Affine Combinations of Adaptive Filters

Renato Candido, Magno T. M. Silva, and Vítor H. Nascimento
Electronic Systems Engineering Department, Escola Politécnica, University of São Paulo
Av. Prof. Luciano Gualberto, trav. 3, nº 158, CEP 05508-900 – São Paulo, SP, Brazil
E-mails: {renatocan, magno, vitor}@lps.usp.br

***Abstract*—We extend the analysis presented in [1] for the affine combination of two least mean-square (LMS) filters to allow for colored inputs and nonstationary environments. Our theoretical model deals, in a unified way, with any combinations based on the following algorithms: LMS, normalized LMS (NLMS), and recursive-least squares (RLS). Through the analysis, we observe that the affine combination of two algorithms of the same family with close adaptation parameters (step-sizes or forgetting factors) provides a 3 dB gain in relation to its best component filter. We study this behavior in stationary and nonstationary environments. Good agreement between analytical and simulation results is always observed. Furthermore, a simple geometrical interpretation of the affine combination is investigated. A model for the transient and steady-state behavior of two possible algorithms for estimation of the mixing parameter is proposed. The model explains situations in which adaptive combination algorithms may achieve good performance.**

***Index Terms*—Adaptive filters, affine combination, steady-state analysis, transient analysis, LMS algorithm.**

I. Introduction

Recently, an affine combination of two least mean-square (LMS) adaptive filters was proposed and its transient performance analyzed [1]. This method combines linearly the outputs of two LMS filters operating in parallel with different step-sizes. The purpose of the combination is to obtain an adaptive filter with fast convergence and reduced steady-state excess mean-square error (EMSE). Since the mixing parameter is not restricted to the interval [0, 1], this method can be interpreted as a generalization of the convex combination of two LMS filters of [2], [3].

In this paper, we extend the results of [1] by providing a unified analysis, which is valid for colored inputs, nonstationary environments, and combinations based on LMS, NLMS, and RLS algorithms. To explain the behavior of the affine combination of two algorithms, we present a simple geometrical interpretation. Furthermore, we also explain why fast-adaptation of the mixing parameter in general leads to a quite large variance around the optimum value. Then, we find a model for the transient and steady-state behavior of two possible algorithms for estimation of the mixing parameter. In order to simplify the arguments, we assume that all quantities are real.

II. Problem Formulation

A combination of two adaptive filters is depicted in Figure 1. In this scheme, the output of the overall filter is given by

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

where $\eta(n)$ is the mixing parameter, $y_i(n)$, $i = 1, 2$ are the outputs of the transversal filters, i.e., $y_i(n) = \mathbf{u}^T(n)\mathbf{w}_i(n-1)$, $\mathbf{u}(n) \in \mathbb{R}^M$ is the common regressor vector, and $\mathbf{w}_i(n-1) \in \mathbb{R}^M$ are the weight vectors of each length-M component filter.

We focus on the affine combination of two adaptive algorithms of the following general class

$$\mathbf{w}_i(n) = \mathbf{w}_i(n-1) + \rho_i(n)\,\mathbf{M}_i(n)\mathbf{u}(n)e_i(n), \;\; i = 1, 2, \tag{2}$$

This work is partly supported by CNPq under grants No. 136050/2008-5 and No. 303361/2004-2 and by FAPESP under grants No. 2008/00773-1 and No. 2008/04828-5.

where $\rho_i(n)$ is a step-size, $\mathbf{M}_i(n)$ is a symmetric non-singular matrix, $e_i(n) = d(n) - y_i(n)$ is the estimation error, and $d(n)$ is the desired response. The LMS, NLMS, and RLS algorithms employ the step-sizes $\rho_i(n)$ and the matrices $\mathbf{M}_i(n)$ as in Table I. In this table, μ_i, $\tilde{\mu}_i$ and ϵ are positive constants, $\|\cdot\|$ is the Euclidian norm, $\mathbf{I}$ is the $M \times M$ identity matrix, and $0 \ll \lambda_i < 1$ is a forgetting factor. For RLS, $\mathbf{M}_i(n) = \widehat{\mathbf{R}}_i^{-1}(n)$ is obtained via the matrix inversion lemma [4, Eq.(2.6.4)] applied to $\widehat{\mathbf{R}}_i(n)$, which is an estimate (with forgetting factor λ_i) of the autocorrelation matrix of the input signal, i.e., $\mathbf{R} \triangleq \mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}$, where $\mathrm{E}\{\cdot\}$ is the expectation operator.

We assume that $d(n)$ and $\mathbf{u}(n)$ are related via a linear regression model, that is, $d(n) = \mathbf{u}^T(n)\mathbf{w}_\mathrm{o}(n-1) + v(n)$, where $\mathbf{w}_\mathrm{o}(n-1)$ is the time-variant optimal solution and $v(n)$ is an i.i.d. (independent and identically distributed) and zero mean random process with variance $\sigma_v^2 = \mathrm{E}\{v^2(n)\}$, which plays the role of a disturbance independent of $\mathbf{u}(n)$ [4, Sec. 6.2.1]. Furthermore, the sequences $\{u(n)\}$ and $\{v(n)\}$ are assumed stationary.

