}} = \frac{1}{\sqrt{100}}\sqrt{\frac{1}{100}\sum_{n=1}^{100}(\text{SSM}_{\text{TA}}^{(n)} - \overline{\text{SSM}}_{\text{TA}})^2} = \frac{\hat{\sigma}_{\text{SSM}_{\text{TA}}}}{\sqrt{100}}, \tag{65}$$

where $\hat{\sigma}_{\text{SSM}_{\text{TA}}}$ is an estimator of the population standard deviation of SSM_{TA}. If the estimate from (65) is "small", we can trust on the result given by (64) as a good estimate of the SSM, since the estimator $\overline{\text{SSM}}_{\text{TA}}$ will be subject to "small" fluctuations. This will be important to help us determine whether or not the algorithm is in an unstable situation, and hence, provide better insight to the tightness of the proposed stability bound. On the other hand, the required correlation matrices used to compute the proposed stability bounds in each scenario (e.g., matrices $\mathbf{J}_{xx}$ and $\mathbf{L}_{xx}$ for the LMS and SR algorithms, respectively) are estimated with a large ensemble of independent realizations of the input. This will let us cover several scenarios where no closed-form expressions exist for these matrices.

We will test the different algorithms with filter lengths $M=32$ and 128. In order to test different input statistics, we will generate the input signals according to distributions such as Gaussian and uniform. We will also test the algorithm using input regressors generated according to a spherically invariant random vector (SIRV) model [17], which is very important for speech [18] and radar applications [19]. In particular we will use a SIRV with characteristic pdf given by a chi-square random variable with 16 degrees of freedom.

We will also include the effect of input correlation. All three input distributions will be generated so that the sequence $\{\mathbf{x}_i\}$ will have an associated correlation matrix given by

$$\mathbf{R}_{xx} = \sigma_x^2 \begin{bmatrix} 1 & a & \cdots & a^{M-1} \\ a & 1 & \cdots & a^{M-2} \\ \vdots & \ddots & \ddots & \vdots \\ a^{M-1} & \cdots & a & 1 \end{bmatrix} \tag{66}$$

with $0 \le a < 1$. The value $a=0$ leads to uncorrelated data, whereas $a=0.9$ corresponds to highly correlated data. It is known that in the Gaussian and SIRV case, the desired correlation can be generated through an appropriate linear transformation [19]. In the uniform case, this can be accomplished using the Spearman correlation coefficient [20]. In every case, the power of the system output $y_i = \mathbf{w}^T\mathbf{x}_i$ and background noise v_i were set to $\sigma_y^2 = 1$ and $\sigma_v^2 = 0.001$, respectively, with the noise being zero-mean, white, and Gaussian.

In Fig. 2 we show the results for the LMS algorithm with $M=32$ and $a=0$ (uncorrelated data). The vertical dashed line represents the numerically computed bound μ_0 according to (31). The x axis is segmented in intervals of length $0.05\mu_0$. According to the overly conservative theoretical result (34), the stability bound under Gaussian regressors in this scenario would be 0.0208. As derived in Section 5.1, the less restrictive result (33) should be

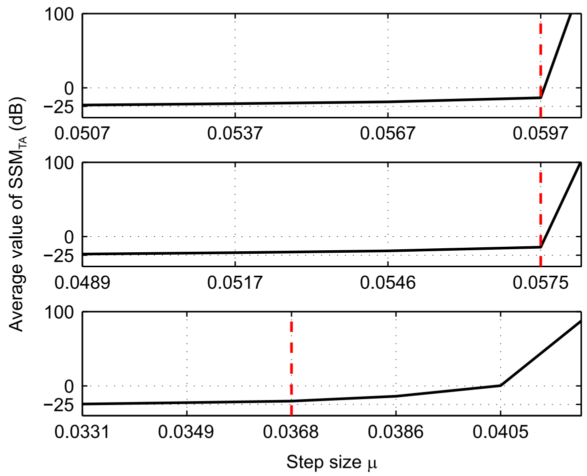

Fig. 2. $\overline{\text{SSM}}_{\text{TA}}$ (estimate of the steady-state mismatch) for the LMS algorithm. $M=32$, $a=0$ (uncorrelated data). Input distributions are uniform (top), Gaussian (middle), and SIRV (bottom). SSM_{TA} was averaged over 100 independent realizations. The vertical dashed line represents the numerically computed bound μ_0 according to (31). The x axis is segmented in intervals of length $0.05\mu_0$.

necessary and sufficient in this scenario, and its corresponding bound is 0.0588. It can be seen that this value is very close to the simulated one obtained by applying our model for the Gaussian input. Moreover, a similar value was obtained for the bound with uniform input. These bounds appear to be very tight (they are at most within 5% of the actual stability limit).

