Signal Processing 91 (2011) 2615–2622

Contents lists available at ScienceDirect

Signal Processing

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

Stochastic analysis of an error power ratio scheme applied to the affine combination of two LMS adaptive filters ☆

José Carlos M. Bermudez [a,*], Neil J. Bershad [b], Jean-Yves Tourneret [c]

[a] Depto. Engenharia Elétrica, Universidade Federal de Santa Catarina, Trindade, Florianópolis, SC 88040-970, Brazil
[b] University of California Irvine, CA, USA
[c] IRIT – ENSEEIHT – TésA, Toulouse, France

ARTICLE INFO

Article history:
Received 23 November 2010
Received in revised form
9 May 2011
Accepted 27 May 2011
Available online 2 June 2011

Keywords:
Adaptive filtering
System identification
Least mean-square algorithm
Affine combination

ABSTRACT

The affine combination of two adaptive filters that simultaneously adapt on the same inputs has been actively investigated. In these structures, the filter outputs are linearly combined to yield a performance that is better than that of either filter. Various decision rules can be used to determine the time-varying parameter for combining the filter outputs. A recently proposed scheme based on the ratio of error powers of the two filters has been shown by simulation to achieve nearly optimum performance. The purpose of this paper is to present a first analysis of the statistical behavior of this error power scheme for white Gaussian inputs. Expressions are derived for the mean behavior of the combination parameter and for the adaptive weight mean-square deviation. Monte Carlo simulations show good to excellent agreement with the theoretical predictions.

1. Introduction

The design of many adaptive filters requires a trade-off between convergence speed and steady-state mean-square error (MSE). In general, a faster (slower) convergence speed yields a larger (smaller) steady-state mean-square deviation (MSD) and MSE. This trade-off is usually controlled by some design parameter of the weight update, such as a step size, a regularization parameter or a forgetting factor. Variable step-size modifications of the basic adaptive algorithms offer a possible solution to this design problem [1–4].

Recently, a novel scheme has been proposed in [5] which uses a convex combination of two fixed step-size adaptive filters as shown in Fig. 1. The key to this scheme is the selection of the scalar mixing parameter $\lambda(n)$ for combining the two filter outputs. The mixing parameter is defined in [6] as a sigmoid function whose free parameter is adaptively optimized using a stochastic gradient search which minimizes the quadratic error of the overall filter. The performance of this adaptive scheme has been recently studied in [7,8]. The convex combination performed as well as the best of its components in the MSE sense. These results indicate that a combination of adaptive filters can lead both to fast convergence rates and good steady-state performance, an attribute that is usually obtained only in variable step-size algorithms. Thus, there is great interest in learning more about the properties of such adaptive structures.

More recently [9] the convex combination has been generalized to an affine combination, in which $\lambda(n)$ is not restricted to the interval (0,1). Fig. 1 was interpreted from the viewpoint of a linear combiner where each adaptive filter is estimating the unknown channel impulse response using the same input data. The achievable performance

☆ This work was partly supported by CNPq under Grant nos. 305377/2009-4 and 473123/2009-6.

* Corresponding author.
E-mail addresses: j.bermudez@ieee.org (J.C.M. Bermudez), bershad@ece.uci.edu (N.J. Bershad), Jean-Yves.Tourneret@enseeiht.fr (J.-Y. Tourneret).

0165-1684/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2011.05.017

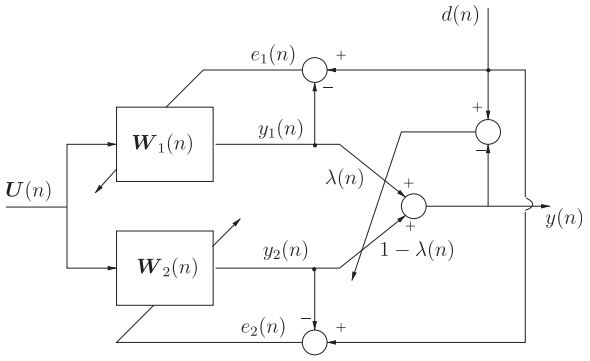

Fig. 1. Adaptive combining of two transversal adaptive filters.

was studied for an affine combination of two LMS adaptive filters with a white Gaussian input. The optimal combining parameter sequence $\lambda_o(n)$ was determined which minimizes the weight mean-square deviation (MSD). This optimal affine combiner is not realizable, as its design requires the knowledge of the unknown response. Nevertheless, its performance provides an upper bound on the performance of any realizable affine combiner. The analysis of the affine combination of adaptive filters has been further expanded in [10,11]. Affine and convex combinations have been compared in [12–14].

Two realizable schemes for updating $\lambda(n)$ were proposed in [9]. The first scheme is based on a stochastic gradient approximation to $\lambda_o(n)$. This scheme has been analyzed in [11]. The second scheme is based on the relative values of averaged estimates of the individual error powers. Though simulation results using this scheme have shown performances very close to the optimum, so far there is no available analytical model for its performance.

This paper presents[1] a first analysis of the stochastic behavior of the error power scheme proposed in [9]. The analysis assumes a stationary environment and a white Gaussian input signal.[2] The mean behavior of $\lambda(n)$ is studied by replacing the involved random variables by their means resulting in expressions for $E[\lambda(n)]$ ($E[\cdot]$ meaning statistical expectation) involving the time-varying weight MSDs of the individual filters, the cross MSD between filters and the background noise. Monte Carlo simulations show excellent agreement with the theoretical predictions based on the analytical model.

This paper is organized as follows. Section 2 briefly reviews the main expressions derived in [9] for the behavior of an affine combination for a white Gaussian input, as well as the expression for $\lambda(n)$ that characterizes the error power scheme. Section 3 presents the derivation of the analytical models for the behaviors of $E[\lambda(n)]$ and the weight MSD of the combination. Section 4 presents simulation results that verify the quality of the analytical model. Section 5 concludes the paper.

2. The affine combiner

The system under investigation is shown in Fig. 1. Each individual filter uses the LMS adaptation rule but with different step sizes μ_i, $i=1,2$:

$$\boldsymbol{W}_i(n+1)=\boldsymbol{W}_i(n)+\mu_i e_i(n)\boldsymbol{U}(n), \quad i=1,2 \tag{1}$$

where

$$e_i(n)=d(n)-\boldsymbol{W}_i^T(n)\boldsymbol{U}(n), \quad i=1,2 \tag{2}$$

$$d(n)=e_o(n)+\boldsymbol{W}_o^T\boldsymbol{U}(n) \tag{3}$$

where $\boldsymbol{W}_o$ is the unknown weight vector, $\boldsymbol{W}_i(n)$, $i=1,2$, are the N-dimensional adaptive coefficient vectors, and $e_o(n)$ is assumed zero-mean, i.i.d. with variance σ_o^2 and statistically independent of any other signal in the system. $\boldsymbol{U}(n)=[u(n),\ldots,u(n-N+1)]^T$ is the input vector. The input process $u(n)$ is assumed to be white, Gaussian, with zero-mean, variance σ_u^2 and conditional correlation matrix $\boldsymbol{R}_u=E[\boldsymbol{U}(n)\boldsymbol{U}^T(n)|\boldsymbol{W}_1(n),\boldsymbol{W}_2(n)]$.