In the affine combination, the mixing parameter $\eta(n)$ is not restricted to the interval [0, 1] and can be adapted via

$$\eta(n+1) = \eta(n) + \mu_\eta e(n)[y_1(n) - y_2(n)], \tag{3}$$

where μ_η is a step-size, and $e(n) = d(n) - y(n)$ is the estimation error of the overall filter. The recursion (3) was obtained in [1], using a stochastic gradient search to minimize the instantaneous mean-square error (MSE) cost function. In [1], $\eta(n)$ was constrained to be less than or equal to 1 for all n, to ensure stability of (3). In this paper, we applied this constraint when using (3). The constraint was not necessary when the normalized version of (3) was used (see Sec. V).

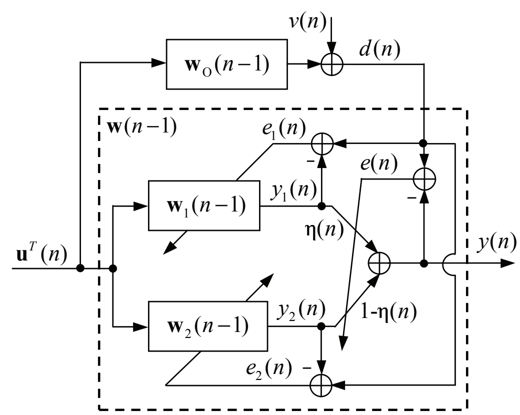

Fig. 1. Affine combination of two transversal adaptive filters.

III. Steady-State performance of adaptive filters

We assume that in a nonstationary environment, the variation in the optimal solution $\mathbf{w}_\mathrm{o}$ follows a random-walk model [4, p. 359], that is, $\mathbf{w}_\mathrm{o}(n) = \mathbf{w}_\mathrm{o}(n-1) + \mathbf{q}(n)$. In this model, $\mathbf{q}(n)$ is an i.i.d. vector

TABLE I
PARAMETERS OF THE CONSIDERED ALGORITHMS.

Alg.	$\rho_i(n)$	$\mathbf{M}_i^{-1}(n)$
LMS	μ_i	$\mathbf{I}$
NLMS	$\tilde{\mu}_i/\left(\epsilon+\|\mathbf{u}(n)\|^2\right)$	
RLS	1	$\widehat{\mathbf{R}}_i(n)=\sum_{l=1}^{n}\lambda_i^{n-l}\mathbf{u}(l)\mathbf{u}^T(l)$

TABLE II
ANALYTICAL EXPRESSIONS FOR CROSS-EMSE OF THE CONSIDERED COMBINATIONS.

Combination	ζ_{12}
μ_1-LMS and μ_2-LMS	$\dfrac{\mu_1\mu_2\sigma_v^2\mathrm{Tr}(\mathbf{R})+\mathrm{Tr}(\mathbf{Q})}{\mu_1+\mu_2-\mu_1\mu_2\mathrm{Tr}(\mathbf{R})}$
$\tilde{\mu}_1$-NLMS and $\tilde{\mu}_2$-NLMS	$\dfrac{\mathrm{Tr}(\mathbf{R})\left[\tilde{\mu}_1\tilde{\mu}_2\sigma_v^2\alpha_u+\mathrm{Tr}(\mathbf{Q})\right]}{\tilde{\mu}_1+\tilde{\mu}_2-\tilde{\mu}_1\tilde{\mu}_2}$
λ_1-RLS and λ_2-RLS	$\dfrac{\tilde{\lambda}_1\tilde{\lambda}_2\left[1+\dfrac{\tilde{\lambda}_1\tilde{\lambda}_2}{1-\lambda_1\lambda_2}\gamma\right]M\sigma_v^2+\mathrm{Tr}(\mathbf{QR})}{\tilde{\lambda}_1+\tilde{\lambda}_2-\tilde{\lambda}_1\tilde{\lambda}_2}$
λ_1-RLS and μ_2-LMS	$\mu_2\tilde{\lambda}_1\,\sigma_v^2\,\mathrm{Tr}(\mathbf{\Sigma})+\mathrm{Tr}(\mathbf{Q\Sigma})$

with positive-definite autocorrelation matrix $\mathbf{Q}=\mathrm{E}\{\mathbf{q}(n)\mathbf{q}^T(n)\}$, independent of the initial conditions $\{\mathbf{w}_\mathrm{o}(-1),\mathbf{w}(-1),\eta(-1)\}$ and of $\{\mathbf{u}(l),d(l)\}$ for all l [4, Sec. 7.4].

One measure of the performance of each component filter is given by the excess MSE (EMSE), defined as

$$\zeta_i(n)\triangleq\mathrm{E}\{e_{a,i}^2(n)\},\qquad\qquad \zeta_i\triangleq\lim_{n\to\infty}\zeta_i(n),$$

where ζ_i is the steady-state value of $\zeta_i(n)$, $e_{a,i}(n)=\mathbf{u}^T(n)\widetilde{\mathbf{w}}_i(n-1)$, and $\widetilde{\mathbf{w}}_i(n-1)=\mathbf{w}_\mathrm{o}(n-1)-\mathbf{w}_i(n-1)$. On the other hand, the overall filter performance can be measured by

$$\zeta(n)\triangleq\mathrm{E}\{e_a^2(n)\},\qquad\qquad \zeta\triangleq\lim_{n\to\infty}\zeta(n),$$

where

$$e_a(n)=\eta(n)e_{a,1}(n)+[1-\eta(n)]e_{a,2}(n). \tag{4}$$

To obtain analytical expressions for ζ, we need expressions for ζ_i, $i=1,2$ and for the cross-EMSE [3], [5]

$$\zeta_{12}(n)\triangleq\mathrm{E}\{e_{a,1}(n)e_{a,2}(n)\},\qquad \zeta_{12}\triangleq\lim_{n\to\infty}\zeta_{12}(n).$$