On the other hand, the LMS algorithm with μ close to 0.0588 and SIRV input data would be absolutely unstable. The more conservative result 0.0208 would actually lead to stable behavior for the three inputs, but would be overly conservative for the uniform and Gaussian data. In fact, the step-size is chosen in practice to be close to half the value resulting from (34). The consequences of being so conservative will be seen in slower convergence of the algorithm. Finally, it might seem that the bound obtained for the SIRV input is not as tight as the ones in the other cases. We will perform a more thorough analysis later to see whether the increments in SSM as μ is increased are due to the typical dynamics of the LMS algorithm or whether some of the realizations were indeed unstable (or at least exhibited some undesirable behavior).

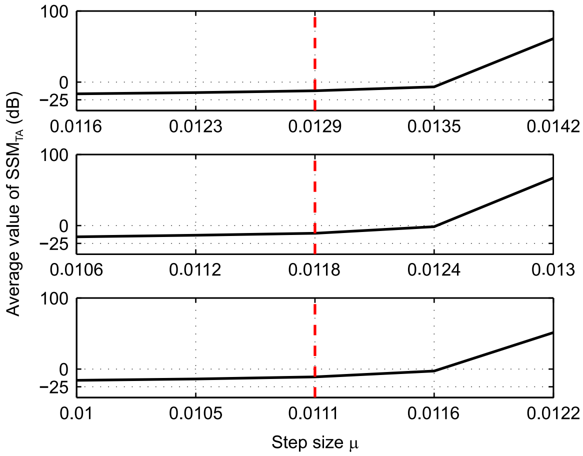

Fig. 3. $\overline{\text{SSM}}_{\text{TA}}$ (estimate of the steady-state mismatch) for the SR algorithm. $M=128$, $a=0.9$ (correlated data). Input distributions are uniform (top), Gaussian (middle), and SIRV (bottom). The vertical dashed line represents the numerically computed bound μ_0 according to (47). The x axis is segmented in intervals of length $0.05\mu_0$.

Fig. 3 shows results for the SR algorithm with a longer filter ($M=128$) and correlated data ($a=0.9$). The same considerations as in Fig. 2 apply, except that the bound μ_0 was computed according to (47). In this scenario, the theoretical result (49) predicts that for Gaussian regressors, the maximum step-size would be 0.0125. Although the bounds we obtained are very close to each other, when SIRV input is used the value $\mu=0.0125$ would lead to large instabilities. In all simulated scenarios for the SR algorithm, we consistently found that the bound for the uniform input was larger than the one for the Gaussian input, which in turn was larger than the one for the SIRV input; however, all three were within 10% to 15% of each other. Therefore, it would be easier than with the LMS algorithm to find a fairly conservative bound that would work in all three scenarios. However, our model already provides a tight bound μ_0 in all cases (we next study whether it is at most within 5 or 10% of the actual stability limit).

It is well known that the SSM of the LMS and SR algorithms will in general increase with μ [1]. However, according to these dynamics, how large of an increase could be considered "normal"? In all the simulated scenarios, the signal-to-noise ratio was set to 30 dB. In Figs. 2 and 3, it can be seen that the resulting $\overline{\text{SSM}}_{\text{TA}}$ values for $\mu<\mu_0$ were close to –20 dB. However, for $\mu>\mu_0$ they approach 0 dB or even larger. If this increase is "normal", it should happen on average for a large number of realizations. Then, the estimated standard error of the mean $\hat{\sigma}_{\overline{\text{SSM}}_{\text{TA}}}$, should be small. Moreover, according to (63), SSM_{TA} is the average of 500 random variables, so by central-limit-theorem arguments we expect it to be normally distributed. If the estimate of its standard deviation $\hat{\sigma}_{\text{SSM}_{\text{TA}}}$ (which in our case is $10\cdot\hat{\sigma}_{\overline{\text{SSM}}_{\text{TA}}}$) increases and become close to or even larger to the estimate of its mean, this would be an indication that the distribution of SSM_{TA} can no longer be seen as gaussian. This could be the case when some realizations of the algorithm show an unstable behavior, so the distribution becomes more heavy-tailed. Therefore, an increase in $\hat{\sigma}_{\overline{\text{SSM}}_{\text{TA}}}$ would be an indication that the source of variability is most likely due to unstable (or undesirable) behavior of the algorithm in some realizations.