In the following analysis, the input vector at time n is assumed statistically independent of the weights at time n (independence theory). Thus, $\boldsymbol{R}_u=E[\boldsymbol{U}(n)\boldsymbol{U}^T(n)]=\sigma_u^2\boldsymbol{I}$. It is also assumed that the errors $e_1(n)$ and $e_2(n)$ are zero-mean, white over time and conditionally Gaussian given $\boldsymbol{W}_1(n)$ and $\boldsymbol{W}_2(n)$. Finally, it will also be assumed, without loss, that $\mu_1\geq\mu_2$, so that $\boldsymbol{W}_1(n)$ will, in general, converge faster than $\boldsymbol{W}_2(n)$. Also, $\boldsymbol{W}_2(n)$ will converge to the lowest individual steady-state weight misadjustment. The stochastic analysis of each individual adaptive filter behavior in (1) is well-known [16–18].

The outputs of the two filters are combined as shown in Fig. 1,

$$y(n)=\lambda(n)y_1(n)+[1-\lambda(n)]y_2(n) \tag{4}$$

where $y_i(n)=\boldsymbol{W}_i^T(n)\boldsymbol{U}(n)$, $i=1,2$, $\lambda(n)$ can be any real number[3] and the overall system error is given by

$$e(n)=d(n)-y(n) \tag{5}$$

The adaptive filter output combination (4) is an affine combination, as $y(n)$ can assume any value on the real line. This setup generalizes the combination of adaptive filter outputs, and can be used to study the properties of the optimal combination.

The optimum instantaneous value $\lambda_o(n)$ for $\lambda(n)$ for white inputs has been determined in [9] to satisfy

$$[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)]^T[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)]\lambda_o(n) = [\boldsymbol{W}_o-\boldsymbol{W}_2(n)]^T[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)] \tag{6}$$

Assuming that, in steady-state, $\lambda_o(n)$ is slowly varying in comparison with other term on the l.h.s. of (6) and

[1] Part of this work has been presented in [15].

[2] The analysis is this paper is restricted to the white input case in order to simplify the understanding and the interpretation of the results. It appears to us that extension to the colored input case is straightforward using the results in [11].

[3] This case corresponds to an affine (as opposed to convex) combination. The output in (4) can have any real value on the line containing $y_1(n)$ and $y_2(n)$. $y(n)$ is restricted to the points on the line between $y_1(n)$ and $y_2(n)$ in the convex combination case.

taking the expectation of both sides yields

$$\lim_{n\to\infty} E\{[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)]^T[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)]\}E[\lambda_o(n)] = \lim_{n\to\infty} E\{[\boldsymbol{W}_o-\boldsymbol{W}_2(n)]^T[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)]\} \quad (7)$$

Thus,

$$\lim_{n\to\infty} E[\lambda_o(n)] \simeq \lim_{n\to\infty} \frac{E[\boldsymbol{W}_2^T(n)\boldsymbol{W}_2(n)]-E[\boldsymbol{W}_2^T(n)\boldsymbol{W}_1(n)]}{E\{[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)]^T[\boldsymbol{W}_1(n)-\boldsymbol{W}_2(n)]\}} \quad (8)$$

It has also been determined in [9] that

$$\begin{aligned} E[\boldsymbol{W}_i^T(n+1)\boldsymbol{W}_j(n+1)] &= [1-(\mu_i+\mu_j)\sigma_u^2+(N+2)\mu_i\mu_j\sigma_u^4] \times E[\boldsymbol{W}_i^T(n)\boldsymbol{W}_j(n)]+\mu_i\sigma_u^2[1-(N+2)\mu_j\sigma_u^2]\boldsymbol{W}_o^T E[\boldsymbol{W}_j(n)] \\ &+\mu_j\sigma_u^2[1-(N+2)\mu_i\sigma_u^2]\boldsymbol{W}_o^T E[\boldsymbol{W}_i(n)] \\ &+\mu_i\mu_j\sigma_u^4\left[N\left(\frac{\sigma_o^2}{\sigma_u^2}\right)+(N+2)\boldsymbol{W}_o^T\boldsymbol{W}_o\right]. \end{aligned} \quad (9)$$

where [16]

$$E[\boldsymbol{W}_k(n)] = (1-\mu_k\sigma_u^2)^n E[\boldsymbol{W}_k(0)]+[1-(1-\mu_k\sigma_u^2)^n]\boldsymbol{W}_o \quad (10)$$

for $k=1,2$, $i=1,2$ and $j=1,2$,

$$\lim_{n\to\infty} E[\boldsymbol{W}_2^T(n)\boldsymbol{W}_1(n)] = \boldsymbol{W}_o^T\boldsymbol{W}_o + \frac{\mu_1\mu_2 N\sigma_o^2}{(\mu_1+\mu_2)-\mu_1\mu_2(N+2)\sigma_u^2} \quad (11)$$

and

$$\lim_{n\to\infty} E[\boldsymbol{W}_i^T(n)\boldsymbol{W}_i(n)] = \boldsymbol{W}_o^T\boldsymbol{W}_o + \frac{\mu_i N\sigma_o^2}{2-\mu_i(N+2)\sigma_u^2}, \quad i=1,2 \quad (12)$$

Rewriting (8) as a function of (11) and (12) yields

$$\lim_{n\to\infty} E[\lambda_o(n)] \simeq \lim_{n\to\infty} \frac{1}{1+\dfrac{E[\boldsymbol{W}_1^T(n)\boldsymbol{W}_1(n)]-E[\boldsymbol{W}_2^T(n)\boldsymbol{W}_1(n)]}{E[\boldsymbol{W}_2^T(n)\boldsymbol{W}_2(n)]-E[\boldsymbol{W}_2^T(n)\boldsymbol{W}_1(n)]}} \quad (13)$$

It has been shown in [11] that (13) can be approximated by

$$\lim_{n\to\infty} E[\lambda_o(n)] \simeq \frac{\delta(2-\mu_1 N\sigma_u^2)}{2(\delta-1)} \quad (14)$$

where $\delta=(\mu_2/\mu_1)<1$.

