Transient and steady-state analysis of the affine combination of two adaptive filters

Renato Candido, Magno T. M. Silva, *Member, IEEE,* and Vítor H. Nascimento, *Member, IEEE*

***Abstract*—In this paper, we propose an approach to the transient and steady-state analysis of the affine combination of one fast and one slow adaptive filters. The theoretical models are based on expressions for the excess mean-square error (EMSE) and cross-EMSE of the component filters, which allows their application to different combinations of algorithms, such as least mean-squares (LMS), normalized LMS (NLMS), and constant modulus algorithm (CMA), considering white or colored inputs and stationary or nonstationary environments. Since the desired universal behavior of the combination depends on the correct estimation of the mixing parameter at every instant, its adaptation is also taken into account in the transient analysis. Furthermore, we propose normalized algorithms for the adaptation of the mixing parameter that exhibit good performance. Good agreement between analysis and simulation results is always observed.**

***Index Terms*—Adaptive filters, affine combination, tracking, transient analysis, least mean square methods, unsupervised learning.**

I. Introduction

COMBINATION schemes constitute an interesting way to improve adaptive filter performance [3]–[15]. Among these schemes, the convex combination of two fixed step-size adaptive filters has received attention due to its relative simplicity, and the proof that it is universal in steady-state, i.e., the combined estimate is at least as good as the best of the component filters [8].

Convex combination schemes were proposed to improve the fundamental tradeoff between convergence rate and steady-state excess mean-square error (EMSE) in adaptive filters [16]–[19]. Furthermore, such schemes have been exploited in nonstationary environments to improve tracking performance, considering, e.g, the algorithm proposed in [8] or the combination of algorithms with different tracking capabilities of [10]. The correct adjustment of the step-size for the updating of the mixing parameter depends on some characteristics of the filtering scenario, such as the input signal and additive noise powers, or the step-sizes of the adaptive filters considered in the combination. This problem was addressed in [9], where a novel normalized scheme was proposed. It was shown that the new update rule preserves the good features of the existing scheme and is more robust to changes in the filtering scenario.

Using a similar approach to that of [8], the authors of [12] proposed an affine combination of two least mean-square (LMS) algorithms, where the condition on the mixing parameter is relaxed, allowing it to be negative. Thus, this scheme can be interpreted as a generalization of the convex combination since the mixing parameter is not restricted to the interval $[0, 1]$. This approach allows for smaller EMSE in theory, but suffers from larger gradient noise in some situations. Under certain conditions, the optimum mixing parameter was proved to be negative in steady-state. Although the optimal linear combiner is unrealizable, two realizable algorithms were introduced. One is based on a stochastic gradient search and is referred to here as η-LMS algorithm. The other is based on the ratio of the average error powers from each individual adaptive filter. Under some circumstances, both algorithms present performance close to the optimum. In the analysis of [12], white Gaussian inputs and stationary environments are assumed. Furthermore, the behavior of the mean-square deviation is studied only after the fast filter has converged but the slow filter has not yet converged.

Similarly to the convex combination, the correct adjustment of the step-size for the updating of the mixing parameter in the affine combination, denoted by μ_η, depends on some characteristics of the filtering scenario. Hence, the desired universal behavior of the affine combination cannot always be ensured. To illustrate, Fig. 1 shows the EMSE as a function of time for two LMS filters with step-sizes $\mu_1 = 0.01$ (μ_1-LMS) and $\mu_2 = 0.001$ (μ_2-LMS), and their affine combination. In this scenario, it is necessary to use a high value for the step-size of the η-LMS algorithm (e.g., $\mu_\eta = 3$) in order to enable the switching from the slow filter to the fast one. A large value of μ_η may, however, cause instability during the initial convergence of the algorithm, thus [12] constrains $\eta(n)$ to be less than or equal to 1. Unfortunately, even with this constraint, the higher the step-size μ_η, the higher the variance of the mixing parameter during the initial iterations. Therefore, the combination performance deviates from universal, as shown in Fig. 1-(a). On the other hand, if the step-size is small (e.g., $\mu_\eta = 0.1$), the combination performs better in the initial iterations but does not switch as fast as needed from the slow filter to the fast one, as shown in Fig. 1-(b). In fact, we will show that μ_η depends on $\alpha(n) \triangleq \mathrm{E}\{[y_1(n) - y_2(n)]^2\}$, where $\mathrm{E}\{\cdot\}$ represents the expectation operation and $y_i(n)$, $i = 1, 2$ are the outputs of the filters. As we can observe in Fig. 1-(c), $\alpha(n)$ suffers a large variation during the adaptation of the filters. Thus, a transient analysis and alternative algorithms to

This work was partly supported by FAPESP under Grants 2008/00773-1 and 2008/04828-5, and by CNPq under Grants 302633/2008-1 and 303361/2004-2. Some preliminary parts of this work appeared as conference papers [1], [2].

The authors are with the Electronic Systems Engineering Department, Escola Politécnica, University of São Paulo, São Paulo, SP, Brazil, e-mails: {renatocan, magno, vitor}@lps.usp.br, ph. +55-11-3091-5606, fax: +55-11-3091-5718.

SP EDICS: 1.ASP-ANAL, 2.ASP-APPL, 3.SPC-BLND

Corresponding author: Renato Candido.

adapt the mixing parameter are two important key issues for the practical application of the affine combination of adaptive filters.

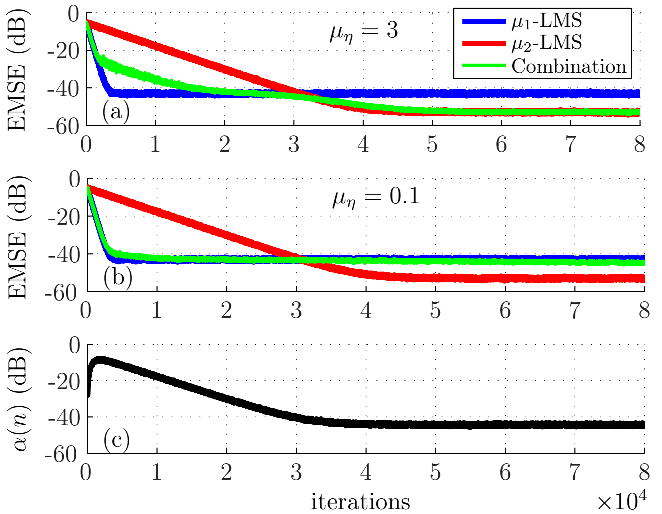

Fig. 1. EMSE for μ_1-LMS, μ_2-LMS and their affine combination a) $\mu_\eta = 3.0$; b) $\mu_\eta = 0.1$; c) ensemble average of $[y_1(n) - y_2(n)]^2$; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $M = 7$, identification of the system $\mathbf{w}_o = [0.9003\ {-0.5377}\ 0.2137\ {-0.0280}\ 0.7826\ 0.5242\ {-0.0871}]^T$, $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs.

A. Contributions and organization of the paper

The present paper extends previous results in four ways:

1) providing a steady-state analysis for the optimum affine combination of adaptive filters, which is valid for white or colored inputs, stationary or nonstationary environments, and combinations based on different algorithms, such as LMS, normalized LMS (NLMS), and the constant modulus algorithm (CMA);
2) proposing a simple geometrical interpretation to explain the behavior of the affine combination;
3) providing a transient analysis of the combination, taking into account the adaptation of the component filters and also the adaptation of the mixing parameter with the η-LMS algorithm;
4) using the results of the transient analysis to facilitate the adjustment of the free parameters of the scheme and to propose two normalized algorithms to update the mixing parameter.

To the best of our knowledge, all these are novel contributions.

The paper is organized as follows. In the next section, we describe the affine combination of two adaptive filters for both supervised (LMS and NLMS) and blind (CMA) algorithms. In Section III, analytical expressions for the optimum mixing parameter and the optimum EMSE of the combination are obtained. In the steady-state analysis of Section IV, the results of Section III are particularized for optimum combinations of two LMS filters, two NLMS filters, and two CMA equalizers, considering stationary and nonstationary environments. In Section V, transient analyses taking into account realizable schemes are presented. Initially, we obtain an analytical expression for the EMSE of the combination, which depends on the transient models of the combined algorithms and also on the algorithm used to adapt the mixing parameter. We summarize results for the transient analysis of LMS, NLMS, and CMA. Then, in Section V-A, we present the transient analysis of the η-LMS algorithm. The resulting analysis suggests the normalization procedure presented in Section V-B and an algorithm with partial instantaneous normalization, proposed and analyzed in Section V-C. Comparisons between analytical and experimental results are shown through simulations in Section VI. Section VII provides a summary of the main conclusions of the paper.

II. Problem Formulation

This section is divided into three parts. We first describe the affine combination of one fast and one slow supervised algorithms. In the sequel, the combination of two CMA equalizers is presented. Then, we propose a common formulation for the affine combination of supervised (LMS and NLMS) or blind (CMA) algorithms.

A. Combination of supervised algorithms

The linear combination of two supervised adaptive filters is depicted in Fig. 2, where the filter weights are adjusted to minimize the mean-square error cost function, obtaining at the output an estimate of the given "desired signal" $d(n)$. 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 and $y_i(n)$, $i = 1, 2$ are the outputs of two transversal filters, i.e., $y_i(n) = \mathbf{u}^T(n)\mathbf{w}_i(n-1)$. The superscript T denotes transposition, $\mathbf{w}_i(n-1)$, $i = 1, 2$ represent the length-M coefficient column-vectors characterizing the component filters, and $\mathbf{u}(n)$ is their common input regressor column-vector.

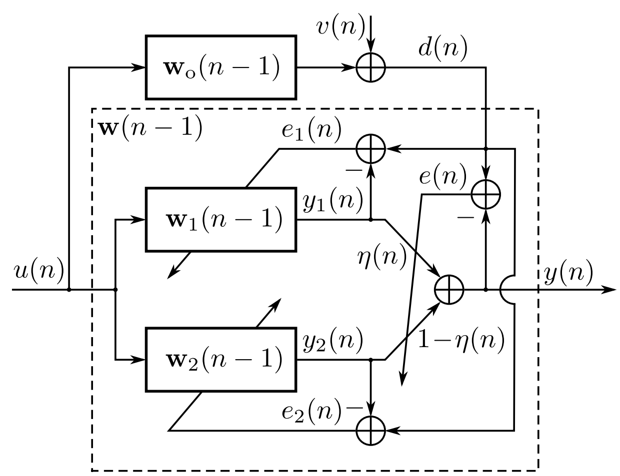

Fig. 2. Linear combination of two supervised adaptive filters.

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

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

where $\rho_i(n)$ is a step-size and $e_i(n)$ is the estimation error. Many algorithms can be written as in (2), by proper choices of $\rho_i(n)$ and $e_i(n)$. For some algorithms $\rho_i(n)$ can even be a matrix, as is the case of the recursive-least squares (RLS) algorithm, where $\rho_i(n)$ is an estimate of the inverse autocorrelation matrix of the input signal. In supervised adaptive filtering, a "desired signal" $d(n)$ is available such that

$$e_i(n) = d(n) - y_i(n) \tag{3}$$

and a linear regression model holds, i.e.,

$$d(n) = \mathbf{u}^T(n)\mathbf{w}_{\rm o}(n-1) + v(n) \tag{4}$$

with $\mathbf{w}_{\rm o}(n-1)$ being the time-variant optimal solution and $v(n)$ a zero-mean random process uncorrelated with $\mathbf{u}(n)$, whose variance is denoted by $\sigma_v^2 = \mathrm{E}\{v^2(n)\}$. In order to make performance analyses more tractable, the sequences $\{\mathbf{u}(n)\}$ and $\{v(n)\}$ are assumed stationary and we will use the common assumption that $v(n)$ is independent of $\mathbf{u}(n)$ (not only uncorrelated) [17, Sec. 6.2.1]. Defining the weight-error vectors $\widetilde{\mathbf{w}}_i(n) = \mathbf{w}_{\rm o}(n) - \mathbf{w}_i(n)$, the *a priori* errors $e_{a,i}(n) = \mathbf{u}^T(n)\widetilde{\mathbf{w}}_i(n-1)$, and using the linear model (4), we find that

$$e_i(n) = e_{a,i}(n) + v(n). \tag{5}$$

An important consequence of this model is that $v(k)$ will be independent of all $\mathbf{w}_i(j)$, $\widetilde{\mathbf{w}}_i(j)$, and $e_{a,i}(k)$, $i = 1, 2$, $j < k$, for any particular time instant k [17, Lemma 6.2.1].

Considering the combination of two LMS filters and the minimization of the overall instantaneous square error $e^2(n) = [d(n) - y(n)]^2$, [12] proposed the following gradient-based algorithm

$$\eta(n+1) = \eta(n) + \mu_\eta e(n)\left[y_1(n) - y_2(n)\right]. \tag{6}$$

To obtain a tradeoff between stability of this recursion and the algorithm's tracking capability in the initial phase of adaptation, $\eta(n)$ in (6) must be constrained to be less than or equal to 1 for all n. We should remark that this kind of constraint is not needed in the normalized algorithms proposed in Sections V-B and V-C.