There have been several works in the literature on the tracking performance of adaptive algorithms (see, e.g., [4], [6]–[10] and their references). Analytical expressions for the EMSE of LMS [4], [9], NLMS [4], [7], and RLS [6] algorithms can be obtained from the first three lines in Table II, using $\mu_2=\mu_1$, $\tilde{\mu}_2=\tilde{\mu_1}$ or $\tilde{\lambda}_2=\tilde{\lambda}_1\triangleq(1-\lambda_1)$, where $\mathrm{Tr}(\mathbf{A})$ stands for the trace of matrix $\mathbf{A}$, $\alpha_u\triangleq\mathrm{E}\left\{\|\mathbf{u}(n)\|^{-2}\right\}$, $\gamma=\mathrm{var}\{u^2(n)\}/(\mathrm{var}\{u(n)\})^2$, and $\mathrm{var}\{\cdot\}$ is the variance. For gaussian inputs, $\gamma=2$ and α_u can be approximated by $1/[\mathrm{var}\{u(n)\}(M-2)]$ [11].

The cross-EMSE for the combination of two LMS filters was estimated in [3] using energy conservation arguments. Using the traditional analysis method[1], analytical expressions for ζ_{12} for the combinations of two RLS filters and of one RLS with one LMS were obtained in [5]. For the combination of two RLS filters, another expression for ζ_{12} can be obtained using similar assumptions to those of [6]. Since the resulting expression is more accurate than that of [5], mainly for smaller forgetting factors, we use it here. Analytical expressions for ζ_{12} considering the combination of two NLMS filters are given, for white regressors, in [12]. We give here a straightforward extension for correlated inputs. All these results are summarized in Table II, where $\mathbf{\Sigma}\triangleq[\tilde{\lambda}_1\mathbf{I}+\mu_2\mathbf{R}]^{-1}\mathbf{R}$.

IV. A STEADY-STATE ANALYSIS OF AFFINE COMBINATIONS

To obtain an analytical expression for the optimum mixing parameter $\eta_\mathrm{o}(n)$ at the steady-state[2], we differentiate the mean-square error cost function $\mathrm{E}\{e^2(n)\}$ with respect to $\eta(n)$ and set the derivative equal to zero, i.e.,

$$\mathrm{E}\{e(n)\,[e_1(n)-e_2(n)]\}=0. \tag{5}$$

Using the linear regression model for $d(n)$, the estimation errors $e_i(n)$, $i=1,2$ are related to the *a priori* errors $e_{a,i}(n)$ via

$$e_i(n)=e_{a,i}(n)+v(n). \tag{6}$$

Then, using (1) and (6), (5) can be rewritten as

$$\mathrm{E}\{\eta_\mathrm{o}(n)[e_{a,1}(n)-e_{a,2}(n)]^2\}=\mathrm{E}\{e_{a,2}(n)[e_{a,2}(n)-e_{a,1}(n)]\}. \tag{7}$$

To proceed, we assume that

A1. $\eta_\mathrm{o}(n)$ is independent of $e_{a,i}(n),\ i=1,2$ at the steady-state. This assumption requires the optimum mixing parameter to be independent of the *a priori* errors when $n\to\infty$.

Thus, using A1 and taking the limit for $n\to\infty$ of both sides of (7), we arrive at

$$\bar{\eta}_\mathrm{o}(\infty)\triangleq\lim_{n\to\infty}\mathrm{E}\{\eta_\mathrm{o}(n)\}\approx\frac{\Delta\zeta_2}{\Delta\zeta_1+\Delta\zeta_2}, \tag{8}$$

where $\Delta\zeta_i=\zeta_i-\zeta_{12}$, $i=1,2$. The accuracy of (8) depends on the the accuracy of the analytical expressions of ζ_i, $i=1,2$ and ζ_{12}. A similar expression was also obtained in [3, Eq.(29)] for the convex combination of two LMS filters. The difference is that in the convex combination, $\eta(n)$ and consequently $\bar{\eta}_\mathrm{o}(\infty)$ are restricted to the interval [0, 1]. The expressions of Table II were obtained without the assumption of white inputs. Thus, (8) is an extension of [1, Eq. (26)] since it allows for colored inputs, nonstationary environments, and holds for combinations of algorithms of the form (2).

Now we obtain an analytical expression for the steady-state EMSE of an affine combination. By squaring both sides of (4) with $\eta(n)=\eta_\mathrm{o}(n)$, taking expectations, and using A1, we arrive at

$$\begin{aligned}\mathrm{E}\{e_a^2(n)\}=&\mathrm{E}\{\eta_\mathrm{o}^2(n)\}\mathrm{E}\{e_{a,1}^2(n)\}+\mathrm{E}\{[1-\eta_\mathrm{o}(n)]^2\}\mathrm{E}\{e_{a,2}^2(n)\}\\&+2\mathrm{E}\{\eta_\mathrm{o}(n)[1-\eta_\mathrm{o}(n)]\}\mathrm{E}\{e_{a,1}(n)e_{a,2}(n)\}.\end{aligned} \tag{9}$$

To proceed, we assume for now that

A2. the variance of $\eta_\mathrm{o}(n)$ is sufficiently small at the steady-state such that $\lim_{n\to\infty}\mathrm{E}\{\eta_\mathrm{o}^2(n)\}\approx\bar{\eta}_\mathrm{o}^{\,2}(\infty)$.