3. Error power based mixing parameter updating scheme

A function of time averaged error powers has been shown in [9] to be a good candidate for an estimator of the optimum $\lambda(n)$ for each n. The individual adaptive error powers are good indicators of the contribution of each adaptive output to the quality of the present estimation of $d(n)$. These errors are readily available and do not need an estimate of the additive noise power.

Consider a uniform sliding time average of the instantaneous errors

$$\hat{e}_1^2(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} e_1^2(m) \quad (15)$$

$$\hat{e}_2^2(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} e_2^2(m) \quad (16)$$

where K is the averaging window length. Then, the instantaneous value of $\lambda(n)$ is determined as

$$\lambda(n) = 1-\kappa g[\xi(n)], \quad \xi(n) = \frac{\hat{e}_1^2(n)}{\hat{e}_2^2(n)} \quad (17)$$

where $g[\xi(n)]$ is a nonlinear function that tends to one as $\xi(n)$ tends to infinity and to a value much less than one as $\xi(n)$ tends to zero. Thus, $y(n)$ tends to $y_1(n)$ when $e_1^2(n) \ll e_2^2(n)$ and to $y_2(n)$ when $e_2^2(n) \ll e_1^2(n)$, which is the desired combined adaptive behavior when $\mu_1>\mu_2$ in Fig. 1 [9].

This paper generalizes the error power ratio scheme proposed in [9] and analyzed in [15] to a generic nonlinearity $g[\xi(n)]$. Examples of functions $g[\xi(n)]$ with the desirable properties described above are the hyperbolic tangent function

$$g_1[\xi(n)] = \tanh\left[\frac{\xi(n)}{2}\right] \quad (18)$$

and the erf function,

$$g_2[\xi(n)] = \mathrm{erf}[\xi(n)] = \frac{2}{\sqrt{\pi}}\int_0^{\xi(n)} e^{-t^2}\,dt \quad (19)$$

Eqs. (17) and either (18) or (19) allow $\lambda(n)$ to vary smoothly over $(1-\kappa,1)$ as required.

3.1. The value of κ

The value of κ in (17) controls the value of $\lambda(n)$ as $n\to\infty$. Ideally, it should be selected so that

$$\lim_{n\to\infty} E[\lambda(n)] \simeq \lim_{n\to\infty} E[\lambda_o(n)] \quad (20)$$

Taking expectations of both sides of (17) and approximating the random variables by their means, a first order approximation of $E[\lambda(n)]$ is obtained as

$$E[\lambda(n)] \simeq 1-\kappa g\left\{\frac{E[\hat{e}_1^2(n)]}{E[\hat{e}_2^2(n)]}\right\} \quad (21)$$

However,

$$\begin{aligned} E[\hat{e}_i^2(n)] &= \frac{1}{K}\sum_{m=n-K+1}^{n} E[e_i^2(m)] \\ &= \sigma_o^2+\frac{\sigma_u^2}{K}\sum_{m=n-K+1}^{n}\mathrm{MSD}_i(m), \quad i=1,2 \end{aligned} \quad (22)$$

where

$$\begin{aligned} \mathrm{MSD}_i(m) &= E\{[\boldsymbol{W}_o-\boldsymbol{W}_i(m)]^T[\boldsymbol{W}_o-\boldsymbol{W}_i(m)]\} \\ &= \boldsymbol{W}_o^T\boldsymbol{W}_o-2\boldsymbol{W}_o^T E[\boldsymbol{W}_i(m)]+E[\boldsymbol{W}_i^T(m)\boldsymbol{W}_i(m)], \quad i=1,2 \end{aligned} \quad (23)$$

Using (22) in (21) and taking the limit as $n\to\infty$ yields

$$\lim_{n\to\infty} E[\lambda(n)] \simeq 1-\kappa g\left[\frac{\sigma_o^2+\sigma_u^2\,\mathrm{MSD}_1(\infty)}{\sigma_o^2+\sigma_u^2\,\mathrm{MSD}_2(\infty)}\right] \quad (24)$$

Finally, equating (24) to (14) and solving for κ yields

$$\kappa = \left[1-\frac{\delta(2-\mu_1 N\sigma_u^2)}{2(\delta-1)}\right]\left\{g\left[\frac{\sigma_o^2+\sigma_u^2\,\mathrm{MSD}_1(\infty)}{\sigma_o^2+\sigma_u^2\,\mathrm{MSD}_2(\infty)}\right]\right\}^{-1} \quad (25)$$

where $\mathrm{MSD}_1(\infty)$ and $\mathrm{MSD}_2(\infty)$ can be obtained from (9) and (10) for a given σ_o^2, σ_u^2, μ_1 and μ_2.

3.2. Mean behavior of $\lambda(n)$

An analytical model for the mean behavior of (17) is derived by approximating the expected value of $g[\xi(n)]$ by a second order expansion. Defining

$$\eta(n)=E[\xi(n)] \quad \text{and} \quad \sigma_\xi^2(n)=E[\xi^2(n)]-\eta^2(n) \tag{26}$$

a second order approximation of $E\{g[\xi(n)]\}$ is [19, p. 113, Eq. (5.55)]

$$E\{g[\xi(n)]\} \simeq g[\eta(n)]+\frac{\sigma_\xi^2(n)}{2}g''[\eta(n)] \tag{27}$$

Thus,

$$E[\lambda(n)] \simeq 1-\kappa\left\{g[\eta(n)]+\frac{\sigma_\xi^2(n)}{2}g''[\eta(n)]\right\} \tag{28}$$

Now, writing

$$\hat{e}_i^2(n)=E[\hat{e}_i^2(n)]+\varepsilon_i(n)=m_i(n)+\varepsilon_i(n), \quad i=1,2 \tag{29}$$

the mean $\eta(n)$ is approximated as

$$\eta(n)=E\left[\frac{m_1(n)+\varepsilon_1(n)}{m_2(n)+\varepsilon_2(n)}\right] \simeq \frac{m_1(n)}{m_2(n)}=\frac{E[\hat{e}_1^2(n)]}{E[\hat{e}_2^2(n)]} \tag{30}$$

where the fluctuations of $\hat{e}_i^2$, $i=1,2$ have been neglected. This is because these quantities are closely approximated by their mean values for sufficiently large K. The numerator and denominator of (30) are given by (22) for $i=1$ and 2, respectively.