B. Combination of blind algorithms

Fig. 3 shows a simplified communications system with a combination of two blind equalizers. In this case, the signal $a(n)$, assumed i.i.d. (independent and identically distributed) and non Gaussian, is transmitted through an unknown channel, whose model is constituted by an FIR (finite impulse response) filter and additive white Gaussian noise. From the received signal $u(n)$ and the known statistical properties of the transmitted signal, the blind equalizer must mitigate the channel effects and recover the signal $a(n)$ for some delay τ_d. We also assume that the equalization algorithms are implemented in $T/2$-fractionally spaced form, due to its inherent advantages (see, e.g., [20]–[23] and the references therein). This type of implementation is widely used in the literature since it ensures perfect equalization in a noise-free environment, under certain well-known conditions. The output of the overall equalizer of Fig. 3 is also given by (1).

Algorithms based on the constant modulus cost function [24], [25] define the "estimation error" as

$$e_i(n) = [r - y_i^2(n)]y_i(n), \tag{7}$$

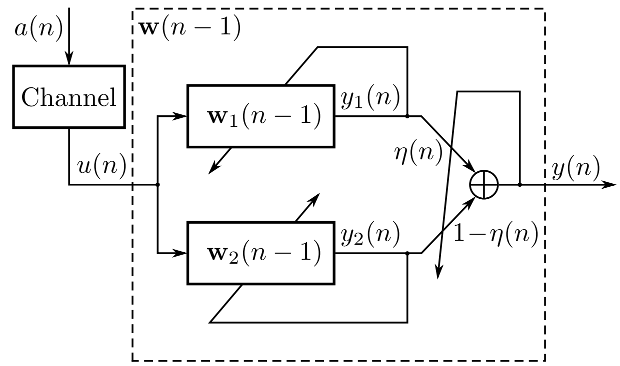

Fig. 3. Simplified communications system with a linear combination of two blind equalizers.

where $r = \mathrm{E}\{a^4(n)\}/\mathrm{E}\{a^2(n)\}$. Using (7), CMA can also be written as in (2). However, given the nonlinear nature of CMA, additional assumptions are necessary to obtain a model as simple as (5): essentially, large signal-to-noise ratio, circular symmetry of the transmitted constellation, and an initial condition close to the zero-forcing solution (see Appendix A). These assumptions were used in [10] and [26] to obtain simple linear models that capture the behavior of CMA close to an optimum solution. Thus, (7) was approximated by

$$e_i(n) \approx \gamma(n)e_{a,i}(n) + \beta(n), \tag{8}$$

where

$$\gamma(n) \triangleq 3a^2(n-\tau_d) - r \tag{9}$$

and

$$\beta(n) \triangleq ra(n-\tau_d) - a^3(n-\tau_d). \tag{10}$$

The variable $\beta(n)$ is identically zero for constant-modulus constellations, so the variability in the modulus of $a(n)$ (as measured by $\beta(n)$) plays the role of measurement noise for constant-modulus based algorithms. Model (8) was proposed in [10] to study convex combinations of constant-modulus-based algorithms and extended in [26] to obtain explicit stability conditions for CMA. The main assumptions and the derivation of this model are summarized in Appendix A.

To update the mixing parameter in order to combine two CMA equalizers, we could use a gradient rule to minimize the instantaneous constant-modulus cost $\hat{J}_{cm}(n) = [r - y^2(n)]^2$, as considered in the convex combination of [27]. However, we observed through simulations that the resulting algorithm does not always ensure the desired universal behavior of the combination, specially for nonconstant modulus signals. Thus, we propose a stochastic gradient algorithm to minimize the instantaneous square decision error $\hat{J}_d(n) = e_d^2(n)$, where $e_d(n) \triangleq \hat{a}(n-\tau_d) - y(n)$ and $\hat{a}(n-\tau_d)$ is the estimate of the transmitted signal at the output of the decision device. This results in the following update equation

$$\eta(n+1) = \eta(n) + \mu_\eta e_d(n)\left[y_1(n) - y_2(n)\right]. \tag{11}$$

We observed through simulations that the decision-error-based adaptation ensures a more adequate behavior than that of the constant-modulus-based adaptation, even in presence of noise and/or when both component filters are far from convergence. Assuming that $\hat{a}(n-\tau_d) = a(n-\tau_d)$ and that the optimal solution achieves perfect equalization (see Assumption B1 in Appendix A), the minimization of $J_d(n)$ is equivalent to the minimization of the square *a priori* error, since under these assumptions $e_d(n) \approx e_a(n)$.

C. A common formulation

Comparing (8) to (5), we can write the following general expression

$$e_i(n) = \kappa(n) e_{a,i}(n) + \varphi(n), \quad i = 1, 2, \tag{12}$$

where $\kappa = 1$ and $\varphi(n) = v(n)$ for a supervised algorithm or $\kappa(n) = \gamma(n)$ and $\varphi(n) = \beta(n)$ for a blind one. In both cases $\mathrm{E}\{\varphi(n)\} = 0$. This model also holds for the overall scheme, i.e.,

$$e(n) = \kappa(n) e_a(n) + \varphi(n), \tag{13}$$

where $e(n)$ represents the error of the combined filter: $e(n) = d(n) - y(n)$ for supervised algorithms or $e(n) = [r - y^2(n)]y(n)$ for constant-modulus-based algorithms, and $e_a(n)$ is the *a priori* error of the overall scheme. It should be noticed that (12) and (13) are approximations in the blind case. For the sake of simplicity, we use the equality sign here and in the expressions derived from (12) and (13).

The supervised LMS and NLMS algorithms and the blind CMA employ the step-sizes $\rho_i(n)$ and the estimation errors $e_i(n)$ as in Table I, where ϵ_N is a regularization factor and $\|\cdot\|$ represents the Euclidean norm. The models for the errors $e_i(n)$ of these algorithms are also shown in this table for convenient reference. The step-size interval which ensures the convergence and stability is different for each algorithm. For the LMS and NLMS algorithms, the step-size intervals are well-known in the literature [16], [17], whereas for CMA, the derivation of this interval was shown recently in [26].

TABLE I
PARAMETERS OF THE CONSIDERED ALGORITHMS AND ERROR MODELS.

Alg.	$\rho(n)$	$e_i(n)$	Model for $e_i(n)$
LMS	μ_i	$d(n) - y_i(n)$	$e_{a,i}(n) + v(n)$
NLMS	$\dfrac{\mu_i}{\epsilon_N + \|\mathbf{u}(n)\|^2}$		
CMA	μ_i	$[r - y_i^2(n)]y_i(n)$	$\gamma(n)e_{a,i}(n) + \beta(n)$

Using model (5) in the supervised case, and the fact that $e_d(n) \approx e_a(n)$ in the blind case, we can write a general expression for updating the mixing parameter, i.e.,

$$\eta(n+1) = \eta(n) + \mu_\eta e_g(n)[y_1(n) - y_2(n)], \tag{14}$$

where

$$e_g(n) = e_a(n) + b(n) \tag{15}$$

and $b(n) = v(n)$ for the combination of supervised algorithms or $b(n) = 0$ for the combination of constant-modulus-based algorithms. In both cases, $\eta(n)$ is constrained to be less than or equal to 1 for all n [12]. Algorithm (14) is denoted here by η-LMS.

To close this section, it is important to observe that:

1) In order to simplify the arguments, we assume that all the quantities are real. In the case of blind equalization of complex constellations, complex extensions may be developed, provided signal circularity conditions are satisfied [28];
2) The analyses provided here can be extended straightforwardly to the affine combination of two RLS filters [1], of two Shalvi-Weinstein equalizers [10], [29], and also to the combination of algorithms of different families, as is the case of the combination of one LMS with one RLS or of the combination of one CMA with one Shalvi-Weinstein algorithm [10];
3) Besides the η-LMS algorithm, [12] proposed a scheme based on error powers to update the mixing parameter. Although this scheme also presents performance close to the optimum under certain circumstances, its structure is significantly different from that of η-LMS, so we leave its analysis for a future work.

III. THE OPTIMUM MIXING PARAMETER AND EMSE

An analytical expression for the optimum mixing parameter $\eta_o(n)$ can be obtained equating to zero the expected value of the gradient used to update $\eta(n)$ in (14), i.e.,

$$\mathrm{E}\left\{e_g(n)[y_1(n) - y_2(n)]\right\} = 0. \tag{16}$$

The error $e_g(n)$ in (16) can be rewritten as a function of the *a priori* errors $e_{a,i}(n)$, $i = 1, 2$, as follows.

Using (1), (12), and (13), the *a priori* error $e_a(n)$ of the overall scheme can be written as

$$\begin{aligned} e_a(n) &= \eta(n)e_{a,1}(n) + [1 - \eta(n)]e_{a,2}(n) \\ &= e_{a,2}(n) + \eta(n)\left[e_{a,1}(n) - e_{a,2}(n)\right]. \end{aligned} \tag{17}$$

Replacing (17) in (15), and remarking that $y_1(n) - y_2(n) = e_{a,2}(n) - e_{a,1}(n)$, (16) can be rewritten as

$$\begin{aligned} &\mathrm{E}\left\{e_{a,2}^2(n) - e_{a,1}(n)e_{a,2}(n)\right\} \\ &- \mathrm{E}\left\{\eta_o(n)[e_{a,2}(n) - e_{a,1}(n)]^2\right\} \\ &+ \mathrm{E}\left\{b(n)[e_{a,2}(n) - e_{a,1}(n)]\right\} = 0. \end{aligned} \tag{18}$$

In the blind case, $b(n) = 0$ and in the supervised case, $b(n) = v(n)$, which is assumed independent of $e_{a,i}(n)$, $i = 1, 2$. Hence, in both cases the third term on the l.h.s. of (18) is equal to zero.

To proceed, we remark that the EMSE of the component filters and the cross-EMSE can be calculated [8], respectively as

$$\zeta_{ii}(n) \triangleq \mathrm{E}\{e_{a,i}^2(n)\}, \quad i = 1, 2, \quad \text{and} \tag{19}$$

$$\zeta_{12}(n) \triangleq \mathrm{E}\{e_{a,1}(n)e_{a,2(n)}\}. \tag{20}$$

Introducing the differences

$$\Delta\zeta_{ii}(n) \triangleq \zeta_{ii}(n) - \zeta_{12}(n), \quad i = 1, 2, \tag{21}$$

and using (19)-(21) in (18), we arrive at

$$\eta_{\rm o}(n) = \frac{\Delta\zeta_{22}(n)}{\Delta\zeta_{11}(n) + \Delta\zeta_{22}(n)}. \tag{22}$$

A similar expression was also obtained in [8, Eq.(29)] for the convex combination of two LMS filters at the steady-state. We should notice that (22) is more general: it holds for all $n \geq 0$ (not only at the steady-state) and the mixing parameter is not restricted to the interval $[0,\, 1]$.

Defining the EMSE of the overall combined scheme as

$$\zeta(n) = {\rm E}\{e_a^2(n)\}, \tag{23}$$

we now obtain an analytical expression for its optimum value. By squaring both sides of (17) with $\eta(n) = \eta_{\rm o}(n)$ and taking expectations, we arrive at

$$\begin{aligned}{\rm E}\{e_a^2(n)\} =& \eta_{\rm o}^2(n){\rm E}\{e_{a,1}^2(n)\} + [1-\eta_{\rm o}(n)]^2{\rm E}\{e_{a,2}^2(n)\} \\ &+ 2\eta_{\rm o}(n)[1-\eta_{\rm o}(n)]{\rm E}\{e_{a,1}(n)e_{a,2}(n)\}.\end{aligned} \tag{24}$$

Using (19)-(22) in (24), we obtain

$$\zeta_{\rm o}(n) = \zeta_{22}(n) - \eta_{\rm o}(n)\Delta\zeta_{22}(n). \tag{25}$$

After some algebraic manipulations, (25) can be rewritten as

$$\zeta_{\rm o}(n) = \zeta_{12}(n) + \frac{\Delta\zeta_{11}(n)\Delta\zeta_{22}(n)}{\Delta\zeta_{11}(n) + \Delta\zeta_{22}(n)}. \tag{26}$$

This expression was obtained in [8, Eq. (33)] for the convex combination of two LMS filters at the steady-state, but again it also holds for all $n \geq 0$.

As already mentioned in [8], (22) and (26) hold for the combination of any two algorithms that satisfy (12). The values of $\Delta\zeta_{ii}(n)$, $i = 1, 2$ however do depend on the actual algorithms that are being combined. Thus, provided approximations for $\zeta_{ij}(n)$, $i, j = 1, 2$ are available, (22) and (26) can be applied to the affine combination of different algorithms, including combinations of algorithms of different families.