Table 1 collects the results of $\overline{\text{SSM}}_{\text{TA}}$ and $\hat{\sigma}_{\overline{\text{SSM}}_{\text{TA}}}$ in all the scenarios with uniform input. In most cases, the tightness of μ_0 is evident. For the LMS case, $M=32$, $a=0.9$, and $\mu=1.1\mu_0$, $\overline{\text{SSM}}_{\text{TA}}$ increased 150% with respect

Table 1
$\overline{\text{SSM}}_{\text{TA}}\pm\hat{\sigma}_{\overline{\text{SSM}}_{\text{TA}}}$ for Uniform input. Under the conditions LMS, $M=32$, $a=0.9$, and $\mu=1.1\mu_0$, the range of SSM_{TA} (interval from minimum to maximum values of $\text{SSM}_{\text{TA}}^{(n)}$ computed over the 100 realizations) was equal to [0.0035,0.1976], whereas for $\mu=1.15\mu_0$, $\overline{\text{SSM}}_{TA}\pm\hat{\sigma}_{\overline{\text{SSM}}_{\text{TA}}}$ was equal to 0.1289 ± 0.0694. Under the conditions SR, $M=128$, $a=0.9$, and $\mu=1.05\mu_0$, the range was equal to [0.1151,0.7516]. Under the conditions LMS, $M=128$, $a=0.9$, and $\mu=1.05\mu_0$, the range was equal to [0.0296,0.2008].

Conditions	$\mu=0.95\mu_0$	$\mu=\mu_0$	$\mu=1.05\mu_0$	$\mu=1.1\mu_0$
SR, $M=32$, $a=0$	0.0133 ± 0.0005	0.0452 ± 0.0039	$(8.4\pm6.3)\times10^{26}$	–
LMS, $M=32$, $a=0$	0.0127 ± 0.0003	0.0454 ± 0.0021	$(8.4\pm8.3)\times10^{55}$	–
SR, $M=32$, $a=0.9$	0.0149 ± 0.0005	0.0321 ± 0.0016	1.22 ± 0.46	$(1.6\pm1.5)\times10^{50}$
LMS, $M=32$, $a=0.9$	0.0024 ± 0.0001	0.0036 ± 0.0001	0.0059 ± 0.0003	0.0149 ± 0.0028
SR, $M=128$, $a=0$	0.0099 ± 0.0002	0.0206 ± 0.0005	5.78 ± 1.81	$(10.1\pm5.4)\times10^{20}$
LMS, $M=128$, $a=0$	0.01 ± 0.0001	0.0231 ± 0.0003	$(2.2\pm2)\times10^{3}$	–
SR, $M=128$, $a=0.9$	0.0332 ± 0.0006	0.0596 ± 0.0014	0.211 ± 0.0082	$(1.3\pm0.2)\times10^{6}$
LMS, $M=128$, $a=0.9$	0.0119 ± 0.0002	0.0205 ± 0.0004	0.0575 ± 0.0018	$(3.1\pm2.8)\times10^{14}$

Table 2
$\overline{\mathrm{SSM}}_{\mathrm{TA}} \pm \hat{\sigma}_{\overline{\mathrm{SSM}}_{\mathrm{TA}}}$ for Gaussian input. Under the conditions LMS, $M=32$, $a=0.9$, and $\mu=1.15\mu_0$, $\overline{\mathrm{SSM}}_{\mathrm{TA}} \pm \hat{\sigma}_{\overline{\mathrm{SSM}}_{\mathrm{TA}}}$ was equal to 0.02 ± 0.0098, with the range being equal to [0.0018,0.9299]. Under the conditions LMS, $M=128$, $a=0.9$, and $\mu=1.05\mu_0$, the range was equal to [0.0121,0.0884].