Using the same reasoning as above, the variance $\sigma_\xi^2(n)$ is approximated as follows:

$$\begin{aligned}\sigma_\xi^2(n) &= E\left\{\left(\frac{\hat{e}_1^2(n)}{\hat{e}_2^2(n)}\right)^2\right\}-\eta^2(n)\\ &\simeq \frac{E\{[\hat{e}_1^2(n)]^2\}}{E\{[\hat{e}_2^2(n)]^2\}}-\left\{\frac{E[\hat{e}_1^2(n)]}{E[\hat{e}_2^2(n)]}\right\}^2\end{aligned} \tag{31}$$

Using (29), (31) can be written as

$$\sigma_\xi^2(n) \simeq \frac{m_2^2(n)E[\varepsilon_1^2(n)]-m_1^2(n)E[\varepsilon_2^2(n)]}{\{m_2^2(n)+E[\varepsilon_2^2(n)]\}m_2^2(n)} \tag{32}$$

Under the assumption that $e_1(n)$ and $e_2(n)$ are zero-mean, white over time, and conditionally Gaussian given $\boldsymbol{W}_1(n)$ and $\boldsymbol{W}_2(n)$,

$$\begin{aligned}E[\varepsilon_i^2(n)] &= \frac{2}{K^2}\sum_{m=n-K+1}^{n}E[e_i^2(m)]^2\\ &= \frac{2}{K^2}\sum_{m=n-K+1}^{n}[\sigma_o^2+\sigma_u^2\,\mathrm{MSD}_i(m)]^2, \quad i=1,2.\end{aligned} \tag{33}$$

Using (22) and (33) in (32) for $\sigma_\xi^2(n)$, (22) in (30) for $\eta(n)$, and finally (25), (30) and (32) in (28) yields the analytical model for $E[\lambda(n)]$. Evaluation of the final expression for each nonlinear function $g[\xi(n)]$ requires the evaluation of $g[\eta(n)]$ and $g''[\eta(n)]$.

3.3. Mean-square deviation

Using (4) and (5) with $y_i(n)=\boldsymbol{W}_i^T\boldsymbol{U}(n)$, $i=1,2$ and rearranging terms yields

$$\begin{aligned}e(n) &= e_o(n)+\{\lambda(n)[\boldsymbol{W}_o-\boldsymbol{W}_1(n)]\\ &\quad +[1-\lambda(n)][\boldsymbol{W}_o-\boldsymbol{W}_2(n)]\}^T\boldsymbol{U}(n)\end{aligned} \tag{34}$$

Neglecting the statistical dependence between $\lambda(n)$ and the adaptive weights,[4] squaring and averaging (34) yields an approximation for the MSD of the adaptive filter combination:

$$\begin{aligned}\mathrm{MSD}_c(n) &= E[e^2(n)]-\sigma_o^2\\ &\simeq \sigma_u^2\{E[\lambda^2(n)]\mathrm{MSD}_1(n)\\ &\quad +\{1-2E[\lambda(n)]+E[\lambda^2(n)]\}\mathrm{MSD}_2(n)\\ &\quad +2\{E[\lambda(n)]-E[\lambda^2(n)]\}\mathrm{MSD}_{21}(n)\}\end{aligned} \tag{35}$$

where

$$\begin{aligned}\mathrm{MSD}_{21}(n) &= E\{[\boldsymbol{W}_o-\boldsymbol{W}_2(n)]^T[\boldsymbol{W}_o-\boldsymbol{W}_1(n)]\}\\ &= \boldsymbol{W}_o^T\boldsymbol{W}_o-\boldsymbol{W}_o^TE[\boldsymbol{W}_1(n)]-\boldsymbol{W}_o^TE[\boldsymbol{W}_2(n)]+E[\boldsymbol{W}_2^T(n)\boldsymbol{W}_1(n)].\end{aligned} \tag{36}$$

Thus, an approximation for $E[\lambda^2(n)]$ is necessary. From (17),

$$E[\lambda^2(n)]=1-2\kappa\,E\{g(\xi(n))\}+\kappa^2E\{g^2[\xi(n)]\} \tag{37}$$

$E\{g[\xi(n)]\}$ has already been evaluated in Section 3.2 above. Again, using a second degree approximation for $E\{g^2[\xi(n)]\}$ and defining

$$q[\xi(n)]=g^2[\xi(n)] \tag{38}$$

yields

$$E\{q[\xi(n)]\} \simeq q[\eta(n)]+\frac{\sigma_\xi^2}{2}q''[\eta(n)] \tag{39}$$

which completes the model, except for the evaluations of $g''[\eta(n)]$ and $q''[\eta(n)]$ for the specific nonlinearity.

3.4. Models using different nonlinear functions

The complete analytical model for different nonlinear functions $g[\xi(n)]$ only requires the evaluation of $g''[\xi(n)]$ and $q''[\xi(n)]$. Table 1 shows these expressions for $g_1[\xi(n)]$ and $g_2[\xi(n)]$ in (18) and (19). Extension to other nonlinear functions is straightforward.

4. Simulation results

This section presents simulation results to verify the accuracy of the models derived for the mean behavior of $\lambda(n)$ (Eq. (28)) and for the $\mathrm{MSD}_c(n)$ (Eq. (35)).

Consider a system identification setup in which $d(n)$ is the system's output and the system's impulse response $\boldsymbol{W}_o=[w_{o_1},\ldots,w_{o_N}]^T$ is given by the raised-cosine function [20] whose k-th coefficient is

$$w_{o_k}=\frac{\sin[2\pi f_o(k-\Delta)]}{2\pi f_o(k-\Delta)}\frac{\cos[2\pi r f_o(k-\Delta)]}{1-4r f_o(k-\Delta)}, \quad k=1,\ldots,N \tag{40}$$

[4] This assumption is reasonable given the small fluctuations of $\lambda(n)$, as shown later on in the simulation examples.

In (40), N is the number of coefficients, Δ is the right shift delay relative to the even function case, r is the roll-off factor ($0 \le r \le 1$) and $f_o = 3\alpha/N$ where α is the expansion factor.

In all the following simulations, $N=32$, $K=100$, $\sigma_u^2 = 1$ and the value of κ was obtained from (25). The expected values were estimated from 50 Monte Carlo runs.

Table 1
Expressions of $g''[\xi(n)]$ and $q''[\xi(n)]$ for $g[\xi(n)] = g_1[\xi(n)]$ and $g[\xi(n)] = g_2[\xi(n)]$.