IV. Steady-state analysis of the optimum combiner

In this section, the optimum mixing parameter and the optimum EMSE of the combination, given respectively by expressions (22) and (26), are particularized for the combination of two LMS filters, two NLMS filters, and two CMA equalizers in steady-state for stationary and nonstationary environments. We do not rederive the steady-state expressions for $\zeta_{ij}(\infty)$, $i, j = 1, 2$ here, only use the best approximations from the literature. As in [12], we assume that the algorithm which updates the mixing parameter is able to achieve the optimum value in steady-state. Realizable schemes for adaptation of $\eta(n)$ are taken into account in Section V.

We assume that in a nonstationary environment, the variation in the optimal solution $\mathbf{w}_{\rm o}$ follows a random-walk model [17, p. 359], that is,

$$\mathbf{w}_{\rm o}(n) = \mathbf{w}_{\rm o}(n-1) + \mathbf{q}(n). \tag{27}$$

In this model, $\mathbf{q}(n)$ is an i.i.d. vector with positive-definite autocorrelation matrix $\mathbf{Q} = {\rm E}\{\mathbf{q}(n)\mathbf{q}^T(n)\}$, independent of the initial conditions $\{\mathbf{w}_{\rm o}(-1), \mathbf{w}(-1), \eta(-1)\}$ and of $\{\mathbf{u}(l)\}$ for all l [17, Sec. 7.4]. In supervised filtering, $\mathbf{q}(n)$ is also assumed independent of the desired response $\{d(l)\}$ for all $l < n$. In blind equalization, $\mathbf{w}_{\rm o}(n)$ represents the zero-forcing solution and $\mathbf{q}(n)$ models the channel variation.

Table II lists analytical expressions of $\zeta_{12}(\infty)$ for some pairs of filters. Expressions for $\zeta_{ii}(\infty)$ can be obtained from Table II making $\mu_1 = \mu_2$. Details about the derivation of these expressions can be found in [8], [10], [13] for the cross-terms and in [16], [17], [23], [28], [30]–[34] for the case $\mu_1 = \mu_2$. In this table, $\mathbf{R} \triangleq {\rm E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}$ is the autocorrelation matrix of the input signal, ${\rm Tr}(\mathbf{A})$ stands for the trace of matrix $\mathbf{A}$, and $\nu_u \triangleq {\rm E}\left\{\|\mathbf{u}(n)\|^{-2}\right\}$. For Gaussian inputs and large number of coefficients, ν_u can be approximated by $1/[\sigma_u^2(M-2)]$ with $\sigma_u^2 = {\rm E}\{u^2(n)\}$ [13], [35]. The constants σ_β^2, $\bar{\gamma}$, and ξ, which appear in the expression for the EMSE of CMA, depend on statistics of the transmitted signal and are defined in Appendix A.

TABLE II

Analytical expressions for the steady-state cross-EMSE of the considered combinations.

Combination	$\zeta_{12}(\infty)$
μ_1-LMS and μ_2-LMS	$\dfrac{\mu_1\mu_2\sigma_v^2{\rm Tr}(\mathbf{R}) + {\rm Tr}(\mathbf{Q})}{\mu_1 + \mu_2 - \mu_1\mu_2{\rm Tr}(\mathbf{R})}$
μ_1-NLMS and μ_2-NLMS	$\dfrac{{\rm Tr}(\mathbf{R})\left[\mu_1\mu_2\sigma_v^2\nu_u + {\rm Tr}(\mathbf{Q})\right]}{\mu_1+\mu_2-\mu_1\mu_2}$
μ_1-CMA and μ_2-CMA	$\dfrac{\mu_1\mu_2\sigma_\beta^2{\rm Tr}(\mathbf{R}) + {\rm Tr}(\mathbf{Q})}{\bar{\gamma}(\mu_1 + \mu_2) - \mu_1\mu_2{\rm Tr}(\mathbf{R})\xi}$

A. Stationary environments

Replacing the expressions of Table II with $\mathbf{Q} = \mathbf{0}$ in (22) and (26), we obtain analytical expressions for the steady-state optimum mixing parameter $\eta_{\rm o}(\infty)$ and for the steady-state optimum EMSE $\zeta_{\rm o}(\infty)$ in stationary environments. The resulting expressions are shown in Table III, where we defined $\delta \triangleq \mu_2/\mu_1$ with $0 < \delta < 1$. It is worth to notice that the second filter of the combination is always assumed to be the slower filter ($\mu_2 < \mu_1$) which consequently presents the lower steady-state EMSE in a stationary environment.

TABLE III

Analytical expressions for $\eta_{\rm o}(\infty)$ and $\zeta_{\rm o}(\infty)$.

Combination	$\eta_{\rm o}(\infty)$	$\zeta_{\rm o}(\infty)$
(μ_1 and μ_2)-LMS	$\dfrac{\delta[2-\mu_1{\rm Tr}(\mathbf{R})]}{2(\delta-1)}$	$\dfrac{1}{2}\left[\dfrac{\mu_2\sigma_v^2{\rm Tr}(\mathbf{R})}{\delta+1-\mu_2{\rm Tr}(\mathbf{R})}\right]$
(μ_1 and μ_2)-NLMS	$\dfrac{\delta[2-\mu_1]}{2(\delta-1)}$	$\dfrac{1}{2}\left[\dfrac{{\rm Tr}(\mathbf{R})\mu_2\sigma_v^2\nu_u}{\delta+1-\mu_2}\right]$
(μ_1 and μ_2)-CMA	$\dfrac{\delta[2-\mu_1{\rm Tr}(\mathbf{R})\xi(\bar{\gamma})^{-1}]}{2(\delta-1)}$	$\dfrac{1}{2}\left[\dfrac{\mu_2\sigma_\beta^2{\rm Tr}(\mathbf{R})}{(\delta+1)\bar{\gamma}-\mu_2{\rm Tr}(\mathbf{R})\xi}\right]$

The expressions of Table III show two interesting properties:

i) $\eta_{\rm o}(\infty)$ is negative for all considered combinations, which can be verified through the stability conditions of the algorithms. To ensure the stability of μ_1-LMS and μ_1-NLMS, the step-sizes should be chosen respectively in the following ranges $0 < \mu_1 < 2/{\rm Tr}(\mathbf{R})$ and

$0 < \mu_1 < 2$ [19]. In the case of μ_1-CMA, assuming model (8), it was shown in [26, Eq. (14)] that the range of step-sizes $0 < \mu_1 < 2\bar{\gamma}/[3\text{Tr}(\mathbf{R})\xi]$ guarantees good performance. Choosing the step-sizes in these ranges, we can verify from the expressions of Table III that $\eta_o(\infty) < 0$.

ii) $\delta \approx 1$ yields $\zeta_o(\infty) \approx \zeta_2(\infty)/2$. Since $\zeta_2(\infty) < \zeta_1(\infty)$ for all 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 [12] for the combination of two LMS filters, assuming Gaussian and white inputs, and additional assumptions equivalent to choosing the LMS step-size for maximum convergence speed. If we consider $\mu_1 = 1/\text{Tr}(\mathbf{R})$ in the expression of Table III, we recover the result of [12, Eq.(26)]. The same property was observed in [13] for the affine combination of two NLMS filters, also assuming white and Gaussian inputs.

An intuitive explanation for Property ii) can be found as follows. Using (12), the overall steady-state error is written as

$$e(n) = \underbrace{e_2(n)}_{d_\eta(n)} + \eta(n) \underbrace{\kappa(n)[\mathbf{w}_2(n) - \mathbf{w}_1(n)]^T \mathbf{u}(n)}_{-u_\eta(n)}. \tag{28}$$

From the point of view of the computation of $\eta(n)$, $d_\eta(n)$ represents the signal which has to be estimated, and $u_\eta(n)$ plays the role of input signal. Assuming that $\mathbf{w}_i(n)$, $i = 1, 2$ vary slowly compared to $\eta(n)$, (28) has a simple geometric interpretation as shown in Fig. 4. The affine combination seeks the best weight vector in the line $\mathbf{w}_2 + \eta(\mathbf{w}_1 - \mathbf{w}_2)$. In Fig. 4-(a), the best linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$ is $\mathbf{w}$. In the case of close step-sizes, we also have close coefficient vectors in steady-state[1], i.e., $\mathbf{w}_1 \approx \mathbf{w}_2$ (Fig. 4-(b)), and η has to assume a large value to take the combined vector close to $\mathbf{w}$, since the input signal $u_\eta(n)$ depends on the difference between $\mathbf{w}_1$ and $\mathbf{w}_2$. Thus, if $(\mathbf{w}_1 - \mathbf{w}_2) \to 0$, $|\eta| \to \infty$.

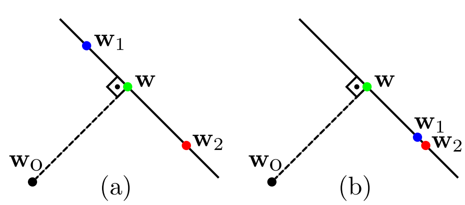

Fig. 4. Geometric interpretation of the affine combination.

B. Nonstationary environments

In a nonstationary environment, the largest EMSE reduction of the affine combination in relation to its components occurs when $\zeta_{11}(\infty) \approx \zeta_{22}(\infty)$. This can happen in two situations (see Table IV):

i) when the step-sizes are not close to one another and $\text{Tr}(\mathbf{Q}) = q_{12}$, where q_{12} is the value of $\text{Tr}(\mathbf{Q})$ for which $\zeta_{11}(\infty) \approx \zeta_{22}(\infty)$;

ii) when the component filters are adapted with close step-sizes ($\delta \approx 1$). However, when $\delta \approx 1$ and $\text{Tr}(\mathbf{Q}) \approx q_{12}$, the gain is small.

Replacing the expressions of Table II under the small step-size approximation[2] in (22) and (26), we obtain analytical expressions for q_{12} and $\zeta_o(\infty)$ shown in Table IV. From these expressions, we can observe that the EMSE reduction in all cases is limited by 3 dB. A reduction close to 3 dB will occur when $\delta \to 0$ in case (i) or when the environment tends to be stationary ($\text{Tr}(\mathbf{Q}) \approx 0$) in case (ii). It is relevant to notice that case (i) also occurs in the convex combination of adaptive filters since in this case $0 < \eta_o(\infty) < 1$. On the other hand, case (ii) occurs only in the affine combination since $\eta_o(\infty)$ does not lie in the interval $[0, 1]$.

The 3 dB gain is an interesting property inherent to the affine combination. However, we should emphasize that using the affine combination with the filters adapted with different step-sizes is more worthwhile than using it with close step-sizes. In the stationary case for $\mu_2 < \mu_1$, the closer $\zeta_{22}(\infty)$ to $\zeta_{11}(\infty)$ the closer the EMSE gain to 3 dB. Although a gain increase can be obtained with close step-sizes, the EMSE of the combination is higher in absolute terms as $\zeta_{22}(\infty)$ becomes closer to $\zeta_{11}(\infty)$. On the other hand, for a single adaptive filter in nonstationary environments there is an optimal value of the step-size for which the steady-state EMSE is minimum [16], [17]. The EMSE of the combination achieves its smallest value when one of its component filters is adapted with this optimum step-size. In this case, the combined estimate is as good as that of the optimum component, and there is no EMSE gain, as is illustrated in a simulation of Section VI-D (see Fig. 17). Moreover, it was shown analytically in [36] that a combination of two filters from the same family (i.e., two LMS or two RLS filters) cannot improve the performance over that of a single filter of the same type with optimal selection of the step-size (or forgetting factor).

TABLE IV
ANALYTICAL EXPRESSIONS FOR q_{12} AND $\zeta_o(\infty)$ FOR CASES (i) AND (ii) IN A NONSTATIONARY ENVIRONMENT.

Combination	(i)		(ii)
	q_{12}	$\zeta_o(\infty)$	$\zeta_o(\infty)$
μ_1-LMS and μ_2-LMS	$\mu_1\mu_2\sigma_v^2 \times\text{Tr}(\mathbf{R})$	$\zeta_{22}(\infty)/2 + \dfrac{2\delta\zeta_{22}(\infty)}{(1+\delta)^2}$	$\zeta_{22}(\infty)/2 + \dfrac{\sigma_v^2\text{Tr}(\mathbf{R})\text{Tr}(\mathbf{Q})}{2\zeta_{22}(\infty)}$
μ_1-NLMS and μ_2-NLMS	$\mu_1\mu_2\sigma_v^2 \times\nu_u$	$\zeta_{22}(\infty)/2 + \dfrac{2\delta\zeta_{22}(\infty)}{(1+\delta)^2}$	$\zeta_{22}(\infty)/2 + \dfrac{\sigma_v^2[\text{Tr}(\mathbf{R})]^2\text{Tr}(\mathbf{Q})\nu_u}{2\zeta_{22}(\infty)}$
μ_1-CMA and μ_2-CMA	$\mu_1\mu_2\sigma_\beta^2 \times\text{Tr}(\mathbf{R})$	$\zeta_{22}(\infty)/2 + \dfrac{2\delta\zeta_{22}(\infty)}{(1+\delta)^2}$	$\zeta_{22}(\infty)/2 + \dfrac{\sigma_\beta^2\text{Tr}(\mathbf{R})\text{Tr}(\mathbf{Q})}{2\bar{\gamma}\zeta_{22}(\infty)}$

[1]It is possible to prove using the results of Tables II and V that if two stable adaptive filters are initialized with the same vector and adapted with close step-sizes the following limit holds

$$\lim_{\delta\to 1}\left\{\lim_{n\to\infty}\frac{\text{E}\{\|\widetilde{\mathbf{w}}_1(n)-\widetilde{\mathbf{w}}_2(n)\|^2\}}{\text{E}\{\|\widetilde{\mathbf{w}}_1(n)\|^2\}}\right\} = 0.$$

In other words, close step-sizes imply close coefficient vectors in steady-state.