Using A2 and taking the limit of both sides of (9) for $n\to\infty$, we arrive at

$$\zeta\approx\zeta_{12}+\frac{\Delta\zeta_1\Delta\zeta_2}{\Delta\zeta_1+\Delta\zeta_2}. \tag{10}$$

This expression was obtained in [3, Eq. (33)] for the convex combination of two LMS filters, but also holds for different affine combinations of algorithms of the form (2).

[1] In the traditional method, one computes a recursion for the autocorrelation matrix of the weight-error vector of a filter.

[2] Note that we use the subscript "o" in $\eta_\mathrm{o}(n)$ to denote the optimum mixing parameter. It is optimum in the mean-square error sense.

A. Stationary environments

In an stationary environment ($\mathbf{Q} = \mathbf{0}$), the expressions (8) and (10) for the combinations of two LMS or two NLMS filters are shown in Table III, where $\delta \triangleq \mu_2/\mu_1$ with $0 < \delta < 1$, and $\tilde{\delta} \triangleq \tilde{\mu}_2/\tilde{\mu}_1$ with $0 < \tilde{\delta} < 1$. The expressions of Table III show two interesting properties:

i) $\bar{\eta}_o(\infty)$ for both combinations is negative, since to ensure the stability of the μ_1-LMS and $\tilde{\mu}_1$-NLMS, the step-sizes are chosen respectively in the following ranges $0 < \mu_1 < 2/\mathrm{Tr}(\mathbf{R})$ and $0 < \tilde{\mu}_1 < 2$;

ii) $\delta \approx 1$ (resp., $\tilde{\delta} \approx 1$) yields $\zeta \approx \zeta_2/2$ for the combination of two LMS filters (resp., NLMS). Since $\zeta_2 < \zeta_1$ for both combinations, the affine combination provides a 3dB gain in relation to the best component filter. In this case, $\eta_o(\infty) \to -\infty$.

Property i) was observed in [1] for the combination of two LMS filters, assuming gaussian, white inputs, and the LMS step-size for maximum convergence speed. Note that, if we also consider the LMS step-size for maximum speed, i.e., $\mu_1 = 1/\mathrm{Tr}(\mathbf{R})$ in the expression of Table III, the steady-state optimum mixing parameter for the combination of two LMS filters will reduce to $\bar{\eta}_o(\infty) = \delta/[2(\delta - 1)]$, which coincides to the result of [1, Eq.(26)]. Although we exemplify these properties for the combinations of two LMS or two NLMS algorithms, they also hold for all the combinations considered here. For the combinations of two RLS or one RLS with one LMS, (8) and (10) do not reduce to simple expressions as those of Table III even for stationary environments, and are not presented here for lack of space.

TABLE III
ANALYTICAL EXPRESSIONS FOR $\bar{\eta}_o(\infty)$ AND ζ IN THE STATIONARY CASE.

Combination	$\bar{\eta}_o(\infty)$	ζ
μ_1-LMS and μ_2-LMS	$\dfrac{\delta[2 - \mu_1 \mathrm{Tr}(\mathbf{R})]}{2(\delta - 1)}$	$\dfrac{1}{2}\left[\dfrac{\mu_2 \sigma_v^2 \mathrm{Tr}(\mathbf{R})}{\delta + 1 - \mu_2 \mathrm{Tr}(\mathbf{R})}\right]$
$\tilde{\mu}_1$-NLMS and $\tilde{\mu}_2$-NLMS	$\dfrac{\tilde{\delta}[2 - \tilde{\mu}_1]}{2(\tilde{\delta} - 1)}$	$\dfrac{1}{2}\left[\dfrac{\mathrm{Tr}(\mathbf{R})\tilde{\mu}_2 \sigma_v^2 \alpha_u}{\tilde{\delta} + 1 - \tilde{\mu}_2}\right]$

In order to explain the behavior of the affine combination when the adaptation parameters are close (e.g., $\mu_1 \approx \mu_2$), the overall steady-state error is written as

$$e(n) = \underbrace{e_{a,2}(n) + v(n)}_{d(n)} + \eta(n) \underbrace{[\mathbf{w}_2(n) - \mathbf{w}_1(n)]^T \mathbf{u}(n)}_{-x(n)}. \qquad (11)$$

From the point of view of the computation of $\eta(n)$, $d(n)$ represents the signal which has to be estimated, and $x(n)$ plays the role of input signal. Assuming that $\mathbf{w}_i$, $i = 1, 2$ vary slowly compared to η, (11) has a simple geometric interpretation as shown in Fig. 2. The affine combination seeks the best weight vector in the line $\mathbf{w}_2 + \eta(\mathbf{w}_1 - \mathbf{w}_2)$. In Fig. 2-(a), the best linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$ is $\mathbf{w}$. In the case of close adaptation parameters (e.g., $\mu_1 \approx \mu_2$ or $\lambda_1 \approx \lambda_2$), we also have $\mathbf{w}_1 \approx \mathbf{w}_2$ (Fig. 2-(b)), and η has to assume a large value to take the combined vector close to $\mathbf{w}$, since the input signal $x(n)$ depends on the difference between $\mathbf{w}_1$ and $\mathbf{w}_2$. Thus, if $(\mathbf{w}_1 - \mathbf{w}_2) \to 0$, $|\eta| \to \infty$.