Conditions	$\mu=0.95\mu_0$	$\mu=\mu_0$	$\mu=1.05\mu_0$	$\mu=1.1\mu_0$
SR, $M=32$, $a=0$	0.0137 ± 0.0007	0.0450 ± 0.005	$(4 \pm 2.9) \times 10^{20}$	–
LMS, $M=32$, $a=0$	0.0122 ± 0.0004	0.0387 ± 0.0029	$(2.1 \pm 2) \times 10^{42}$	–
SR, $M=32$, $a=0.9$	0.0210 ± 0.0009	0.0755 ± 0.0085	$(4.3 \pm 2.5) \times 10^{7}$	–
LMS, $M=32$, $a=0.9$	0.0015 ± 0.0001	0.0017 ± 0.0001	0.0025 ± 0.0002	0.0031 ± 0.0002
SR, $M=128$, $a=0$	0.0097 ± 0.0002	0.0229 ± 0.0009	6.03 ± 1.69	$(5.8 \pm 1.7) \times 10^{17}$
LMS, $M=128$, $a=0$	0.0096 ± 0.0001	0.0213 ± 0.0003	116 ± 18	$(6.8 \pm 2.8) \times 10^{28}$
SR, $M=128$, $a=0.9$	0.0442 ± 0.0011	0.0850 ± 0.0028	0.7 ± 0.06	$(5 \pm 1.6) \times 10^{6}$
LMS, $M=128$, $a=0.9$	0.0080 ± 0.0001	0.0129 ± 0.0004	0.0219 ± 0.0011	0.1501 ± 0.0271

Table 3
$\overline{\mathrm{SSM}}_{\mathrm{TA}} \pm \hat{\sigma}_{\overline{\mathrm{SSM}}_{\mathrm{TA}}}$ for SIRV input. Under the conditions LMS, $M=32$, $a=0.9$, and $\mu=1.15\mu_0$, $\overline{\mathrm{SSM}}_{\mathrm{TA}} \pm \hat{\sigma}_{\overline{\mathrm{SSM}}_{\mathrm{TA}}}$ was equal to 0.0111 ± 0.0083. Under the conditions LMS, $M=128$, $a=0$, and $\mu=1.05\mu_0$, the range was equal to [0.0035,0.0700]. Under the conditions LMS, $M=128$, $a=0.9$, and $\mu=1.05\mu_0$, the range was equal to [0.0047,0.1520].

Conditions	$\mu=0.95\mu_0$	$\mu=\mu_0$	$\mu=1.05\mu_0$	$\mu=1.1\mu_0$
SR, $M=32$, $a=0$	0.0121 ± 0.0006	0.0348 ± 0.0045	$1.7 \pm 1.5) \times 10^{3}$	–
LMS, $M=32$, $a=0$	0.0055 ± 0.0005	0.0091 ± 0.0007	0.0394 ± 0.0155	1.09 ± 0.66
SR, $M=32$, $a=0.9$	0.0183 ± 0.0007	0.0515 ± 0.0059	2.43 ± 0.93	$(1.4 \pm 1.3) \times 10^{27}$
LMS, $M=32$, $a=0.9$	$(15 \pm 0.4) \times 10^{-4}$	0.0011 ± 0.0001	0.0019 ± 0.0003	0.0016 ± 0.0001
SR, $M=128$, $a=0$	0.0096 ± 0.0002	0.0191 ± 0.0009	2.3 ± 1.79	$(2.6 \pm 2.1) \times 10^{14}$
LMS, $M=128$, $a=0$	0.0038 ± 0.0001	0.0068 ± 0.0004	0.0153 ± 0.0013	0.1588 ± 0.0542
SR, $M=128$, $a=0.9$	0.0401 ± 0.001	0.0794 ± 0.0028	0.52 ± 0.05	$(1.4 \pm 0.6) \times 10^{5}$
LMS, $M=128$, $a=0.9$	0.005 ± 0.0002	0.0079 ± 0.0006	0.0128 ± 0.0017	0.0565 ± 0.0223

to the value when $\mu=1.05\mu_0$ (whereas in the other conditions analyzed in this table the increase was below 50%), while $\hat{\sigma}_{\overline{\mathrm{SSM}}_{\mathrm{TA}}}$ increased 10 times. Also, while the minimum value of $\mathrm{SSM}_{\mathrm{TA}}^{(n)}$ was close to the mean value when $\mu=0.95\mu_0$ (stable), the maximum value was 50 times larger. All of these facts lead us to believe that the behavior of the algorithm under these conditions is no longer stable. A similar analysis can be done for the SR case.