[2]The small step-size approximation was assumed in order to obtain simpler expressions.

V. Transient analysis of realizable schemes

In this section, we take into account the adaptation of $\eta(n)$ in the analysis. Our focus will be on how much a realizable estimate for $\eta_o(n)$ deviates from the optimum, and how this affects the combination's overall performance.

By squaring both sides of (17) and taking expectations, we obtain

$$\mathrm{E}\{e_a^2(n)\} = \mathrm{E}\{e_{a,2}^2(n)\} + \mathrm{E}\{\eta^2(n)[e_{a,1}(n) - e_{a,2}(n)]^2\} + 2\mathrm{E}\{\eta(n)\left[e_{a,2}(n)e_{a,1}(n) - e_{a,2}^2(n)\right]\}. \quad (29)$$

To proceed, we assume that:

A1. The adaptation of $\eta(n)$ is slow so that the correlation between it and $e_{a,i}(n)e_{a,j}(n)$, $i,j = 1,2$ can be disregarded.

This assumption follows from observations: simulations show that $\eta(n)$ converges slowly compared to variations in the input $\mathbf{u}(n)$ and thus to variations on the *a-priori* errors.

Using A1, (19)-(21) and (23), we can rewrite (29) as

$$\zeta(n) \approx \zeta_{22}(n) + \mathrm{E}\{\eta^2(n)\}\alpha(n) - 2\mathrm{E}\{\eta(n)\}\Delta\zeta_{22}(n), \quad (30)$$

where we defined

$$\alpha(n) \triangleq \mathrm{E}\{[y_1(n) - y_2(n)]^2\} = \Delta\zeta_{11}(n) + \Delta\zeta_{22}(n). \quad (31)$$

To estimate the EMSE of the combination for all $n \geq 0$ using (30), analytical expressions for $\zeta_{ij}(n)$, $i = 1,2$, $\mathrm{E}\{\eta(n)\}$, and $\mathrm{E}\{\eta^2(n)\}$ should be obtained.

It is common in the literature to evaluate the EMSE as

$$\zeta_{ij}(n) \triangleq \mathrm{E}\{e_{a,i}(n)e_{a,j}(n)\} \approx \mathrm{Tr}(\mathbf{R}\mathbf{S}_{ij}(n-1)), \quad (32)$$

where

$$\mathbf{S}_{ij}(n) \triangleq \mathrm{E}\{\widetilde{\mathbf{w}}_i(n)\widetilde{\mathbf{w}}_j^T(n)\}, \quad i = 1,2 \quad (33)$$

is the covariance ($i = j$) or the cross-variance ($i \neq j$) matrix of the weight-error vector. This approach is based on the independence assumption between the regressor vector $\mathbf{u}(n)$ and weight-error vectors $\widetilde{\mathbf{w}}_i(n-1)$, $i = 1,2$ and is justified for small step-sizes due to the different time-scales for variations in $\mathbf{u}(n)$ and $\widetilde{\mathbf{w}}_i(n-1)$. This condition is a part of the widely used independence assumptions in adaptive filter theory [16]–[19], [37]. Recursions for $\mathbf{S}_{ii}(n)$, $i = 1,2$ are generally obtained in the transient analysis of adaptive filters (see, e.g, [16], [18], [26], [35] and their references). In the transient analysis of linear combinations of two adaptive filters, an estimate of $\mathbf{S}_{ij}(n)$, $i \neq j$ should also be obtained, which is a straightforward extension from the case $i = j$ [10]. In Table V, we show the recursions for the cross-variance matrix $\mathbf{S}_{12}(n)$. Expressions for the covariance matrix $\mathbf{S}_{ii}(n)$, $i = 1,2$ can be obtained from this table, making $\mu_1 = \mu_2$. Using the expressions of Table V in conjunction with (32), $\zeta_{ij}(n)$, $i,j = 1,2$ can be estimated for all $n \geq 0$. The expression for $\mathbf{S}_{12}(n)$ considering the combination of two NLMS was derived using the approach from [35], under the assumptions of Gaussian inputs and large number of coefficients. We should notice that steady-state approximations for the EMSE and cross-EMSE of the component filters can be obtained from the expressions of Table V, i.e., $\zeta_{ij}(\infty) \approx \mathrm{Tr}(\mathbf{R}\mathbf{S}_{ij}(\infty))$, $i,j = 1,2$. However, this procedure leads to more complex expressions than those of Table III.

TABLE V
Recurrent expressions for cross-variance matrix $\mathbf{S}_{12}(n)$.

Combination	$\mathbf{S}_{12}(n)$
μ_1-LMS and μ_2-LMS	$\mathbf{S}_{12}(n) \approx \mathbf{S}_{12}(n-1) - \mu_1\mathbf{R}\mathbf{S}_{12}(n-1) - \mu_2\mathbf{S}_{12}(n-1)\mathbf{R} + \mu_1\mu_2[2\mathbf{R}\mathbf{S}_{12}(n-1)\mathbf{R} + \mathbf{R}\mathrm{Tr}(\mathbf{R}\mathbf{S}_{12}(n-1)) + \sigma_v^2\mathbf{R}] + \mathbf{Q}$
μ_1-NLMS and μ_2-NLMS	$\mathbf{S}_{12}(n) \approx \mathbf{S}_{12}(n-1) - \dfrac{\mu_1}{\sigma_u^2(M-2)}\mathbf{R}\mathbf{S}_{12}(n-1) - \dfrac{\mu_2}{\sigma_u^2(M-2)}\mathbf{S}_{12}(n-1)\mathbf{R} + \dfrac{\mu_1\mu_2}{\sigma_u^4(M-2)(M-4)} \times [2\mathbf{R}\mathbf{S}_{12}(n-1)\mathbf{R} + \mathbf{R}\mathrm{Tr}(\mathbf{R}\mathbf{S}_{12}(n-1)) + \sigma_v^2\mathbf{R}] + \mathbf{Q}$
μ_1-CMA and μ_2-CMA	$\mathbf{S}_{12}(n) \approx \mathbf{S}_{12}(n-1) - \mu_1\bar{\gamma}\mathbf{R}\mathbf{S}_{12}(n-1) - \mu_2\bar{\gamma}\mathbf{S}_{12}(n-1)\mathbf{R} + \mu_1\mu_2[2\xi\mathbf{R}\mathbf{S}_{12}(n-1)\mathbf{R} + \xi\mathbf{R}\mathrm{Tr}(\mathbf{R}\mathbf{S}_{12}(n-1)) + \sigma_\beta^2\mathbf{R}] + \mathbf{Q}$

Expressions for $\mathrm{E}\{\eta(n)\}$ and $\mathrm{E}\{\eta^2(n)\}$ depend on the mixing parameter adaptation. In the next section, we assume that $\eta(n)$ is updated with the η-LMS algorithm.

A. Adaptation of the mixing parameter using η-LMS

Replacing (17) in (15), we get

$$e_g(n) = e_{a,2}(n) - \eta(n)[e_{a,2}(n) - e_{a,1}(n)] + b(n). \quad (34)$$

Using (34) and remarking that $y_1(n) - y_2(n) = e_{a,2}(n) - e_{a,1}(n)$, the update equation of η-LMS, given by (14), can be rewritten as

$$\eta(n+1) = \overbrace{\eta(n)\left(1 - \mu_\eta[e_{a,2}(n) - e_{a,1}(n)]^2\right)}^{\mathcal{A}} + \overbrace{\mu_\eta[e_{a,2}^2(n) - e_{a,1}(n)e_{a,2}(n)]}^{\mathcal{B}} + \overbrace{\mu_\eta b(n)[e_{a,2}(n) - e_{a,1}(n)]}^{\mathcal{C}}. \quad (35)$$

Using (35), we can obtain recursions for the first and the second moments of $\eta(n)$.

1) First-order analysis: Using the same arguments of Section III, we remark that $\mathrm{E}\{\mathcal{C}\} = 0$. Assuming A1 and taking expectations in (35), we get

$$\boxed{\mathrm{E}\{\eta(n+1)\} = \mathrm{E}\{\eta(n)\}[1 - \mu_\eta\alpha(n)] + \mu_\eta\Delta\zeta_{22}(n). \quad (36)}$$

Since the constraint $\eta(n) \leq 1$ is imposed in the η-LMS algorithm, we truncate at each iteration the theoretical value of $\mathrm{E}\{\eta(n+1)\}$ estimated by (36), so that $\mathrm{E}\{\eta(n+1)\} \leq 1$.

Taking the limit for $n \to \infty$ on both sides of (36), we obtain

$$\lim_{n\to\infty} \mathrm{E}\{\eta(n)\} = \eta_o(\infty). \quad (37)$$

Thus, as observed in [12], the η-LMS algorithm converges in the average to the optimum mixing parameter at the steady-state.

A sufficient condition for the exponential stability of (36) is given by [38, p. 73]

$$|1-\mu_\eta\alpha(n)| < 1-\varepsilon, \ \ \forall n, \tag{38}$$

where ε is a small positive constant. In particular, for a constant step-size, a sufficient condition is

$$0 < \mu_\eta < \frac{2}{\max\{\alpha(n)\}}. \tag{39}$$

2) Second-order analysis: Squaring (35) and taking expectations, we obtain

$$\begin{aligned}\mathrm{E}\big\{\eta^2(n+1)\big\} = \mathrm{E}\big\{\mathcal{A}^2\big\}+\mathrm{E}\big\{\mathcal{B}^2\big\}+\mathrm{E}\big\{\mathcal{C}^2\big\}+\mathrm{E}\{2\mathcal{A}\mathcal{B}\}\\ +\,\mathrm{E}\{2\mathcal{A}\mathcal{C}\}+\mathrm{E}\{2\mathcal{B}\mathcal{C}\}.\end{aligned} \tag{40}$$

To evaluate the terms of (40), we assume that

A2. The *a priori* errors $e_{a,1}(n)$ and $e_{a,2}(n)$ are jointly Gaussian with zero-mean, which implies [39]

$$\mathrm{E}\left\{e^3_{a,i}(n)e_{a,j}(n)\right\} = 3\zeta_{ii}(n)\zeta_{ij}(n), \ \ i,j=1,2, \tag{41}$$

$$\mathrm{E}\left\{[e_{a,2}(n)-e_{a,1}(n)]^4\right\} = 3\alpha^2(n), \tag{42}$$

$$\mathrm{E}\left\{e^2_{a,1}(n)e^2_{a,2}(n)\right\} = \zeta_{11}(n)\zeta_{22}(n)+2\zeta^2_{12}(n). \tag{43}$$

Although this condition is violated in general, it is frequently used to make the transient analysis of adaptive filters more tractable [16]–[19]. This assumption tends to be reasonable for small step-sizes and long filters [17].

Now, using A1 and A2 we can evaluate the terms of (40):

$\mathrm{E}\{\mathcal{A}^2\}$: Using A1 and (42), we obtain

$$\begin{aligned}\mathrm{E}\{\mathcal{A}^2\} &= \mathrm{E}\left\{\eta^2(n)\big(1-\mu_\eta[e_{a,2}(n)-e_{a,1}(n)]^2\big)^2\right\}\\ &\approx \mathrm{E}\{\eta^2(n)\}\left[1-2\mu_\eta\alpha(n)+3\mu_\eta^2\alpha^2(n)\right].\end{aligned} \tag{44}$$

$\mathrm{E}\{\mathcal{B}^2\}$: Using (41) and (43), we have

$$\begin{aligned}\mathrm{E}\{\mathcal{B}^2\} &= \mu^2_\eta\mathrm{E}\left\{[e^2_{a,2}(n)-e_{a,1}(n)e_{a,2}(n)]^2\right\}\\ &\approx \mu^2_\eta\zeta_{22}(n)\alpha(n)+2\mu^2_\eta\Delta\zeta^2_{22}(n).\end{aligned} \tag{45}$$

$\mathrm{E}\{\mathcal{C}^2\}$: Since $b(n)$ is assumed independent of $e_{a,i}(n)$, $i=1,2$, $\mathrm{E}\{b(n)\}=0$, and $\mathrm{E}\{b^2(n)\}=\sigma^2_b$, we get

$$\begin{aligned}\mathrm{E}\left\{\mathcal{C}^2\right\} &= \mathrm{E}\big\{\mu^2_\eta b^2(n)[e_{a,2}(n)-e_{a,1}(n)]^2\big\}\\ &\approx \mu^2_\eta\sigma^2_b\alpha(n).\end{aligned} \tag{46}$$

For the combination of two CMA equalizers this term is null, since $b(n)\equiv 0$.