B. Nonstationary environments

In a nonstationary environment, the largest EMSE reduction of the affine combination in relation to its components occurs when $\zeta_1 \approx \zeta_2$. This can happen in two situations: (i) when $\mathrm{Tr}(\mathbf{Q}) = q_{12}$

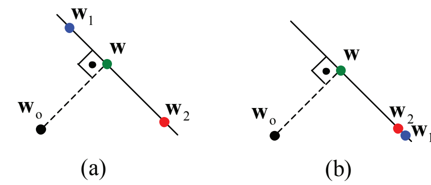

Fig. 2. Geometric interpretation of the affine combination.

or when (ii) the component filters have close adaptation parameters. In Table IV, we show the analytical expressions for q_{12} and ζ for the combinations of two LMSs or two NLMSs[3]. From these expressions, we can observe that the EMSE reduction in both cases is limited by 3 dB. A reduction close to 3 dB will occur when $\delta \to 0$ (or $\tilde{\delta} \to 0$) in case (i) or when the environment tends to be stationary ($\mathrm{Tr}(\mathbf{Q}) \approx 0$) in case (ii).

TABLE IV
ANALYTICAL EXPRESSIONS FOR q_{12} AND ζ FOR THE CASES (i) AND (ii) IN A NONSTATIONARY ENVIRONMENT.

Combination	(i)		(ii)
	q_{12}	ζ	ζ
μ_1-LMS and μ_2-LMS	$\mu_1\mu_2\sigma_v^2 \times \mathrm{Tr}(\mathbf{R})$	$\zeta_2/2 + \dfrac{2\delta\zeta_2}{(1+\delta)^2}$	$\zeta_2/2 + \dfrac{\sigma_v^2 \mathrm{Tr}(\mathbf{R})\mathrm{Tr}(\mathbf{Q})}{2\zeta_2}$
$\tilde{\mu}_1$-NLMS and $\tilde{\mu}_2$-NLMS	$\tilde{\mu}_1\tilde{\mu}_2\sigma_v^2 \times \alpha_u$	$\zeta_2/2 + \dfrac{2\tilde{\delta}\zeta_2}{(1+\tilde{\delta})^2}$	$\zeta_2/2 + \dfrac{\sigma_v^2 [\mathrm{Tr}(\mathbf{R})]^2 \mathrm{Tr}(\mathbf{Q})\alpha_u}{2\zeta_2}$

V. TRANSIENT ANALYSIS

At each instant, the combination parameter η is adapted based on the projections y_1 and y_2 of $\mathbf{w}_1$ and $\mathbf{w}_2$ in the direction of the regressor $\mathbf{u}$. If one tries to adapt η quickly, e.g., using the normalized LMS algorithm instead of (3), a problem arises when $\mathbf{u}$ is close to orthogonal to $(\mathbf{w}_1 - \mathbf{w}_2)$, as shown in Fig. 3. We show in the figure the situation for two possible values of the optimum solution, $\mathbf{w}_o$ and $\mathbf{w}_o'$. Note that for both values of the optimum solution, the best value of η is 1/2. However, looking at the projections on $\mathbf{u}$, in one case one would choose $\eta \approx 0$ and for the other case, $\eta \approx -\rho$, where ρ is a positive number. This example explains why fast-adaptation of the combination parameter in general leads to a quite large variance around η_o.

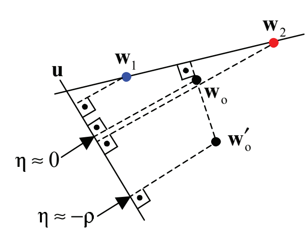

Fig. 3. Updating of $\eta(n)$ when the regressor $\mathbf{u}$ is close to orthogonal to $(\mathbf{w}_1 - \mathbf{w}_2)$.

We now find a model for the transient behavior of the combination. Assuming that $\eta(n)$ is independent of the *a priori* errors, we can show

[3]Again, expressions for combinations which involve the RLS algorithm are not shown due to space reasons.

that (10) still holds, and $\mathrm{E}\{\eta(n)\} = \bar{\eta}_o(n)$ and

$$\mathrm{E}\{e^2(n)\} = \sigma_o^2(n) + \sigma_\eta^2(n)[\zeta_1(n) + \zeta_2(n) - 2\zeta_{12}(n)], \quad (12)$$

where we have defined

$$\sigma_o^2(n) = \mathrm{E}\{e^2(n)|_{\eta(n)=\bar{\eta}_o(n)}\} = \sigma_v^2 + \zeta(n), \text{ and} \quad (13)$$
$$\sigma_\eta^2(n) = \mathrm{E}\{\eta^2(n)\} - \bar{\eta}_o^2(n). \quad (14)$$

To simplify the notation, we sometimes omit the time index n in the following discussion.

Note that the largest MSE reduction will occur when $\mathrm{E}\{e^2(n)\} \to \sigma_o^2(n)$. This happens, for example, when the adaptation parameters are close (e.g., $\mu_1 \approx \mu_2$ or $\lambda_1 \approx \lambda_2$), since in this case $\zeta_1 \approx \zeta_{12} \approx \zeta_2$, and the effect of a possibly large σ_η^2 is reduced. On the other hand, if $\zeta_1 \gg \zeta_{12} \gg \zeta_2$ (or vice-versa), the second term of the r.h.s. of (12) will be approximately proportional to the largest of ζ_1, ζ_2, and σ_η^2 will have to be smaller than $\min\{\zeta_1, \zeta_2\}/\max\{\zeta_1, \zeta_2\}$ to make the combination competitive with the best filter.