Tables 2 and 3 show comparable results for Gaussian and SIRV inputs, respectively. The same ideas used for analyzing the previous table can be applied here. The LMS case, $M=32$, $a=0.9$ proved to be the least tight, requiring a 15% increase from μ_0 to lose stability.

It can be seen that in all scenarios tested, the algorithms were stable when using $\mu=0.95\mu_0$, so the sufficiency of the proposed bound μ_0 is well established. Moreover, μ_0 was within 5% of the actual stability limit in all but one of the tested conditions.

7. Conclusions

We have presented an analysis that allows us to evaluate the mean and mean-square stability of a large class of adaptive filters. Without a full transient model, which is the usual approach in the literature, we were able to obtain sufficient conditions on the stability, and without restricting to the Gaussian case. In several cases of interest, the conditions obtained are tight enough or even necessary. Some well-known results, as well as some new ones, were also obtained for popular adaptive filters. The simulation results show the tightness of the bound derived for several cases of interest.

Appendix A. Proof of Theorem 1

We begin with the following lemma which will be useful:

Lemma 1. *Given a positive doubly-indexed sequence* $g(i,j)$ *such that* $\sum_{j=k}^{i} g(i,j+1) \le N$, $\forall\, i \ge k$ *with* $N>0$, *and such that* $g(i,k+1) \le g(i,j+1)g(j,k+1)\ \forall\, i \ge j \ge k$, $\exists\ 0<\gamma<1$ *and* $L(k)>0$ *such that*

$$g(i,k+1) \le L(k)\gamma^{i-k}. \tag{A.1}$$

Proof. From $g(i,k+1) \le g(i,j+1)g(j,k+1)$, we obtain

$$g(i,j+1) \ge \frac{g(i,k+1)}{g(j,k+1)}. \tag{A.2}$$

Then, it is clear that

$$\sum_{j=k}^{i} g(i,j+1) \ge \sum_{j=k}^{i} \frac{g(i,k+1)}{g(j,k+1)} = g(i,k+1)\sum_{j=k}^{i}\frac{1}{g(j,k+1)}, \tag{A.3}$$

which leads to

$$\sum_{j=k}^{i}\frac{1}{g(j,k+1)} \le \frac{N}{g(i,k+1)}. \tag{A.4}$$

Using (A.4), we can obtain

$$\sum_{j=k}^{i}\frac{1}{g(j,k+1)} \geq \left(1+\frac{1}{N}\right)\sum_{j=k}^{i-1}\frac{1}{g(j,k+1)}. \tag{A.5}$$

Through repeated application of the last reasonings, we can obtain

$$\sum_{j=k}^{i}\frac{1}{g(j,k+1)} \geq \left(1+\frac{1}{N}\right)^{i-k}\frac{1}{g(k,k+1)}. \tag{A.6}$$

Taking (A.4) and (A.6), we get

$$g(i,k+1) \leq Ng(k,k+1)\left(\frac{N}{N+1}\right)^{i-k}, \tag{A.7}$$

from which we obtain the desired result by setting $\gamma = N/(1+N)$ and $L(k) = Ng(k,k+1)$.

Now we will prove Theorem 1. We start by proving the sufficiency. In order to do that we will analyze the four terms in (11) separately. We begin with the term in (14). We have the following

$$\tilde{\mathbf{c}}_0^T\mathbf{D}(i,0)\tilde{\mathbf{c}}_0 \leq \|\tilde{\mathbf{c}}_0\|^2\mathrm{tr}[\mathbf{D}(i,0)] \leq \|\tilde{\mathbf{c}}_0\|^2 N(\boldsymbol{\mu})\gamma^{i+1}(\boldsymbol{\mu}), \tag{A.8}$$

which goes to zero as $i\to\infty$. For the term in (15) we can write

$$\sum_{j=0}^{i}\tilde{\mathbf{c}}_0^T\mathbf{B}^{j+1}\mathbf{D}(i,j+1)\mathbf{c} = \sum_{j=0}^{i}\mathrm{tr}[\mathbf{c}\tilde{\mathbf{c}}_0^T\mathbf{B}^{j+1}\mathbf{D}(i,j+1)]. \tag{A.9}$$