$\mathrm{E}\{2\mathcal{A}\mathcal{B}\}$: Using A1, (41) and (43), we obtain

$$\begin{aligned}\mathrm{E}\{2\mathcal{A}\mathcal{B}\} = 2\mu_\eta\mathrm{E}\{\eta(n)\}\mathrm{E}\Big\{&\big(1-\mu_\eta[e_{a,2}(n)-e_{a,1}(n)]^2\big)\\ &\times[e^2_{a,2}(n)-e_{a,1}(n)e_{a,2}(n)]\Big\}\\ \approx 2\mu_\eta\mathrm{E}\{\eta(n)\}&\left[\zeta_{22}(n)-3\mu_\eta\Delta\zeta_{22}(n)\alpha(n)\right].\end{aligned} \tag{47}$$

$\mathrm{E}\{2\mathcal{A}\mathcal{C}\}$ and $\mathrm{E}\{2\mathcal{B}\mathcal{C}\}$: Since $b(n)$ is assumed independent of $e_{a,i}(n)$, $i=1,2$ and $\mathrm{E}\{b(n)\}=0$, these terms are null. Again, for the combination of two CMA equalizers these terms are null by definition, since $b(n)\equiv 0$.

Replacing the approximations (44)-(47) in (40), we finally arrive at

$$\boxed{\begin{aligned}\mathrm{E}\{\eta^2(n+1)\} \approx \mathrm{E}\{\eta^2(n)\}&\left[1-2\mu_\eta\alpha(n)+3\mu^2_\eta\alpha^2(n)\right]\\ &+2\mu_\eta\mathrm{E}\{\eta(n)\}\left[\zeta_{22}(n)-3\mu_\eta\Delta\zeta_{22}(n)\alpha(n)\right]\\ &+\mu^2_\eta\left(\zeta_{22}(n)+\sigma^2_b\right)\alpha(n)+2\mu^2_\eta\Delta\zeta^2_{22}(n).\end{aligned}} \tag{48}$$

Using (48) and (36) in conjunction with the expressions of Table V, the EMSE of the combination for all $n\geq 0$ considering the η-LMS algorithm can be estimated via (30).

From (48), the range of step-sizes to ensure the mean-square stability of η-LMS is given by [38]

$$0 < \mu_\eta < \frac{2}{3\max\{\alpha(n)\}}, \tag{49}$$

which is more restrictive than (39).

The stability of η-LMS depends on $\alpha(n)$. From Fig. 1, we can see that $\alpha(n)=\mathrm{E}\{[y_1(n)-y_2(n)]^2\}$ is large at first when the fast filter has almost converged but the slow filter is still far from the optimum solution. At this point, μ_η should be small, as required by (49). However, when the EMSE of the slow and fast filters are similar, $\alpha(n)$ is small. At this point, a large μ_η is required so the combination will switch to the slow filter. This is the reason why [12] needs to constrain $\eta(n)\leq 1$. To guarantee that η-LMS switches quickly to the slow filter at the proper time, μ_η must be chosen so large that η-LMS will be unstable at the beginning, when $\alpha(n)$ is large. Therefore, some sort of normalization is necessary for the estimation of η. Thus, we propose in the following sections two normalized algorithms to update the mixing parameter.

B. Adaptation of the mixing parameter using η-PN-LMS

Using an instantaneous normalization, i.e., replacing the step-size by $\mu_\eta(n)=\tilde{\mu}_\eta/[y_1(n)-y_2(n)]^2$, can also lead to divergence (see, e.g, [40]). One possible solution is to normalize the algorithm using an estimate of $\alpha(n)$, as in [9]. The resulting normalized algorithm is called power normalized least mean-square (η-PN-LMS) algorithm and updates the mixing parameter via the recursion

$$\eta(n+1) = \eta(n)+\mu_\eta(n)e_g(n)[y_1(n)-y_2(n)] \tag{50}$$

where

$$\mu_\eta(n) \triangleq \frac{\widetilde{\mu}_\eta}{\epsilon+p(n)}, \tag{51}$$

$$p(n) = \lambda p(n-1)+(1-\lambda)[y_1(n)-y_2(n)]^2 \tag{52}$$

is a low-pass filtered estimate for the power of $y_1(n)-y_2(n)$, ϵ is a small positive constant used to avoid large step-sizes when $p(n)$ becomes small, and $0\ll\lambda<1$ is a forgetting factor. The stability of (50) is ensured for $0<\widetilde{\mu}_\eta<2$ [38] and no constraint on $\eta(n)$ is necessary.

In the analysis of the η-PN-LMS algorithm, we assume that

A3. The forgetting factor λ is sufficiently close to one, so that the variance of $p(n)$ is small and the step-size $\mu_\eta(n)$ is weakly correlated with the *a priori* errors $e_{a,i}(n)$, $i=1,2$ and the mixing parameter $\eta(n)$.

Using A3, the analysis of η-LMS can be directly extended to η-PN-LMS, replacing μ_η^k by $[\mathrm{E}\{\mu_\eta(n)\}]^k$, $k=1,2$ in the expressions of Section V-A. Hence, we only need to estimate $\mathrm{E}\{\mu_\eta(n)\}$, as shown in the sequel.

Expanding $\mu_\eta(n)$ as a Taylor series, around the expected value $\mathrm{E}\{p(n)\} \triangleq \bar{p}(n)$, we obtain

$$\mu_\eta(n) \approx \frac{\widetilde{\mu}_\eta}{\epsilon+\bar{p}(n)} - \frac{\widetilde{\mu}_\eta\,[p(n)-\bar{p}(n)]}{[\epsilon+\bar{p}(n)]^2} + \frac{\widetilde{\mu}_\eta\,[p(n)-\bar{p}(n)]^2}{[\epsilon+\bar{p}(n)]^3}. \quad (53)$$

Taking expectations on both sides of (53), we arrive at

$$\mathrm{E}\{\mu_\eta(n)\} \approx \frac{\widetilde{\mu}_\eta}{\epsilon+\bar{p}(n)} + \frac{\widetilde{\mu}_\eta \sigma_p^2(n)}{[\epsilon+\bar{p}(n)]^3}, \quad (54)$$

where we denoted $\sigma_p^2(n) = \mathrm{E}\{[p(n)-\bar{p}(n)]^2\}$. Assuming A3, the second term on the r.h.s. of (54) can be disregarded, which leads to

$$\mathrm{E}\{\mu_\eta(n)\} \approx \frac{\widetilde{\mu}_\eta}{\epsilon+\bar{p}(n)}. \quad (55)$$

Using the same arguments, the second moment of the step-size $\mu_\eta(n)$ can be approximated by $\mathrm{E}\{\mu_\eta^2(n)\} \approx [\mathrm{E}\{\mu_\eta(n)\}]^2$.

Now, we obtain a recursion for $\bar{p}(n)$. Taking expectations on both sides of (52), we get

$$\bar{p}(n) = \lambda\bar{p}(n-1) + (1-\lambda)\mathrm{E}\left\{[y_1(n)-y_2(n)]^2\right\}. \quad (56)$$

Remarking that $y_1(n)-y_2(n) = e_{a,2}(n)-e_{a,1}(n)$, the following recursion holds

$$\bar{p}(n) = \lambda\bar{p}(n-1) + (1-\lambda)\alpha(n). \quad (57)$$

At steady-state, we have $\bar{p}(\infty) = \alpha(\infty)$.

C. Adaptation of the mixing parameter using η-SR-LMS

Although η-PN-LMS circumvents the problem encountered in the convergence of η-LMS, three parameters must be adjusted: $\widetilde{\mu}_\eta$, λ, and ϵ. The forgetting factor λ is relatively easy to be adjusted (e.g., $\lambda = 0.99$). However, the choice of the step-size $\widetilde{\mu}_\eta$ and of the regularization factor ϵ needs some care, as we show through the simulations of Section VI. In order to avoid these extra adjustments and since a normalization is necessary, we can employ a partial instantaneous normalization using $\mu_\eta(n) = \mu_{\eta s}/|y_1(n)-y_2(n)|$ as step-size. With this choice, the update rule (50) reduces to

$$\eta(n+1) = \eta(n) + \mu_{\eta s} e_g(n)\,\mathrm{sign}\,[y_1(n)-y_2(n)], \quad (58)$$

where $\mathrm{sign}[\cdot]$ is the sign function defined as

$$\mathrm{sign}[x] \triangleq \begin{cases} +1, & x>0 \\ 0, & x=0 \\ -1, & x<0 \end{cases}. \quad (59)$$

We call this algorithm sign regressor least mean-square algorithm (η-SR-LMS).

Using (34) and remarking that $x\,\mathrm{sign}[x] = |x|$ and that $y_1(n)-y_2(n) = e_{a,2}(n)-e_{a,1}(n)$, (58) can be rewritten as

$$\begin{aligned}\eta(n+1) = {} & \overbrace{\eta(n)\,(1-\mu_{\eta s}|e_{a,2}(n)-e_{a,1}(n)|)}^{\mathcal{D}} \\ & + \overbrace{\mu_{\eta s} e_{a,2}(n)\mathrm{sign}\,[e_{a,2}(n)-e_{a,1}(n)]}^{\mathcal{E}} \\ & + \overbrace{\mu_{\eta s} b(n)\mathrm{sign}\,[e_{a,2}(n)-e_{a,1}(n)]}^{\mathcal{F}}. \end{aligned} \quad (60)$$

Using (60), we can obtain recursions for the first and second moments of $\eta(n)$.

1) First-order analysis: Assuming A1, taking expectations in (60), and remarking that $\mathrm{E}\{\mathcal{F}\} = 0$, we obtain

$$\begin{aligned}\mathrm{E}\{\eta(n+1)\} = {} & \mathrm{E}\{\eta(n)\}\big(1-\mu_{\eta s}\mathrm{E}\{|e_{a,2}(n)-e_{a,1}(n)|\}\big) \\ & + \mu_{\eta s}\mathrm{E}\{e_{a,2}(n)\mathrm{sign}\,[e_{a,2}(n)-e_{a,1}(n)]\}. \end{aligned} \quad (61)$$

Assuming A2 and using a special case of Price's theorem (see, e.g, [39], [17, p. 306]), the following approximations hold

$$\mathrm{E}\{|e_{a,2}(n)-e_{a,1}(n)|\} \approx \sqrt{\frac{2\alpha(n)}{\pi}}, \quad (62)$$

and

$$\mathrm{E}\{e_{a,2}(n)\mathrm{sign}\,[e_{a,2}(n)-e_{a,1}(n)]\} \approx \frac{\Delta\zeta_{22}(n)}{\sqrt{\pi\alpha(n)/2}}. \quad (63)$$

Replacing (62) and (63) in (61), we arrive at

$$\boxed{\mathrm{E}\{\eta(n+1)\} \approx \mathrm{E}\{\eta(n)\}\left[1-\mu_{\eta s}\sqrt{\frac{2\alpha(n)}{\pi}}\right] + \mu_{\eta s}\frac{\Delta\zeta_{22}(n)}{\sqrt{\pi\alpha(n)/2}}.} \quad (64)$$

Taking the limit for $n\to\infty$ on both sides of (64), we obtain $\lim_{n\to\infty}\mathrm{E}\{\eta(n)\} = \eta_{\mathrm{o}}(\infty)$. Hence, the η-SR-LMS algorithm also converges in the average to the optimum mixing parameter at the steady-state.