$g_1[\xi(n)]$	$g_1''[\xi(n)] = -\frac{1}{4}\{1-g_1^2[\xi(n)]\}g_1[\xi(n)]$ $q_1''[\xi(n)] = \frac{1}{2}\{1-g_1^2[\xi(n)]\}\{1-3g_1^2[\xi(n)]\}$
$g_2[\xi(n)]$	$g_2''[\xi(n)] = -\frac{4\xi(n)}{\sqrt{\pi}}e^{-\xi^2(n)}$ $q_2''[\xi(n)] = \frac{8}{\sqrt{\pi}}\left\{\frac{e^{-\xi^2(n)}}{\sqrt{\pi}} - \xi(n)e^{-\xi^2(n)}g_2[\xi(n)]\right\}e^{-\xi^2(n)}$

4.1. Example 1

The unknown response $\boldsymbol{W}_o$ was obtained from (40) with $\Delta = 10$, $r=0.2$ and $\alpha = 1.2$. The response is shown in Fig. 2. The nonlinearity $g_2[\xi(n)]$ was used in (17). Fig. 3(a) shows the mean behaviors of $\lambda_o(n)$ (Eq. (6)) and $\lambda(n)$ (Eq. (17)), and the theoretical prediction from (28) for $\sigma_o^2 = 10^{-4}$, $\mu_1 = 1/34$ and step-size ratio $\delta = 0.1$ ($\mu_2 = 0.0029$). Note that (17) and (25) result in a mean behavior quite close to that of $E[\lambda_o(n)]$. Also, the theoretical model very accurately predicts the behavior of $E[\lambda(n)]$. Fig. 3(b) shows that (1) the actual system performance is very close to optimal and (2) the theoretical model (35) is accurate.

4.2. Example 2

The unknown response has parameters $\Delta = 5$, $r=0$ and $\alpha = 3.8$ and is shown in Fig. 4. This unknown response is more peaky with a longer tail than in the first example. The

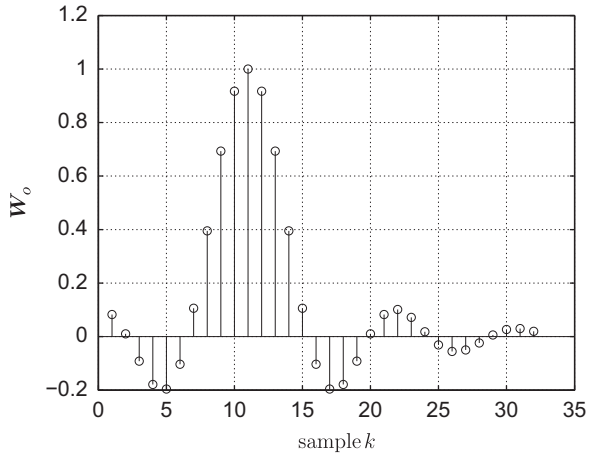

Fig. 2. Unknown impulse response $\boldsymbol{W}_o$ for Example 1.

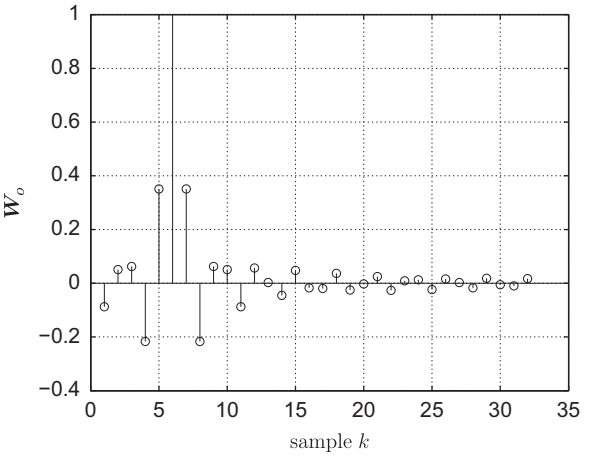

Fig. 4. Unknown impulse response $\boldsymbol{W}_o$ for Example 2.

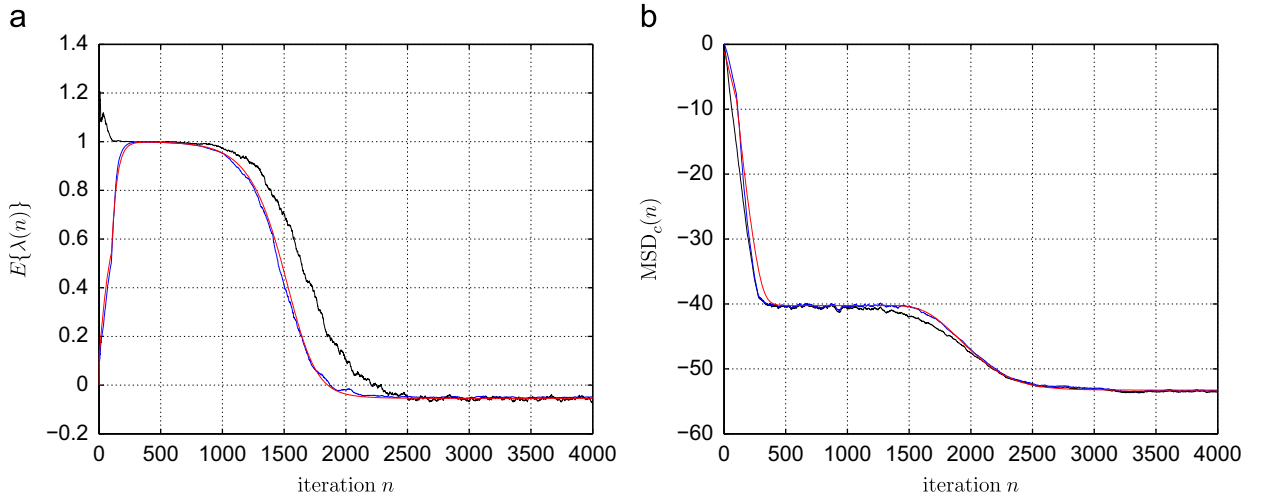

Fig. 3. Monte Carlo simulations averaged over 50 realizations for $g_2[\xi(n)]$ with $\sigma_o^2 = 10^{-4}$, $\delta = 0.1$ and $\mu_1 = 1/34$. (a) Behavior of $E\{\lambda(n)\}$ and $E\{\lambda_o(n)\}$. Black (top curve): $E\{\lambda_o(n)\}$ (optimum). Blue (superimposed by the red curve): $E\{\lambda(n)\}$ from (17). Red (bottom curve): theory using (28). (b) Behavior of $\mathrm{MSD}_c(n)$. Black: $\mathrm{MSD}_c(n)$ for $\lambda_o(n)$. Blue (bottom curve): $\mathrm{MSD}_c(n)$ from simulations. Red (top curve): theory using (35). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