For each of the terms on the RHS of (A.9), we apply the Cauchy–Schwartz inequality and obtain

$$|\mathrm{tr}[\mathbf{c}\tilde{\mathbf{c}}_0^T\mathbf{B}^{j+1}\mathbf{D}(i,j+1)]| \leq |\tilde{\mathbf{c}}_0^T\mathbf{c}| \times \{\mathrm{tr}[\mathbf{D}(i,j+1)\mathbf{B}^{j+1}\mathbf{B}^{j+1}\mathbf{D}(i,j+1)]\}^{1/2}. \tag{A.10}$$

We also have

$$\begin{aligned}|\mathrm{tr}[\mathbf{c}\tilde{\mathbf{c}}_0^T\mathbf{B}^{j+1}\mathbf{D}(i,j+1)]| &\leq |\tilde{\mathbf{c}}_0^T\mathbf{c}|\lambda_{\max}[\mathbf{D}(i,j+1)]\mathrm{tr}[\mathbf{B}^{j+1}\mathbf{B}^{j+1}]^{1/2}\\ &\leq N(\boldsymbol{\mu})\gamma^{i-j}(\boldsymbol{\mu})\mathrm{tr}[\mathbf{B}^{j+1}\mathbf{B}^{j+1}]^{1/2}.\end{aligned} \tag{A.11}$$

We can then write

$$|\mathrm{tr}[\mathbf{c}\tilde{\mathbf{c}}_0^T\mathbf{B}^{j+1}\mathbf{D}(i,j+1)]| \leq |\tilde{\mathbf{c}}_0^T\mathbf{c}|M^{1/2}N(\boldsymbol{\mu})\gamma^{i-j}(\boldsymbol{\mu})|\lambda_{\max}(\mathbf{B})|^{j+1}. \tag{A.12}$$

As (9) guarantees that $|\lambda_{\max}(\mathbf{B})| < 1$, we have

$$|\mathrm{tr}[\mathbf{c}\tilde{\mathbf{c}}_0^T\mathbf{B}^{j+1}\mathbf{D}(i,j+1)]| \leq |\tilde{\mathbf{c}}_0^T\mathbf{c}|M^{1/2}N(\boldsymbol{\mu})\gamma^{i-j}(\boldsymbol{\mu}), \tag{A.13}$$

which implies that the term in (15) is bounded for all i.

For the term (16), we can write

$$\sigma_v^2\sum_{j=0}^{i}\mathrm{tr}[\tilde{\mathbf{F}}_{xx}\mathbf{D}(i,j+1)] \leq \sigma_v^2\lambda_{\max}(\tilde{\mathbf{F}}_{xx})N(\boldsymbol{\mu})\sum_{j=0}^{i}\gamma^{i-j}(\boldsymbol{\mu}) < \infty. \tag{A.14}$$

Then, it only remains to analyze the term (17). Using the same previous reasoning, it can be seen that

$$|\mathbf{c}^T\mathbf{D}(i,k+1)\mathbf{B}^{k-j}\mathbf{c}| \leq \|\mathbf{c}\|^2M^{1/2}N(\boldsymbol{\mu})\gamma^{i-k}(\boldsymbol{\mu})|\lambda_{\max}(\mathbf{B})|^{k-j}, \tag{A.15}$$

$$|\mathbf{c}^T\mathbf{B}^{j-k}\mathbf{D}(i,j+1)\mathbf{c}| \leq \|\mathbf{c}\|^2M^{1/2}N(\boldsymbol{\mu})\gamma^{i-j}(\boldsymbol{\mu})|\lambda_{\max}(\mathbf{B})|^{j-k}. \tag{A.16}$$

Consider the first term in (17):

$$\begin{aligned}(1-\alpha)^2\sum_{j=0}^{i}\sum_{k\geq j}^{i}|\mathbf{c}^T\mathbf{D}(i,k+1)\mathbf{B}^{k-j}\mathbf{c}|\\ \leq \|\mathbf{c}\|^2M^{1/2}N(\boldsymbol{\mu})\sum_{j=0}^{i}\sum_{k=j}^{i}\gamma^{i-k}(\boldsymbol{\mu})|\lambda_{\max}(\mathbf{B})|^{k-j}.\end{aligned} \tag{A.17}$$