The range of step-sizes that guarantees stability of (64) is given by [38]

$$0 < \mu_{\eta s} < \sqrt{\frac{2\pi}{\max\{\alpha(n)\}}}. \quad (65)$$

2) Second-order analysis: Squaring (60) and taking expectations, we obtain

$$\begin{aligned}\mathrm{E}\{\eta^2(n+1)\} = {} & \mathrm{E}\{\mathcal{D}^2\} + \mathrm{E}\{\mathcal{E}^2\} + \mathrm{E}\{\mathcal{F}^2\} + \mathrm{E}\{2\mathcal{D}\mathcal{E}\} \\ & + \mathrm{E}\{2\mathcal{D}\mathcal{F}\} + \mathrm{E}\{2\mathcal{E}\mathcal{F}\}. \end{aligned} \quad (66)$$

Using A1 and A2, we can evaluate the terms of (66):

$\mathrm{E}\{\mathcal{D}^2\}$: Using A1 and (62), we obtain

$$\begin{aligned}\mathrm{E}\{\mathcal{D}^2\} &= \mathrm{E}\left\{\eta^2(n)\left[1-\mu_{\eta s}|e_{a,2}(n)-e_{a,1}(n)|\right]^2\right\} \\ &\approx \mathrm{E}\{\eta^2(n)\}\left[1-\mu_{\eta s}\sqrt{8\alpha(n)/\pi} + \mu_{\eta s}^2\alpha(n)\right]. \end{aligned} \quad (67)$$

$\mathrm{E}\{\mathcal{E}^2\}$ and $\mathrm{E}\{\mathcal{F}^2\}$: Using the fact that $\mathrm{sign}^2[x] = 1$ almost everywhere on the real line, we get

$$\mathrm{E}\{\mathcal{E}^2\} = \mathrm{E}\big\{\mu_{\eta s}^2 e_{a,2}^2(n)\mathrm{sign}^2\left[e_{a,2}(n) - e_{a,1}(n)\right]\big\} \approx \mu_{\eta s}^2 \zeta_{22}(n), \quad (68)$$

and

$$\mathrm{E}\left\{\mathcal{F}^2\right\} = \mathrm{E}\big\{\mu_{\eta s}^2 b^2(n)\mathrm{sign}^2\left[e_{a,2}(n) - e_{a,1}(n)\right]\big\} \approx \mu_{\eta s}^2 \sigma_b^2. \quad (69)$$

$\mathrm{E}\{2\mathcal{D}\mathcal{E}\}$: Using A1 and (63), we obtain

$$\begin{aligned}\mathrm{E}\{2\mathcal{D}\mathcal{E}\} &= 2\mu_{\eta s}\mathrm{E}\Big\{\eta(n)\big(1-\mu_{\eta s}|e_{a,2}(n)-e_{a,1}(n)|\big) \\ &\qquad \times e_{a,2}(n)\mathrm{sign}\left[e_{a,2}(n)-e_{a,1}(n)\right]\Big\} \\ &\approx 2\mu_{\eta s}\mathrm{E}\{\eta(n)\}\left[\frac{\Delta\zeta_{22}(n)}{\sqrt{\pi\alpha(n)/2}} - \mu_{\eta s}\Delta\zeta_{22}(n)\right]. \quad (70)\end{aligned}$$

Since $b(n)$ is assumed independent of $e_{a,i}(n)$, $i = 1, 2$ and $\mathrm{E}\{b(n)\} = 0$, we have $\mathrm{E}\{2\mathcal{D}\mathcal{F}\} \approx 0$ and $\mathrm{E}\{2\mathcal{E}\mathcal{F}\} \approx 0$. Replacing the approximations (67)-(70) in (66), we finally arrive at

$$\boxed{\begin{aligned}\mathrm{E}\{\eta^2(n+1)\} &\approx \mathrm{E}\{\eta^2(n)\}\left[1-\mu_{\eta s}\sqrt{\frac{8\alpha(n)}{\pi}}+\mu_{\eta s}^2\alpha(n)\right] \\ &+ 2\mu_{\eta s}\mathrm{E}\{\eta(n)\}\left[\frac{\Delta\zeta_{22}(n)}{\sqrt{\pi\alpha(n)/2}} - \mu_{\eta s}\Delta\zeta_{22}(n)\right] \\ &+ \mu_{\eta s}^2\left[\sigma_b^2 + \zeta_{22}(n)\right]. \quad (71)\end{aligned}}$$

The range of step-sizes that guarantees stability of (71) is given by [38]

$$0 < \mu_{\eta s} < \sqrt{\frac{8}{\pi \max\{\alpha(n)\}}}. \quad (72)$$

It should be noticed that this range is more restrictive than (65). Although the step size $\mu_{\eta s}$ still depends on an estimate of $\alpha(n)$, this dependence is weaker than that of the η-LMS algorithm due to the square-root in (72). Furthermore, $\mu_{\eta s}$ can be adjusted based on the analytical EMSE of the combination (see Fig. 11). Thus, the η-SR-LMS algorithm can perform better than η-LMS, following quickly the variations on $\eta_{\mathrm{o}}(n)$ with a small EMSE and with only one free parameter to adjust, as shown in the simulations of Section VI-A.

VI. Simulation results

The simulations are divided into four parts. First, we verify the accuracy of the transient analysis for the introductory simulations shown in Fig. 1. We also verify the behavior of the proposed algorithms η-PN-LMS and η-SR-LMS in the same simulation scenario. In the second part, we show some results concerning the analysis of combinations of NLMS filters and CMA equalizers. In the third part, we verify the validity of the analysis of combinations of LMS filters with close step-sizes. Finally, we focus on the tracking analysis and compare the performances of the affine and convex combinations.

A. Recalling the introductory simulation

To verify the validity of the transient analysis in the supervised case, we consider the identification of a time-invariant system. The optimum solution is formed with $M = 7$ independent random values between -1 an 1, and is given by

$$\mathbf{w}_{\mathrm{o}} = [+0.90 \;\; -0.54 \;\; -0.03 \;\; +0.78 \;\; +0.52 \;\; -0.09]. \quad (73)$$

We assume white Gaussian input with variance $1/M$ so that $\mathrm{Tr}(\mathbf{R}) = 1$, and an average of 500 runs. Moreover, i.i.d. noise $v(n)$ with variance $\sigma_v^2 = 0.01$ is added to form the desired signal.

Figures 5 and 6 show the results of the EMSE and the mixing parameter for the affine combination of two LMS filters in the same situations considered in Figures 1-(a) and (b), in which the mixing parameter is updated with the η-LMS algorithm. In Fig. 5, where $\mu_\eta = 3$, the analysis can predict that the performance of the combination is far from universal in the initial iterations. Similarly, with $\mu_\eta = 0.1$, the analysis can predict that the combination is not able to switch to the slow filter, as shown in Fig. 6. We should notice that, due to the constraint imposed in the η-LMS algorithm ($\eta(n) \leq 1$), these situations become difficult to model and there is a small gap between the experimental and theoretical EMSE during the initial iterations. Moreover, the mixing parameter does not achieve the optimum value obtained in the analysis, which is higher than one in the initial iterations.

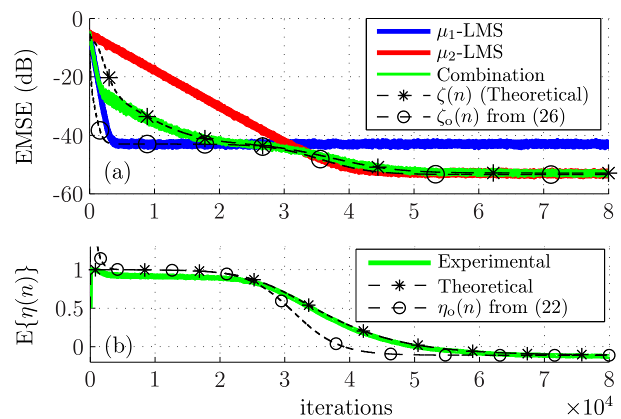

Fig. 5. a) EMSE for μ_1-LMS, μ_2-LMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-LMS algorithm and theoretical $\eta_{\mathrm{o}}(n)$; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $\mu_\eta = 3$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs.

In the same scenario, the algorithms η-PN-LMS or η-SR-LMS can circumvent the problem, as shown in Figures 7 and 8 respectively. These two algorithms have a similar performance, which is predicted by the analysis with a good accuracy in both cases. In addition, the experimental mixing parameter is higher than one in the initial iterations, being far from its theoretical optimum value during the very first iterations, as shown in Figures 7-(c) and 8-(c). However, this does not represent an issue since the combination presents a close to universal performance.

To illustrate the influence of the parameters ϵ and $\widetilde{\mu}_\eta$ in the performance of the η-PN-LMS algorithm, Figures 9 and 10

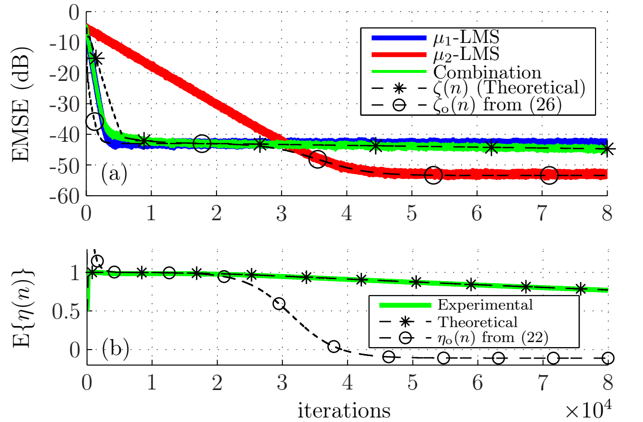

Fig. 6. a) EMSE for μ_1-LMS, μ_2-LMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $\mu_\eta = 0.1$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs.

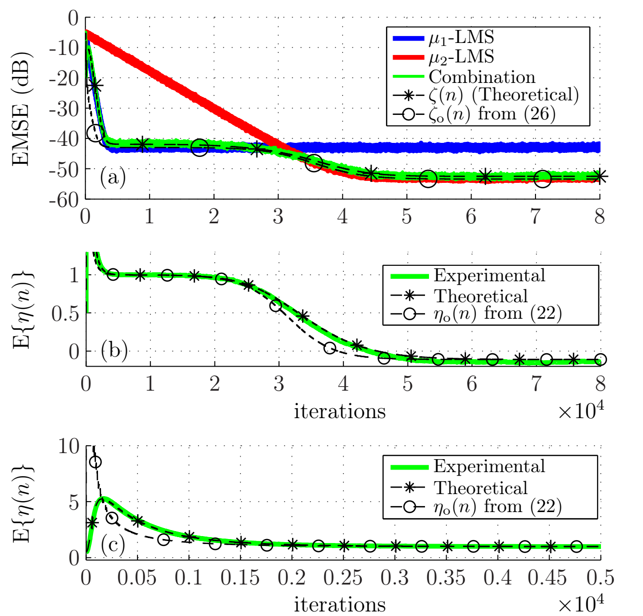

Fig. 7. a) EMSE for μ_1-LMS, μ_2-LMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-PN-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $\widetilde{\mu}_\eta = 3\times 10^{-3}$, $\epsilon = 5\times 10^{-4}$, $\lambda = 0.99$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs; c) detail of b) from $n = 0$ until $n = 0.5\times 10^4$ (note the different x-scaling).

show the theoretical, experimental, and optimal EMSE of the combination at three time instants as a function of ϵ (Fig. 9) and of $\widetilde{\mu}_\eta$ (Fig. 10). The time instants were chosen in order to check the accuracy of the analysis in three different situations: at $n = 15\times 10^3$ the slower filter has not converged yet, the time instant $n = 40\times 10^3$ is close to the switching between the faster to the slower filter, and at $n = 65\times 10^3$ both filters have converged. The same simulation setting of Fig. 7 is considered. Similarly, Fig. 11 shows the results on the influence of $\mu_{\eta s}$ for the η-SR-LMS algorithm, considering the same simulation setting of Fig. 8. The analysis provides an accurate estimation of the EMSE in all cases, which enables the adjustment of the parameters through the analytical results. We can also observe that the optimum value of ϵ, $\widetilde{\mu}_\eta$, or $\mu_{\eta s}$ is different for each time instant considered in the simulations. However, it is possible to choose an intermediate value to obtain a tradeoff in these three situations.

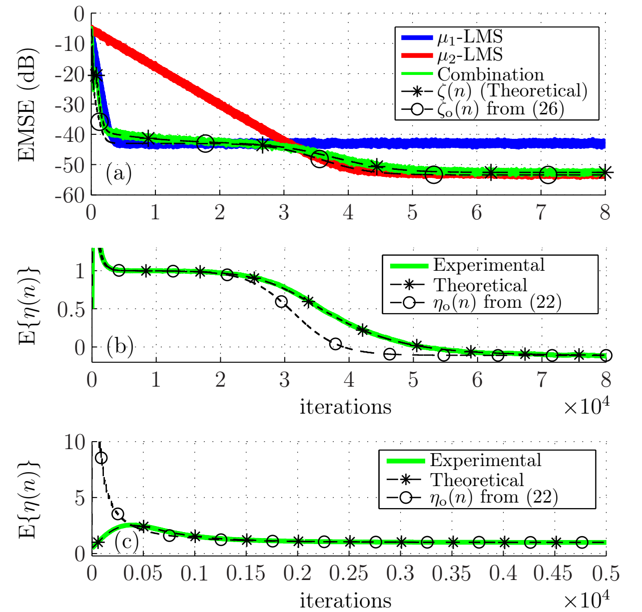

Fig. 8. a) EMSE for μ_1-LMS, μ_2-LMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-SR-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $\mu_{\eta s} = 2.5\times 10^{-2}$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs; c) detail of b) from $n = 0$ until $n = 0.5\times 10^4$ (note the different x-scaling).