A recursion for σ_η^2 can be found by subtracting η_o from both sides of (3), and squaring the result. In the following, we assume that $\mathbf{w}_o$ is constant. Defining $\delta\eta(n) = \bar{\eta}_o(n) - \eta(n)$, we obtain

$$\begin{aligned}\delta\eta(n+1) &= \left[1 - \mu_\eta(e_{a,2}(n) - e_{a,1}(n))^2\right]\delta\eta(n) \\ &+ \mu_\eta\bar{\eta}_o(n)(e_{a,2}(n) - e_{a,1}(n))^2 - \mu_\eta e_{a,2}(n)(e_{a,2}(n) - e_{a,1}(n)) \\ &- \mu_\eta v(n)(e_{a,2}(n) - e_{a,1}(n)). \quad (15)\end{aligned}$$

Taking the expected value of (15), it can be shown that $\mathrm{E}\{\delta\eta(n)\} \to 0$. On the other hand, squaring (15) and taking expected values we obtain, assuming that $e_{a,1}(n)$ and $e_{a,2}(n)$ are Gaussian,

$$\begin{aligned}\sigma_\eta^2(n+1) = &\Big[1 - 2\mu_\eta\left(\Delta\zeta_1(n) + \Delta\zeta_2(n)\right) \quad (16) \\ &+ 3\mu_\eta^2\left(\Delta\zeta_1(n) + \Delta\zeta_2(n)\right)^2\Big]\sigma_\eta^2(n) + \mu_\eta^2\sigma_v^2\left(\Delta\zeta_1(n) + \Delta\zeta_2(n)\right) \\ &+ \mu_\eta^2\left[3\zeta_{12}(n)\left(\Delta\zeta_1(n) + \Delta\zeta_2(n)\right) - 2\left(\zeta_1(n)\zeta_2(n) - \zeta_{12}^2(n)\right)\right].\end{aligned}$$

For stability, we need

$$\mu_\eta < \frac{2}{3[\Delta\zeta_1(n) + \Delta\zeta_2(n)]}, \quad (17)$$

and the steady-state variance is

$$\lim_{n\to\infty}\sigma_\eta^2 = \mu_\eta\frac{3\zeta_{12}\left(\Delta\zeta_1 + \Delta\zeta_2\right) - 2\left(\zeta_1\zeta_2 - \zeta_{12}^2\right) + \sigma_v^2\left(\Delta\zeta_1 + \Delta\zeta_2\right)}{2\left(\Delta\zeta_1 + \Delta\zeta_2\right) - 3\mu_\eta\left(\Delta\zeta_1 + \Delta\zeta_2\right)^2}.$$

The adaptation law (3) is usually not fast enough to follow the necessary quick variations on η, and at the same time avoid a large excess mean-square error. As Fig. 3 shows, using an instantaneous normalization, i.e., replacing the step-size by $\mu_\eta(n) = \tilde{\mu}_\eta/[e_{a,2}(n) - e_{a,1}(n)]^2$, will lead to a very large σ_η^2, or even divergence (see [13]). On the other hand, some form of normalization is necessary, otherwise (3) will either be too slow when both component filters have converged (and $e_{a,2}(n) - e_{a,1}(n)$ is small), or will converge too fast (and diverge) when $e_{a,2}(n) - e_{a,1}(n)$ is large (e.g., when the fast filter has already converged, but the slow filter has still a large misadjustment). One possible solution is to normalize the filter using an estimate of $\mathrm{E}\left\{[e_{a,2}(n) - e_{a,1}(n)]^2\right\}$, as in [14].

Another possibility is to employ a partial instantaneous normalization, using $\mu_\eta(n) = \tilde{\mu}_\eta/|y_1(n) - y_2(n)|$ as step-size (note that $y_1(n) - y_2(n) = e_{a,2}(n) - e_{a,1}(n)$). With this choice, the update rule (3) reduces to

$$\eta(n+1) = \eta(n) + \tilde{\mu}_\eta e(n)\,\mathrm{sign}\,|y_1(n) - y_2(n)|. \quad (18)$$

It can be shown that this recursion also leads to an unbiased estimate of the optimum η_o, with variance

$$\begin{aligned}\sigma_\eta^2(n+1) = &\Big[1 - 2\tilde{\mu}_\eta\sqrt{2/\pi}\sqrt{\Delta\zeta_1 + \Delta\zeta_2} + \tilde{\mu}_\eta^2\left(\Delta\zeta_1 + \Delta\zeta_2\right)\Big] \\ &\times\sigma_\eta^2(n) + \tilde{\mu}_\eta^2\frac{\zeta_1\zeta_2 - \zeta_{12}^2}{\Delta\zeta_1 + \Delta\zeta_2} + \tilde{\mu}_\eta^2\sigma_v^2. \quad (19)\end{aligned}$$

For large step-sizes, (18) leads to smaller σ_η^2 than (3). The situation reverses for small step-sizes. Through simulations, we noticed that recursion (18) is less sensitive to variations in the input power and the value of the step-size.