Making the change of variables $p=k-j$, we write the RHS of (A.17) as

$$\|\mathbf{c}\|^2M^{1/2}N(\boldsymbol{\mu})\sum_{j=0}^{i}\gamma^{i-j}(\boldsymbol{\mu})\sum_{p=0}^{i-j}\left[\frac{|\lambda_{\max}(\mathbf{B})|}{\gamma(\boldsymbol{\mu})}\right]^p. \tag{A.18}$$

The following fact is known

$$\sum_{p=0}^{i-j}\left[\frac{|\lambda_{\max}(\mathbf{B})|}{\gamma(\boldsymbol{\mu})}\right]^p = \begin{cases} i-j+1, & \text{if } \dfrac{|\lambda_{\max}(\mathbf{B})|}{\gamma(\boldsymbol{\mu})}=1,\\[2ex] \dfrac{1-\left[\dfrac{|\lambda_{\max}(\mathbf{B})|}{\gamma(\boldsymbol{\mu})}\right]^{i-j+1}}{1-\dfrac{|\lambda_{\max}(\mathbf{B})|}{\gamma(\boldsymbol{\mu})}}, & \text{otherwise.}\end{cases} \tag{A.19}$$

Using this last result allows us to show that (A.18) is bounded for all i. In the same manner, we can show that the second term in (17) is also bounded for all i. Combining all the results, we conclude that $\lim_{i\to\infty}E[\|\tilde{\mathbf{c}}_{i+1}\|^2] < \infty$, proving the sufficiency part of the theorem.

The necessity can be proved as follows. First, we have

$$E[\|\tilde{\mathbf{c}}_{i+1}\|^2] = E\{\|\tilde{\mathbf{c}}_{i+1}-E[\tilde{\mathbf{c}}_{i+1}]\|^2\} + \|E[\tilde{\mathbf{c}}_{i+1}]\|^2. \tag{A.20}$$

Clearly, if $\lim_{i\to\infty}E[\|\tilde{\mathbf{c}}_{i+1}\|^2] < \infty$, we must have $\lim_{i\to\infty}\|E[\tilde{\mathbf{c}}_{i+1}]\|^2 < \infty$, which implies condition (9). Secondly, looking at (11) and (14)–(17), we see that the only term that is independent of the initial condition and the system to be estimated is (16). In the special case when $\mathbf{c}=\mathbf{0}$, if $\lim_{i\to\infty}E[\|\tilde{\mathbf{c}}_{i+1}\|^2] < \infty$, we must have that (16) is bounded for all i. Obviously the same must be true for any $\mathbf{c}\in\mathbb{R}^M$. In explicit terms

$$\exists N > 0 \text{ such that } \sum_{j=k}^{i}\mathrm{tr}[\tilde{\mathbf{F}}_{xx}\mathbf{D}(i,j+1)] < N,\ \forall\ i \geq k. \tag{A.21}$$

As $\tilde{\mathbf{F}}_{xx} > 0$, we have

$$\lambda_{\min}(\tilde{\mathbf{F}}_{xx})\sum_{j=k}^{i}\mathrm{tr}[\mathbf{D}(i,j+1)] < N, \quad \forall\ i \geq k. \tag{A.22}$$

Defining

$$\mathbf{E}(i) = \alpha\mathbf{I} - \tilde{\mathbf{f}}(\mathbf{x}_i)\tilde{\mathbf{x}}_i^T, \tag{A.23}$$

and assuming that $j \geq k$, we can write

$$\mathrm{tr}[\mathbf{D}(i,k+1)] = \mathrm{tr}\{E[\mathbf{E}^T(k+1)\ldots\mathbf{E}^T(i)\mathbf{E}(i)\ldots\mathbf{E}(k+1)]\}. \tag{A.24}$$

Using the independence assumption, it follows that

$$\mathrm{tr}[\mathbf{D}(i,k+1)] = \mathrm{tr}\{\mathbf{D}(i,j+1)E[\mathbf{E}(j)\ldots\mathbf{E}(k+1)\mathbf{E}^T(k+1)\ldots\mathbf{E}^T(j)]\}. \tag{A.25}$$