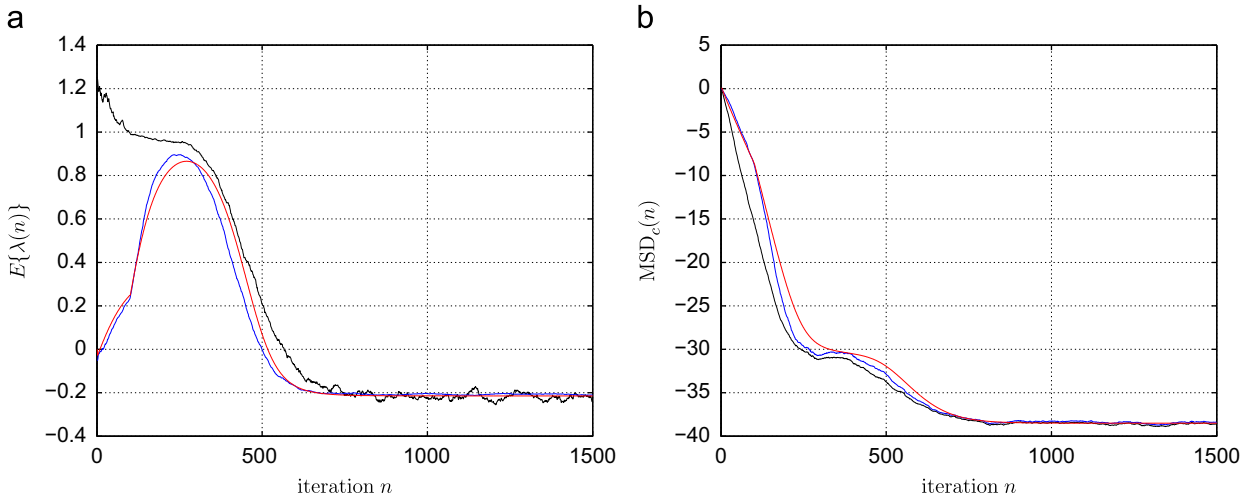

Fig. 5. Monte Carlo simulations averaged over 50 realizations for $g_2[\xi(n)]$ with $\sigma_o^2 = 10^{-3}$, $\delta = 0.3$ and $\mu_1 = 1/34$. (a) Behavior of $E\{\lambda(n)\}$ and $E\{\lambda_o(n)\}$. Black (top curve): $E\{\lambda_o(n)\}$ (optimum). Blue (bottom curve at $n=500$): $E\{\lambda(n)\}$ from (17). Red (middle curve at $n=500$): theory using (28). (b) Behavior of $\mathrm{MSD}_c(n)$. Black (bottom curve): $\mathrm{MSD}_c(n)$ for $\lambda_o(n)$. Blue (middle curve): $\mathrm{MSD}_c(n)$ from simulations. Red (top curve): theory using (35). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

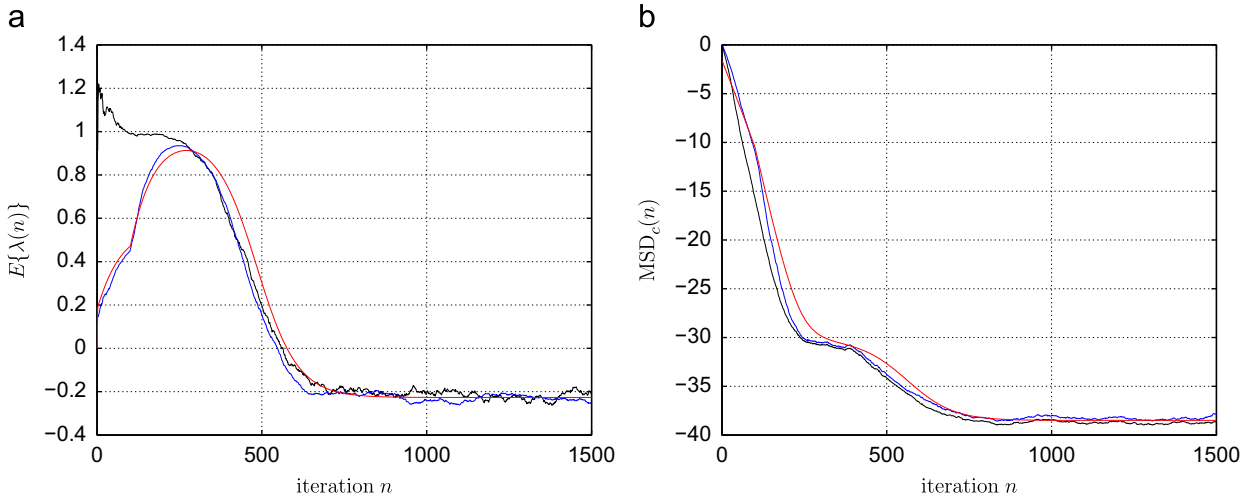

Fig. 6. Monte Carlo simulations averaged over 50 realizations for $g_1[\xi(n)]$ with $\sigma_o^2 = 10^{-3}$, $\delta = 0.3$ and $\mu_1 = 1/34$. (a) Behavior of $E\{\lambda(n)\}$ and $E\{\lambda_o(n)\}$. Black (top curve): $E\{\lambda_o(n)\}$ (optimum). Blue (bottom curve at $n=500$): $E\{\lambda(n)\}$ from (17). Red (middle curve at $n=500$): theory using (28). (b) Behavior of $\mathrm{MSD}_c(n)$. Black (bottom curve): $\mathrm{MSD}_c(n)$ for $\lambda_o(n)$. Blue (middle curve): $\mathrm{MSD}_c(n)$ from simulations. Red (top curve): theory using (35). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

nonlinearity $g_2[\xi(n)]$ was used in (17). The noise level was increased to 10^{-3}, $\mu_1 = 1/34$ was maintained and the step-size ratio was changed to $\delta = 0.3$ ($\mu_2 = 0.0088$). Fig. 5(a) and (b) shows the simulation results and the theoretical predictions for $E[\lambda(n)]$ and $\mathrm{MSD}_c(n)$ for this case. Again, very good agreements between the behaviors of $E[\lambda_o(n)]$ and $E[\lambda(n)]$ and excellent theoretical predictions of the actual system behavior are displayed. Note that a window of only 100 points ($K=100$) corresponds to a very fast response for changing $\lambda(n)$, a desirable property in practical applications.

4.3. Example 3

For this example, the parameters are the same as in Example 2 (Fig. 5), except that $g_1[\xi(n)]$ is used instead of $g_2[\xi(n)]$ for the nonlinearity in (17). The results are shown in Fig. 6. Comparison of Fig. 6 with Fig. 5 suggests that the performance is relatively robust to the exact form of the saturation nonlinearity g.