In order to further improve the convergence speed of the algorithms, we estimated

$$p(n+1) = \lambda_p p(n) + (1 - \lambda_p)[y_1(n) - y_2(n)]^2,$$

where $0 \ll \lambda_p < 1$ is a forgetting factor, and used as step-sizes $\bar{\mu}_\eta = \tilde{\mu}_\eta/(\varepsilon + p(n))$ for (3), where $\varepsilon > 0$ is a regularization constant, and $\bar{\bar{\mu}}_\eta = \tilde{\mu}_\eta/(\varepsilon + \sqrt{p(n)})$ for (18). The algorithm (3) with $\mu_\eta = \bar{\mu}_\eta$ is called power-normalized LMS (PN) and the algorithm (18) with $\tilde{\mu}_\eta = \bar{\bar{\mu}}_\eta$ is called normalized signed regressor LMS (NSR).

VI. Simulations

We consider a system identification application with the initial optimal solution formed with $M = 7$ independent random values between 0 and 1, and given by

$$\mathbf{w}_o^T(0) = [+0.90\ -0.54\ +0.21\ -0.03\ +0.78\ +0.52\ -0.09].$$

The input signal $u(n)$ is generated with a first-order autoregressive model, whose transfer function is $\sqrt{1-\alpha^2}/(1-\alpha z^{-1})$, with $\alpha = 0.8$. This model is fed with an i.i.d. Gaussian random process, whose variance is such that $\mathrm{Tr}(\mathbf{R}) = 1$. Moreover, additive i.i.d. noise $v(n)$ with variance $\sigma_v^2 = 0.01$ is added to form the desired signal. To obtain the results shown in Figs. 4 and 5, the algorithm (3) is used to update the mixing parameter $\eta(n)$.

Fig. 4 shows the EMSE and mixing parameter along the iterations for the combination of two RLS filters in the stationary case. The curves were estimated from the ensemble-average of 500 independent runs and filtered by a moving-average filter with 512 coefficients. The dashed lines in the figure show the steady-state predicted values of ζ for each algorithm and their combination. Since the component filters are adapted with close forgetting factors, i.e., $(1-\lambda_2) = 0.9(1-\lambda_1)$, the affine combination provides an EMSE reduction of approximately 3 dB as predicted by the analysis. In this case, the mixing parameter tends to -7.55, which also agrees with the analysis.

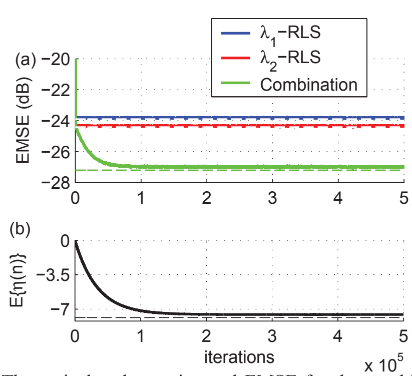

Fig. 4. (a) Theoretical and experimental EMSE for the combination of two RLS filters with $\lambda_1 = 0.9$, $\lambda_2 = 0.91$, and $\mu_\eta = 1$ (b) Ensemble-average of $\eta(n)$, and theoretical value of $\bar{\eta}_o(\infty)$; ensemble-average of 500 independent runs; the theoretical values are indicated by dashed lines.

To illustrate the accuracy of the analysis in a nonstationary environment, we show in Fig. 5 the theoretical and experimental values of the ratio $\zeta/\min\{\zeta_1,\zeta_2\}$, as a function of $\delta=\mu_2/\mu_1$ with fixed $\mu_1=0.1$, considering the combination of two LMS filters and $\mathbf{Q}=\sigma_q^2\mathbf{I}$. As predicted by the expressions of Table IV, the largest EMSE reduction occurs when $\mathrm{Tr}(\mathbf{Q})=\mu_1\mu_2\sigma_v^2\mathrm{Tr}(\mathbf{R})$ or when $\delta\approx 1$, and is limited in both cases by 3 dB. Moreover, for each curve of Fig. 5, there is a value of δ for which $\zeta=\min\{\zeta_1,\zeta_2\}$. At this point, the combination performs as its best component, which is adapted with the optimum step-size $\mu_{\rm o}$ [4, p. 369]. Although the affine combination can provide an EMSE reduction in relation to its components, its minimum EMSE coincides with that of LMS with the optimum step-size $\mu_{\rm o}$. These properties can be exploited to improve the tracking capability of adaptive filters, extending the convex combination of variable step LMS algorithms (CVS-LMS) proposed in [3] to the affine combinations considered here (we intend to pursue this matter elsewhere).

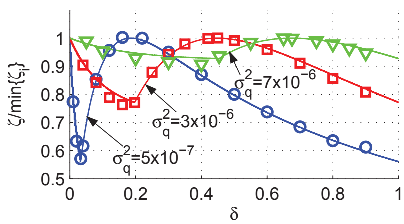

Fig. 5. Theoretical and experimental values of $\zeta/\min\{\zeta_i\}$, $i=1,2$ for the combination of two LMS filters with $\mu_1=0.1$, $\mu_2=\delta\mu_1$, $\mu_\eta=1$, and $\mathbf{Q}=\sigma_q^2\mathbf{I}$. The experimental values are indicated by ◯, □, and ▽; ensemble-average of 50 independent runs.