Finally, from the last equation, it is easy to show that

$$\mathrm{tr}[\mathbf{D}(i,k+1)] \leq \mathrm{tr}[\mathbf{D}(i,j+1)]\mathrm{tr}[\mathbf{D}(j,k+1)], \quad j \geq k. \tag{A.26}$$

This allows us to use Lemma 1 to obtain

$$\mathrm{tr}[\mathbf{D}(i,k+1)] \le \frac{NM}{\lambda_{\min}(\tilde{\mathbf{F}}_{xx})}\left[\frac{N}{\lambda_{\min}(\tilde{\mathbf{F}}_{xx})+N}\right]^{i-k}, \tag{A.27}$$

concluding the proof of the necessity part.

References

[1] S. Haykin, Adaptive Filter Theory, Prentice Hall, Upper Saddle River, NJ, 2001.

[2] B. Farhang-Boroujeny, Adaptive Filters: Theory and Applications, Wiley, New York, 1998.

[3] T.Y. Al-Naffouri, A.H. Sayed, Transient analysis of data normalized adaptive filters, IEEE Trans. Signal Process. 51 (3) (2003) 639–652.

[4] T.Y. Al-Naffouri, A.H. Sayed, Transient analysis of adaptive filters with error nonlinearities, IEEE Trans. Signal Process. 51 (3) (2003) 653–663.

[5] J. Benesty, T. Gaensler, D.R. Morgan, M.M. Sondhi, S.L. Gay, Advances in Network and Acoustic Echo Cancellation, Springer-Verlag, Berlin, 2001.

[6] W.A. Sethares, I.M.Y. Mareels, B.D.O. Anderson, C.R. Johnson, R.R. Bitmead, Excitation conditions for signed regressor least mean squares adaptation, IEEE Trans. Circuits and Syst. 35 (6) (1988) 613–624.

[7] R.C. Bradley, Basic properties of strong mixing conditions: a survey and some open questions, Probab. Surv. 2 (November) (2005) 107–144.

[8] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1985.

[9] A. Feuer, E. Weinstein, Convergence analysis of LMS filters with uncorrelated Gaussian data, IEEE Trans. Acoustics, Speech and Signal Process. 33 (1) (1985) 222–230.

[10] K. Mayyas, T. Aboulnasr, Leaky LMS algorithm: MSE analysis for Gaussian data, IEEE Trans. Signal Process. 45 (4) (1997) 927–934.

[11] A.H. Sayed, T.Y. Al-Naffouri, Mean-square analysis of normalized leaky adaptive filters. In: Proceedings of ICASSP, Salt Lake City, vol. 6, May 2001, pp. 3873–3876.

[12] Ali H. Sayed, Adaptive Filters, John Wiley & Sons, 2008.

[13] T.Y. Al-Naffouri, M. Moinuddin, Exact performance analysis of the ε-NLMS algorithm for colored circular Gaussian inputs, IEEE Trans. Signal Process. 58 (10) (2010) 5080–5090.

[14] R. Price, A useful theorem for nonlinear devices having Gaussian inputs, IRE Trans. Inform. Theory IT-4 (June) (1958) 69–72.

[15] E. Eweda, Analysis and design of signed regressor LMS algorithms for stationary and nonstationary adaptive filtering with correlated Gaussian data, IEEE Trans. Circuits and Syst. 37 (11) (1990) 1367–1374.

[16] S.P. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[17] K. Yao, A representation theorem and its applications to spherically-invariant random processes, IEEE Trans. Inform. Theory 19 (5) (1973) 600–608.

[18] H. Brehm, W. Stammler, Description and generation of spherically invariant speech-model signals, Signal Process. 12 (March) (1987) 119–141.

[19] M. Rangaswamy, D. Weiner, A. Öztürk, Computer generation of correlated non-Gaussian radar clutter, IEEE Trans. Aerosp. Electron. Syst. 31 (1) (1995) 106–116.

[20] P. Embrechts, F. Lindskog, A.J. McNeil, Modelling dependence with copulas and applications to risk management, in: S.T. Rachev (Ed.), Handbook of Heavy Tailed Distributions in Finance, Elsevier, 2001.