4.4. Example 4

For this example, the parameters are the same as in Example 1 (Fig. 3), except that the step size μ_1 was reduced to 1/64 and $\delta = 0.2$. The results are shown in Fig. 7. In this case, the transfer from filter 1 to filter 2 starts, on average, earlier than the ideal. This is what causes the bump in the MSD_c behavior, relative to the optimal. Convergence, however, is achieved after about 2500 iterations and matches the optimal performance. The analytical model

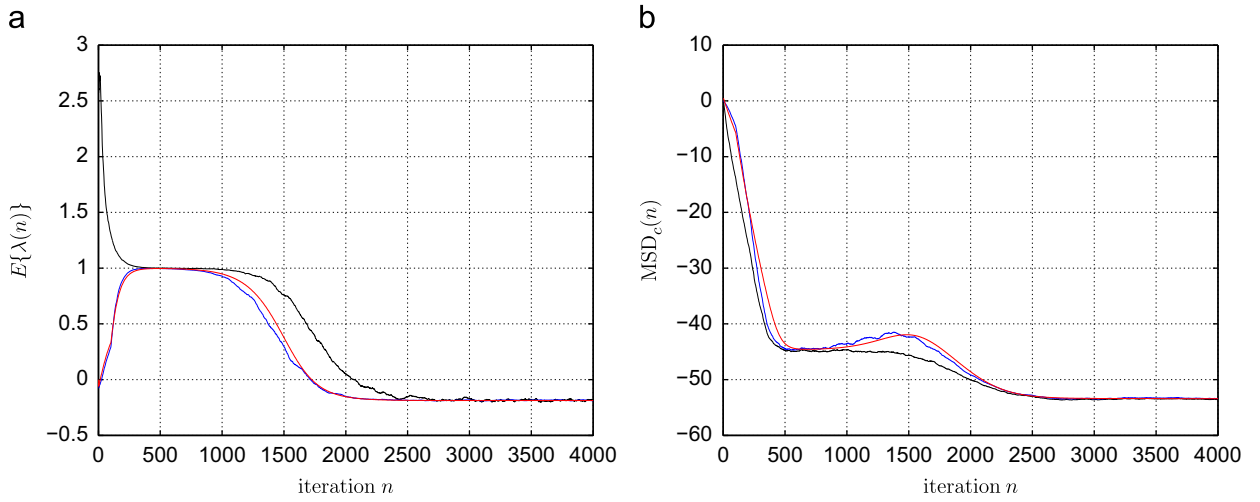

Fig. 7. Monte Carlo simulations averaged over 50 realizations for $g_2[\xi(n)]$ with $\sigma_o^2 = 10^{-4}$, $\delta = 0.2$ and $\mu_1 = 1/64$. (a) Behavior of $E\{\lambda(n)\}$ and $E\{\lambda_o(n)\}$. Black (top curve): $E\{\lambda_o(n)\}$ (optimum). Blue (superimposed by the red curve): $E\{\lambda(n)\}$ from (17). Red (bottom curve): theory using (28). (b) Behavior of $\mathrm{MSD}_c(n)$. Black: $\mathrm{MSD}_c(n)$ for $\lambda_o(n)$. Blue (bottom curve): $\mathrm{MSD}_c(n)$ from simulations. Red (top curve): theory using (35). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

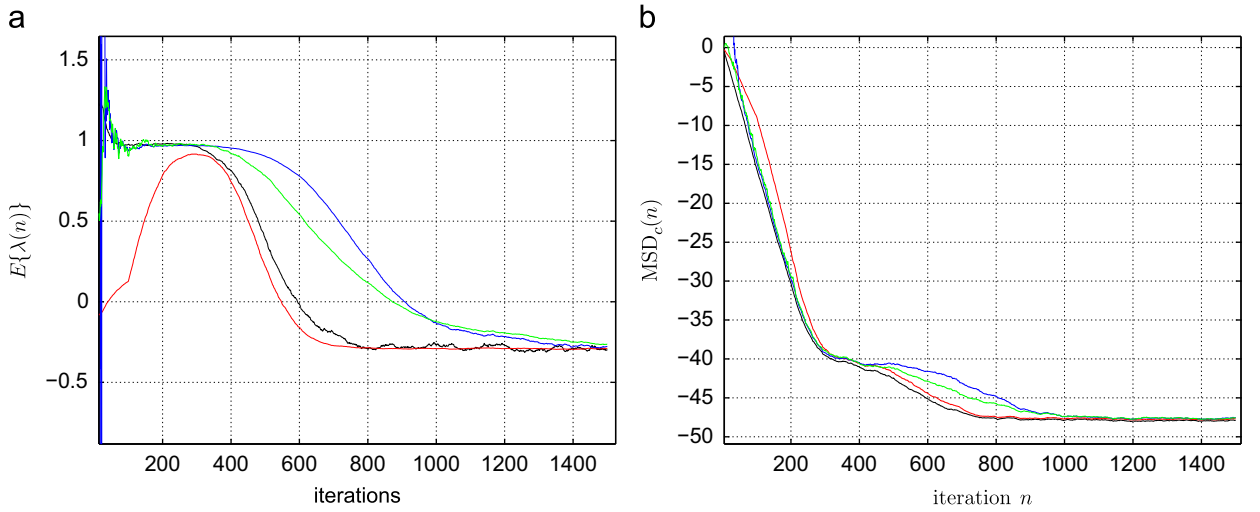

Fig. 8. Comparison the results obtained using the error power scheme with those obtained using η-PN-LMS and η-SR-LMS algorithms in [11] $g_2[\xi(n)]$. (a) Evolution of $E\{\lambda(n)\}$. Black: $E\{\lambda_o(n)\}$. Red: $E\{\lambda(n)\}$ for the error power scheme. Blue: $E\{\lambda(n)\}$ for the η-PN-LMS algorithm in [11]. Green: $E\{\lambda(n)\}$ for the η-SR-LMS algorithm in [11]. (b) Evolution of $\mathrm{MSD}_c(n)$. Black: $\mathrm{MSD}_c(n)$ for $\lambda(n) = \lambda_o(n)$. Red: $\mathrm{MSD}_c(n)$ for the error power scheme. Blue: $\mathrm{MSD}_c(n)$ for the η-PN-LMS algorithm in [11]. Green: $\mathrm{MSD}_c(n)$ for the η-PN-LMS algorithm in [11]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predicts very well the algorithm's behavior. The premature transfer from filter 1 to filter 2 seems to be a typical behavior of the error power transfer scheme as the faster filter step size reduces from the maximum speed value. Nevertheless, the best performance is obtained from the LMS affine combination when filter 1 is designed for maximum convergence speed and filter 2 is designed for the required steady-state performance. A practical design methodology will be illustrated in the following example.

4.5. Example 5

This example is presented in order to compare the performance of the error power scheme to the η-PN-LMS and η-SR-LMS algorithms described in [11]. The unknown response has parameters $\Delta = 15$, $r = 0$ and $\alpha = 2.8$ and the noise floor is $\sigma_o^2 = 10^{-4}$. The nonlinearity $g_2[\xi(n)]$ was used in (17). The steady-state MSE for the combined filters is

$$\xi_c(\infty) = E[e^2(\infty)] = \sigma_o^2(n) + \sigma_u^2 \mathrm{MSD}_c(\infty)$$

Step size $\mu_2 = 0.0106$ was chosen so that $\sigma_u^2\,\mathrm{MSD}_2(\infty) \approx \sigma_u^2\,\mathrm{MSD}_c(\infty) \approx 2 \times 10^{-5}$ has a negligible effect on $\xi_c(\infty)$. Smaller values for μ_2 slow overall convergence without significant reduction in $\xi_c(\infty)$.