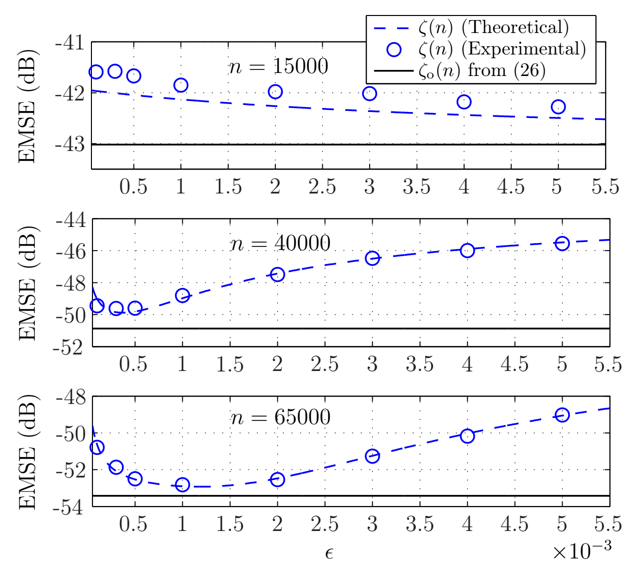

Fig. 9. Theoretical, experimental and optimal EMSE at three different time instants for the affine combination of μ_1-LMS and μ_2-LMS using the η-PN-LMS for different values of ϵ; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $\widetilde{\mu}_\eta = 3$, $\lambda = 0.99$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs; each experimental value was calculated by the mean EMSE of 50 samples around the considered time instant.

B. Combinations of two NLMS filters and two CMA equalizers

To verify that the transient analysis is also accurate for the affine combination of the other algorithms, Fig. 12 and

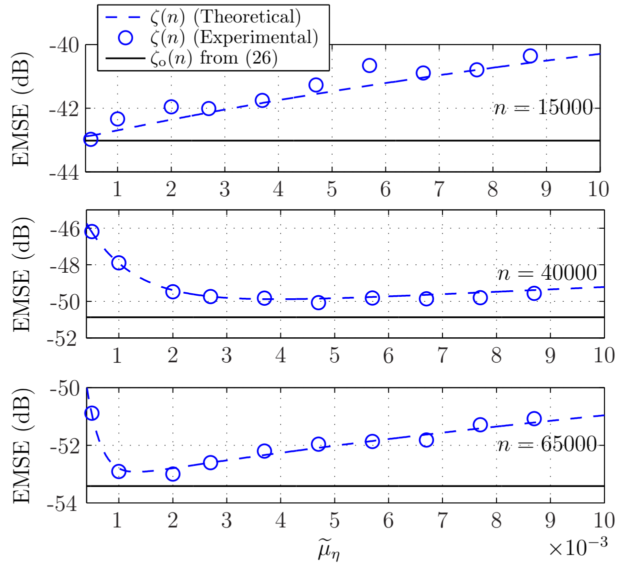

Fig. 10. Theoretical, experimental and optimal EMSE at three different time instants for the affine combination of μ_1-LMS and μ_2-LMS using the η-PN-LMS for different values of $\widetilde{\mu}_\eta$; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $\epsilon = 5\times10^{-4}$, $\lambda = 0.99$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs; each experimental value was calculated by the mean EMSE of 50 samples around the considered time instant.

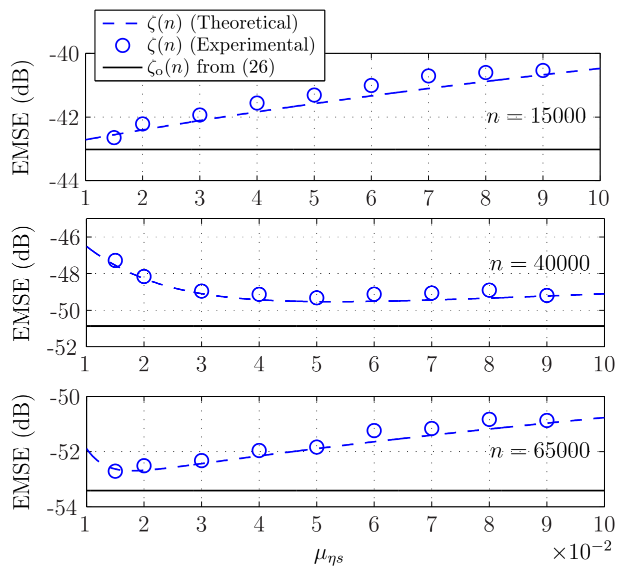

Fig. 11. Theoretical, experimental and optimal EMSE at three different time instants for the affine combination of μ_1-LMS and μ_2-LMS using the η-SR-LMS for different values of $\mu_{\eta s}$; $\mu_1 = 0.01$, $\mu_2 = 0.001$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/7$; 500 independent runs; each experimental value was calculated by the mean EMSE of 50 samples around the considered time instant.

Fig. 13 show the results for combination of two NLMS filters with the η-PN-LMS algorithm and two CMA equalizers with the η-SR-LMS algorithm, respectively. For the NLMS case, to obtain a better estimate for the EMSE of the component filters using the expression of Table V, we consider $M = 32$ coefficients and the optimum solution $(\mathbf{w}_o)$ from [12, Fig. 2]. Again, we can observe a good agreement between analysis and simulation. In the CMA case, we assume the channels $\mathbf{h}_1 = [+0.1\ \ +0.3\ \ +1.0\ \ -0.1\ \ +0.5\ \ +0.2]^T$ and $\mathbf{h}_2 = [+0.25\ \ +0.64\ \ +0.80\ \ -0.55]^T$ [27], [33] in the absence of noise and the transmission of a 4-PAM (pulse amplitude modulation) signal, i.e., $a(n) = \pm1$ or $a(n) = \pm3$, with statistics $r = 8.2$, $\sigma_\beta^2 = 28.8$, and $\bar{\gamma} = 6.8$. In the combination, each component filter has $M = 4$ coefficients implemented as a T/2-fractionally spaced equalizer (FSE) and is initialized with only one non-null and unitary element in the second position. Fig. 13 shows the results for the EMSE and the mixing parameter considering the channel $\mathbf{h}_1$ until $n = 4\times10^4$ and the channel $\mathbf{h}_2$ after that. To smooth the EMSE curves, they were filtered by a moving-average filter of 32 coefficients. Although there is no exact agreement between analysis and simulation, the predicted values model the overall behavior of the combination, considering that a difference of a few dB is common in models of blind algorithms due to the strong assumptions necessary for the analysis.

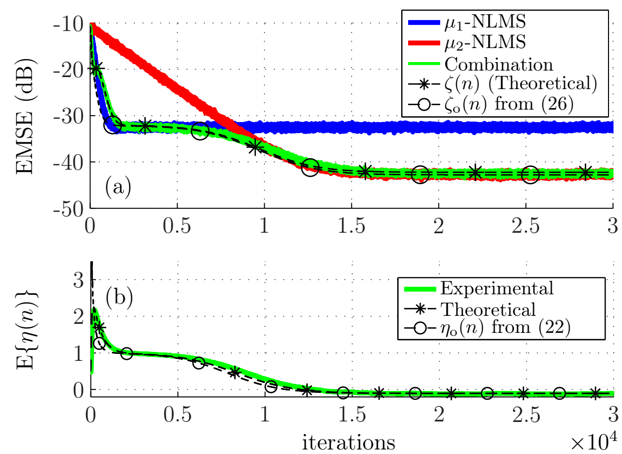

Fig. 12. a) EMSE for μ_1-NLMS, μ_2-NLMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-PN-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 0.1$, $\mu_2 = 0.01$, $\widetilde{\mu}_\eta = 3\times10^{-3}$, $\epsilon = 5\times10^{-4}$, $\lambda = 0.99$, $M = 32$; identification of the system considered in [12], $\sigma_v^2 = 0.01$, white input with variance $\sigma_u^2 = 1/32$, $\mathbf{Q} = \mathbf{0}$; 500 independent runs.

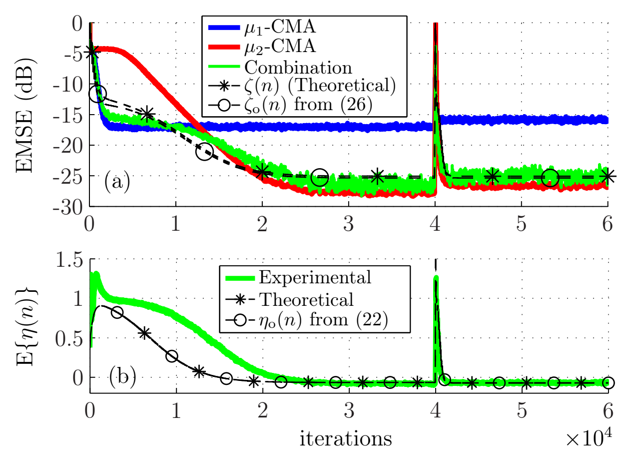

Fig. 13. a) EMSE for μ_1-CMA, μ_2-CMA and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-SR-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 1\times10^{-3}$, $\mu_2 = 1\times10^{-4}$, $\mu_{\eta s} = 0.5$; Equalizers with $M = 4$ as T/2-FSE, initialized with $[0\ 1\ 0\ 0]^T$; channel $\mathbf{h}_1 = [0.1\ 0.3\ 1.0\ -0.1\ 0.5\ 0.2]^T$ until $n = 4\times10^4$ and $\mathbf{h}_2 = [0.25\ 0.64\ 0.80\ -0.55]^T$ after $n = 4\times10^4$; $\mathbf{Q} = \mathbf{0}$; 4-PAM transmitted signal; 500 independent runs; EMSE curves filtered by a moving-average filter of 32 coefficients.

C. Affine combination of filters with close step-sizes

We now consider an affine combination of two LMS filters with close step-sizes in a stationary environment. We assume that the optimum solution is given by (73) and the input $u(n)$ is generated using a first-order autoregressive model, whose transfer function is $\sqrt{1-\varrho^2}/(1-\varrho z^{-1})$, with $\varrho = 0.8$. This model is fed with and i.i.d. Gaussian random process, whose variance is $1/M$, such that $\mathrm{Tr}(\mathbf{R}) = 1$. Again, to form the desired signal, white noise $v(n)$ with variance $\sigma_v^2 = 0.01$ is added. Fig. 14 shows the EMSE and the mixing parameter along the iterations for two LMS filters with step-sizes $\mu_1 = 0.01$ and $\mu_2 = 0.009$, using the η-LMS algorithm with $\mu_\eta = 600$. This high value of μ_η is needed in order to ensure a high convergence rate for the combination since $[y_1(n)-y_2(n)]$ is small. In this situation, the performances of the component filters are very close and the combination provides a 3 dB EMSE gain in steady-state, as shown in Fig. 14-(a) and predicted by the analysis. To smooth the EMSE curves, they were filtered by a moving-average filter with 256 coefficients. We can observe that, due to the constraint ($\eta(n) \leq 1$) imposed in the η-LMS algorithm, the mixing parameter does not achieve its optimum value, which may be close to 25 in some time instants, as shown in Fig. 14-(b). Consequently, the EMSE of the combination is far from the optimum EMSE in some time instants.

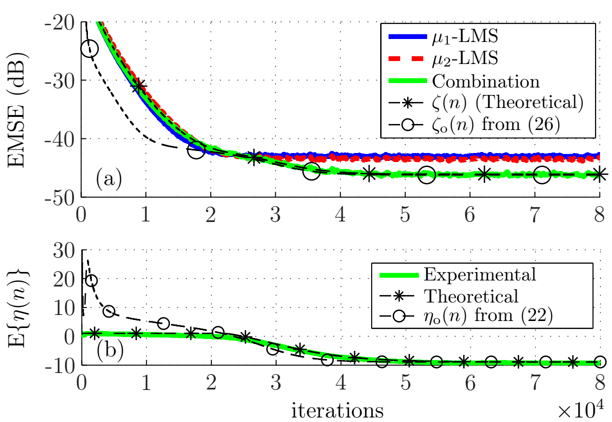

Fig. 14. a) EMSE for μ_1-LMS, μ_2-LMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 0.01$, $\mu_2 = 0.009$, $\mu_\eta = 600$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, colored input (AR model, 1^{st} order, pole at 0.8) with variance $\sigma_u^2 = 1/7$; 500 independent runs; EMSE curves filtered by a moving-average filter with 256 coefficients.

In the same scenario, the η-PN-LMS and η-SR-LMS algorithms circumvent the problem since no constraint in $\eta(n)$ is necessarily used, as show respectively in Figures 15 and 16. A 3 dB EMSE gain can be observed in steady-state and there is also an EMSE gain in the transient, being both well predicted by the analysis.

D. Accuracy of tracking analysis

To verify the validity of the tracking analysis, the affine combination is compared to the convex combination assuming two LMS filters with different step-sizes. The same simulation setting of Fig. 14 is considered, but with fixed $\mu_1 = 0.1$ and

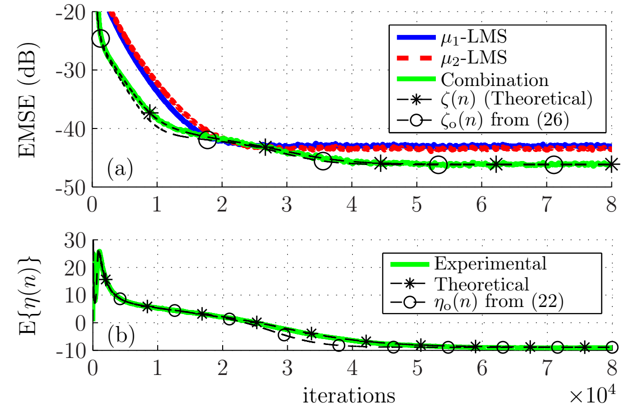

Fig. 15. a) EMSE for μ_1-LMS, μ_2-LMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-PN-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 0.01$, $\mu_2 = 0.009$, $\widetilde{\mu}_\eta = 0.4$, $\epsilon = 9 \times 10^{-4}$, $\lambda = 0.99$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, colored input (AR model, 1^{st} order, pole at 0.8) with variance $\sigma_u^2 = 1/7$; 500 independent runs; EMSE curves filtered by a moving-average filter with 256 coefficients.