Fig. 6 shows the EMSE and mixing parameter for the combination of two LMS filters. We consider a system identification application with the initial optimal solution formed with $M=10$ independent Gaussian random values with zero mean and unit variance. The optimum solution is kept constant, except for a change at $n=75000$ (by adding a vector of random Gaussian variables with variance 0.01). The input signal $u(n)$ is generated as before (again with $\alpha=0.8$). The experimental curves were estimated from the ensemble-average of 100 independent runs. The mixing parameter is adapted with the PN and NSR algorithms. Both algorithms provide an adequate behavior for the combination, with $\mathrm{E}\{\eta(n)\}$ following $\bar{\eta}_{\rm o}(n)$ closely. As predicted by the analysis, the combined scheme attains the lower stationary EMSE of the μ_2-LMS and presents the faster convergence of the μ_1-LMS. The variance of the mixing parameter is usually larger for NSR than for PN. However, the mixing parameter adapted with PN may exhibit peaks at the beginning and when the optimum solution changes. This effect is less pronounced when NSR is used. In addition, NSR is less sensitive to variations in the simulation parameters (such as input and noise power, step-sizes, regularization).

VII. Conclusion

We extended the analysis of [1] and [12] to allow for colored inputs and nonstationary environments, considering affine combinations based on LMS, NLMS, and RLS algorithms. Good agreement between analytical and simulation results is always observed. A simple geometrical interpretation of the affine combination allowed us to explain its behavior in different situations, including when the component filters are adapted with close step-sizes or forgetting factors. Furthermore, we proposed and analysed two new normalized algorithms for updating the mixing parameter. The theoretical model explains situations in which the adaptive combination algorithms may achieve good performance.

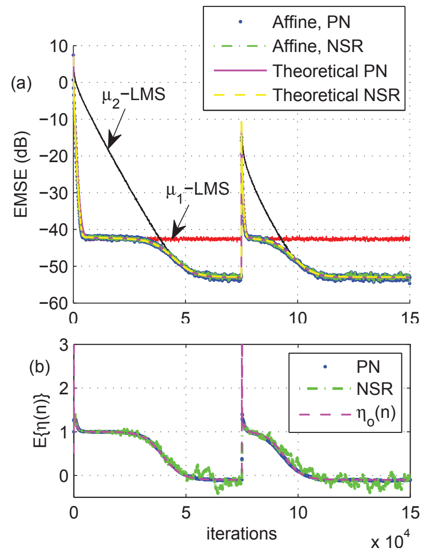

Fig. 6. (a) Experimental and theoretical EMSE for the combination of two LMS filters with $\mu_1=10^{-2}$, $\mu_2=10^{-3}$, using PN ($\tilde{\mu}_\eta=0.01$, $\varepsilon=6\times10^{-4}$, $\lambda_p=0.99$) or NSR ($\tilde{\mu}_\eta=0.0125$, $\varepsilon=0.1$, $\lambda_p=0.99$); (b) Ensemble-average of $\eta(n)$ and $\bar{\eta}_{\rm o}(n)$; $M=10$, $\sigma_v^2=10^{-3}$, correlated regressor with $\mathrm{var}\{u(n)\}=1$ and $\alpha=0.8$, ensemble-average of 100 independent runs.

References

[1] N. J. Bershad, J. C. M. Bermudez, and J.-Y. Tourneret, “An affine combination of two LMS adaptive filters - transient mean-square analysis,” *IEEE Trans. Signal Processing*, vol. 56, pp. 1853–1864, May 2008.

[2] M. Martínez-Ramón *et al.*, “An adaptive combination of adaptive filters for plant identification,” in *Proc. of DSP’2002*, vol. 2. IEEE, 2002, pp. 1195–1198.

[3] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, “Mean-square performance of a convex combination of two adaptive filters,” *IEEE Trans. Signal Processing*, vol. 54, pp. 1078–1090, Mar. 2006.

[4] A. H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, NJ, 2003.

[5] M. T. M. Silva and V. H. Nascimento, “Improving the tracking capability of adaptive filters via convex combination,” *IEEE Trans. Signal Processing*, vol. 56, pp. 3137–3149, Jul. 2008.

[6] E. Eleftheriou and D. D. Falconer, “Tracking properties and steady-state performance of RLS adaptive filter algorithms,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1097–1110, Oct. 1986.

[7] D. T. M. Slock, “On the convergence behavior of the LMS and the normalized LMS algorithms,” *IEEE Trans. Signal Processing*, vol. 41, pp. 2811–2825, Sep. 1993.

[8] E. Eweda, “Comparison of RLS, LMS and sign algorithms for tracking randomly time-varying channels,” *IEEE Trans. Signal Processing*, vol. 42, pp. 2937–2944, Nov. 1994.

[9] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, Upper Saddle River, 2001.

[10] N. R. Yousef and A. H. Sayed, “A unified approach to the steady-state and tracking analyses of adaptive filters,” *IEEE Trans. Signal Processing*, vol. 49, pp. 314–324, Feb. 2001.

[11] M. C. Costa and J. C. M. Bermudez, “An improved model for the normalized LMS algorithm with gaussian inputs and large number of coefficients,” in *Proc. of ICASSP’2002*, vol. II. IEEE, 2002, pp. 1385–1388.

[12] J. C. M. Bermudez, N. J. Bershad, and J.-Y. Tourneret, “An affine combination of two NLMS adaptive filters - transient mean-square analysis,” in *Proc. of 42th Asilomar Conf. on Signals, Systems & Computers*, 2008.

[13] N. J. Bershad, “Analysis of the normalized LMS algorithm with Gaussian inputs,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 793–806, 1986.

[14] L. Azpicueta-Ruiz, A. Figueiras-Vidal, and J. Arenas-Garcia, “A normalized adaptation scheme for the convex combination of two adaptive filters,” in *Proc. of ICASSP’2008*. IEEE, 2008, pp. 3301–3304.