Now, the fastest convergence step size for filter 1 would be $\mu_{1_o} = 1/N\sigma_u^2 = 0.0312$ for this case. Using $\mu_1 = 1/(N+2) = 0.0294$ (for stability reasons), yields $\delta = 0.3590$.

Good step-size parameters for the λ adjustment in [11] (obtained after some optimization by trial and error) are $\tilde{\mu}_{\eta} \approx 0.09$ and $\mu_{\eta s} \approx 0.75$ for this case.[5]

Fig. 8(a) and (b) shows the Monte Carlo simulation results for the new design parameters. Fig. 8(a) shows that $\lambda(n)$ evaluated using the error power scheme is much closer to $\lambda_o(n)$ in the transition region than $\lambda(n)$ for either of the two schemes in [11]. This difference is seen clearly in Fig. 8(b). Increasing the λ-adjustment step sizes to increase the speed in the transition region has led to larger steady-state fluctuations in $\lambda(n)$, resulting in larger steady-state MSD. It can also be verified that $\lambda(n)$ obtained using all three schemes deviates significantly from $\lambda_o(n) \approx 1$ during the initial phase (this property is also discussed in [11]). After the initial and before the transition phase, $\lambda(n)$ provided by either of the two schemes in [11] approaches $\lambda(n) = 1$ faster than the error power scheme.[6] The different behaviors, however, do not appear to have a significant impact on the behavior of $\mathrm{MSD}_c(n)$ before the transition phase.

5. Conclusions

This paper has studied the stochastic behavior of the error power scheme proposed in [9] for an affine combination of two LMS adaptive filters. A new design equation for κ improves the steady-state match between the output combination parameter $\lambda(n)$ and the optimum parameter $\lambda_o(n)$. Analytic models have been derived for the mean behavior of $\lambda(n)$ and for the combined adaptive weights mean-square deviation. Linear and quadratic model approximations have been used. The resulting model was shown to be simple to implement and accurate in predicting the affine combination behavior. Comparison with gradient type schemes [11] indicate a better performance of the studied scheme in the transition region. Possible future works include the analysis of the error power scheme performance for colored input signals.

References

[1] R.W. Harris, D.M. Chabries, F.A. Bishop, Variable step (VS) adaptive filter algorithm, IEEE Trans. Acoust. Speech Signal Process. 34 (1986) 309–316.

[2] R.H. Kwong, E.W. Johnston, A variable step size LMS algorithm, IEEE Trans. Signal Process. 40 (1992) 1633–1642.

[3] T. Aboulnasr, K. Mayyas, A robust variable step-size LMS-type algorithm: analysis and simulations, IEEE Trans. Signal Process. 45 (1997) 631–639.

[4] H.C. Shin, A.H. Sayed, Variable step-size NLMS and affine projection algorithms, IEEE Signal Process. Lett. 11 (2004) 132–135.

[5] M. Martínez-Ramón, J. Arenas-García, A. Navia-Vásquez, A.R. Figueiras-Vidal, An adaptive combination of adaptive filters for plant identification, in: Proceedings of the 14th International Conference on Digital Signal Processing, IEEE, Santorini, Greece, 2002, pp. 1195–1198.

[6] J. Arenas-Garcia, V. Gomez-Verdejo, A.R. Figueiras-Vidal, New algorithms for improved adaptive convex combination of LMS transversal filters, IEEE Trans. Instrum. Meas. 54 (2005) 2239–2249.

[7] J. Arenas-Garcia, A.R. Figueiras-Vidal, A.H. Sayed, Mean-square performance of a convex combination of two adaptive filters, IEEE Trans. Signal Process. 54 (2006) 1078–1090.

[8] V.H. Nascimento, M. Silva, R. Candido, J. Arenas-Garcia, A transient analysis for the convex combination of adaptive filters, in: Proceedings of the IEEE 15th Workshop on Statistical Signal Processing, IEEE, Cardiff, Wales, UK, 2009, pp. 53–56.

[9] N.J. Bershad, J.C.M. Bermudez, J.-Y. Tourneret, An affine combination of two LMS adaptive filters—transient mean-square analysis, IEEE Trans. Signal Process. 56 (2008) 1853–1864.

[10] R. Candido, M.T.M. Silva, V. Nascimento, Affine combinations of adaptive filters, in: Proceedings of the Asilomar Conference on Signals, Systems, and Computers, IEEE, Pacific Grove, CA, USA, 2008.

[11] R. Candido, M.T.M. Silva, V.H. Nascimento, Transient and steady-state analysis of the affine combination of two adaptive filters, IEEE Trans. Signal Process. 58 (2010) 4064–4078.

[12] A.T. Erdogan, S.S. Kozat, A.C. Singer, Comparison of convex combination and affine combination of adaptive filters, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Taiwan, 2009, pp. 3089–3092.

[13] S.S. Kozat, A.C. Singer, A.T. Erdogan, A.H. Sayed, Unbiased model combinations for adaptive filtering, IEEE Trans. Signal Process. 58 (2010) 4421–4427.

[14] S.S. Kozat, A.T. Erdogan, A.C. Singer, A.H. Sayed, Steady-state mse performance analysis of mixture approaches to adaptive filtering, IEEE Trans. Signal Process. 58 (2010) 4050–4063.

[15] N.J. Bershad, J.C.M. Bermudez, J.-Y. Tourneret, An affine combination of two LMS adaptive filters—statistical analysis of an error power ratio scheme, in: Proceedings of the Asilomar Conference on Signals, Systems, and Computers, IEEE, Pacific Grove, CA, USA, 2009.

[16] S. Haykin, Adaptive Filter Theory, fourth ed., Prentice-Hall, New Jersey, 2002.

[17] N.J. Bershad, A. Bist, Fast coupled adaptation for sparse impulse responses using a partial haar transform, IEEE Trans. Signal Process. 53 (2005) 966–976.

[18] A.H. Sayed, Adaptive Filtering, Wiley, IEEE Press, New York, 2008.

[19] A. Papoulis, Probability, Random Variables, and Stochastic Processes, third ed., McGraw-Hill, New York, 1991.

[20] L.W. Couch, III, Digital and Analog Communication Systems, second ed., MacMillan, New York, 1987.

[5] Note that the schemes in [11] require the optimization of two parameters, whereas the error power scheme requires only the choice of K. We have found that K should be significantly larger than N.

[6] The evaluation of $\lambda(n)$ is not recursive for the error power scheme. It is evaluated instantaneously from (17). It starts at zero and quickly jumps to 1.