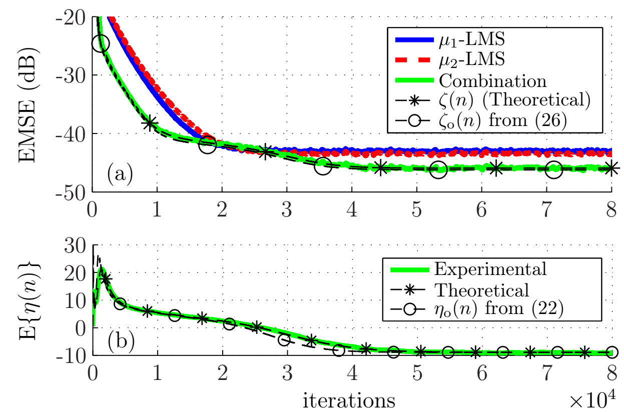

Fig. 16. a) EMSE for μ_1-LMS, μ_2-LMS and their affine combination; b) ensemble average of $\eta(n)$ adapted with the η-SR-LMS algorithm and theoretical $\eta_o(n)$; $\mu_1 = 0.01$, $\mu_2 = 0.009$, $\mu_{\eta s} = 0.5$, $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, colored input (AR model, 1^{st} order, pole at 0.8) with variance $\sigma_u^2 = 1/7$; 500 independent runs; EMSE curves filtered by a moving-average filter with 256 coefficients.

$\mu_2 = \delta\mu_1$. Fig. 17-(a) shows the theoretical and experimental values of $\zeta_{ii}(\infty)$, $i = 1, 2$ for the component filters and the values of $\zeta(\infty)$ for the affine and convex combinations as functions of δ, considering a nonstationary environment with $\mathbf{Q} = 4 \times 10^{-7}\mathbf{I}$. The ratio $\zeta(\infty)/\min\{\zeta_{ii}(\infty)\}$ is also shown in Fig. 17-(b). It can be noticed that there is an EMSE reduction for both the affine and convex combinations when $\mathrm{Tr}(\mathbf{Q}) = q_{12} = \mu_1\mu_2\sigma_v^2\mathrm{Tr}(\mathbf{R})$, which corresponds to $\delta = 0.025$ in this case. An EMSE reduction for the affine combination also occurs when $\mu_1 \approx \mu_2$, i.e., $\delta \approx 1$. In this case, the convex combination can only perform as its best component filter, since the mixing parameter needs to be negative to cause the EMSE reduction, as shown in Fig. 17-(c). In both cases, the reduction is limited to 3 dB, which agrees with the results of Table IV. The theoretical results for the convex combination were obtained truncating the value of the optimal mixing parameter to the interval $[0, 1]$. It is important to remark, though, that both points at which the largest EMSE reduction happens do not represent optimal situations, as can be seen in Fig. 17-(a). For a single LMS filter in a nonstationary environment, there is an optimum value of the step-size that minimizes the EMSE. The minimum EMSE value for the affine and convex combinations (≈ -38 dB) occurs exactly when μ_2 assumes this optimum value, which happens for $\delta = 0.17$ in this example. In this case, both combinations perform as their best component filter μ_2-LMS. Therefore, using the affine combination of filters of the same family updated with different step-sizes is more worthwhile than using it with close step-sizes.

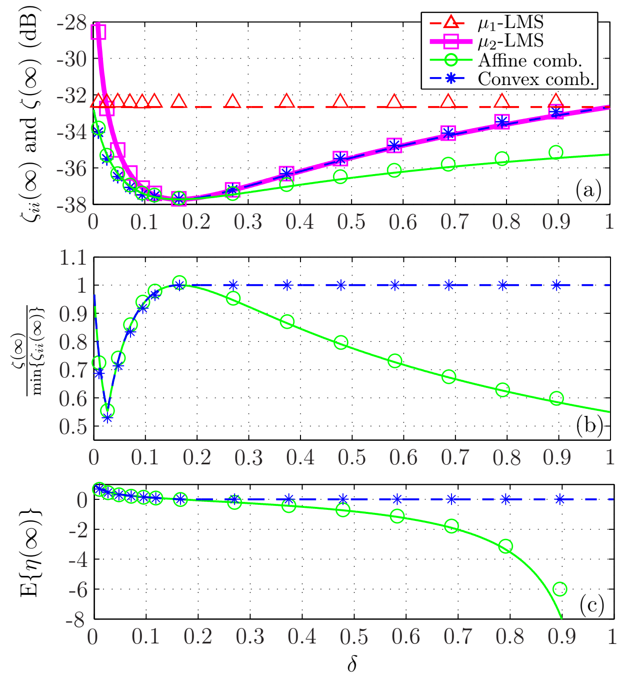

Fig. 17. Theoretical and experimental values of a) $\zeta_{ii}(\infty)$, $i = 1, 2$ and $\zeta(\infty)$; b) $\zeta(\infty)/\min\{\zeta_{ii}(\infty)\}$, $i = 1, 2$; and c) $\mathrm{E}\{\eta(\infty)\}$ for the affine ($\mu_\eta = 1$) and convex ($\mu_a = 100$, $a^+ = 4$ and $\epsilon = 0.1$ [8]) combinations of two LMS filters with $\mu_1 = 0.1$, $\mu_2 = \delta\mu_1$ and $M = 7$; identification of the system given by (73), $\sigma_v^2 = 0.01$, colored input (AR model, 1^{st} order, pole at 0.8) with variance $\sigma_u^2 = 1/7$; $\mathbf{Q} = 4\times10^{-7}\mathbf{I}$; 50 independent runs. The theoretical values are indicated by lines and the experimental values by $\triangle$, $\square$, $\bigcirc$, and $\ast$.

VII. Conclusion

As an extension of [12] and [13], we proposed transient and steady-state analyses for the EMSE and the mixing parameter of the affine combination, based on the theoretical EMSE and cross-EMSE of the component filters and on the adaptation of the mixing parameter. This allows the application to different combinations of algorithms, such as LMS, NLMS and CMA, considering white or colored inputs and stationary or nonstationary environments. Good agreement between the analysis and the simulations was always observed. Moreover, we proposed and analyzed two normalized algorithms for updating the mixing parameter. The theoretical models can predict situations in which these algorithms can achieve a better performance, being useful for the designer.

Appendix A
Assumptions for the CMA analysis

Model (8) is based on the following assumption:

B1. The channel noise power is small enough for the zero-forcing solution $\mathbf{w}_\mathrm{o}$ to be one of the global minimizers of the constant-modulus cost function. In other words, the optimal solution achieves perfect equalization, i.e., $a(n-\tau_d) \approx \mathbf{u}^T(n)\mathbf{w}_\mathrm{o}(n-1)$ [10], [23], [26], [33].

Using B1, the filter output can be approximated by

$$y_i(n) \approx a(n-\tau_d) - e_{a,i}(n), \quad i = 1, 2. \tag{74}$$

Equation (8) is obtained replacing (74) in (7) and assuming that terms depending on $e_{a,i}^k(n)$, $k \geq 2$ are sufficiently small to be disregarded for all $n \geq 0$. In other words, we assume that the deviation between the component equalizers and the zero-forcing solution is always small.

To calculate the first and second moments of the random i.i.d. variables $\gamma(n)$ and $\beta(n)$, we assume that

B2. The constellation used to generate the $a(n)$ has circular symmetry, so that $\mathrm{E}\{a^k(n)\} = 0$ for all odd integers $k > 0$. This assumption is not restrictive, since this condition is true for practical constellations.

Using B2, we find that $\mathrm{E}\{\beta(n)\} = 0$,

$$\sigma_\beta^2 \triangleq \mathrm{E}\{\beta^2(n)\} = \mathrm{E}\{a^6(n) - r^2a^2(n)\}, \tag{75}$$

$$\bar{\gamma} \triangleq \mathrm{E}\{\gamma(n)\} = 3\mathrm{E}\{a^2(n)\} - r, \tag{76}$$

and

$$\xi \triangleq \mathrm{E}\{\gamma^2(n)\} = 3r\mathrm{E}\{a^2(n)\} + r^2. \tag{77}$$

References

[1] R. Candido, M. T. M. Silva, and V. H. Nascimento, “Affine combinations of adaptive filters,” in *Proc. 42th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 2008, pp. 236–240.

[2] ——, “On combinations of CMA equalizers,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Taipei, Taiwan, 2009, pp. 2865–2868.

[3] P. Andersson, “Adaptive forgetting in recursive identification through multiple models,” *International Journal of Control*, vol. 42, pp. 1175–1193, Nov. 1985.

[4] M. Niedźwiecki, “Identification of nonstationary stochastic systems using parallel estimation schemes,” *IEEE Trans. Autom. Control*, vol. 35, pp. 329–334, Mar. 1990.

[5] ——, “Multiple-model approach to finite memory adaptive filtering,” *IEEE Trans. Signal Process.*, vol. 40, pp. 470–473, Feb. 1992.

[6] A. C. Singer and M. Feder, “Universal linear prediction by model order weighting,” *IEEE Trans. Signal Process.*, vol. 47, pp. 2685–2700, Oct. 1999.

[7] M. Martínez-Ramón *et al.*, “An adaptive combination of adaptive filters for plant identification,” in *Proc. 14th Int. Conf. Digital Signal Process.*, vol. 2, Santorini, Greece, 2002, pp. 1195–1198.

[8] 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 Process.*, vol. 54, pp. 1078–1090, Mar. 2006.

[9] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, “A normalized adaptation scheme for the convex combination of two adaptive filters,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Las Vegas, NV, 2008, pp. 3301–3304.

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

[11] J. Arenas-Garcia and A. R. Figueiras-Vidal, “Adaptive combination of proportionate filters for sparse echo cancellation,” *IEEE Trans. Audio, Speech, and Language Process.*, vol. 17, pp. 1087–1098, Aug. 2009.

[12] 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 Process.*, vol. 56, pp. 1853–1864, May 2008.

[13] 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. 42th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 2008, pp. 230–236.

[14] A. T. Erdogan, S. S. Kozat, and A. C. Singer, "Comparison of convex combination and affine combination of adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Taipei, Taiwan, 2009, pp. 3089–3092.

[15] S. S. Kozat and A. C. Singer, "A performance-weighted mixture of LMS filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Taipei, Taiwan, 2009, pp. 3101–3104.

[16] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, Upper Saddle River, 2002.

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

[18] B. Farhang-Boroujeny, *Adaptive Filters - Theory and Applications*. John Wiley & Sons, West Sussex, 1998.

[19] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 3rd ed. Springer, 2008.

[20] J. R. Treichler, I. Fijalkow, and C. R. Johnson Jr., "Fractionally spaced equalizers," *IEEE Signal Process. Mag.*, vol. 13, pp. 65–81, May 1996.

[21] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard (CMA) adaptive equalizers," *IEEE Trans. Signal Process.*, vol. 44, pp. 818–826, Apr. 1996.

[22] M. Mboup and P. A. Regalia, "A gradient search interpretation of the super-exponential algorithm," *IEEE Transactions on Information Theory*, vol. 46, pp. 2731–2734, Nov. 2000.

[23] J. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally spaced blind adaptive equalizers," *IEEE Trans. Signal Process.*, vol. 48, pp. 80–91, Jan. 2000.

[24] D. N. Godard, "Self-recovering equalization and carrier tracking in two dimensional data communication system," *IEEE Trans. Commun.*, vol. 28, pp. 1867–1875, Nov. 1980.

[25] J. R. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 334–358, Apr. 1983.

[26] V. H. Nascimento and M. T. M. Silva, "Stochastic stability for the Constant Modulus Algorithm," *IEEE Trans. Signal Process.*, vol. 56, pp. 4984–4989, Oct. 2008.

[27] J. Arenas-García and A. R. Figueiras-Vidal, "Improved blind equalization via adaptive combination of constant modulus algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. III, Toulouse, France, 2006, pp. 756–759.

[28] M. T. M. Silva and V. H. Nascimento, "Tracking analysis of the Constant Modulus Algorithmn," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Las Vegas, NV, 2008, pp. 3561–3564.

[29] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Transactions on Information Theory*, vol. 39, pp. 504–519, Mar. 1993.

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

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

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

[33] ——, "A feedback analysis of the tracking performance of blind adaptive equalization algorithms," in *Proc. IEEE Conf. on Dec. and Control*, vol. 1, Dec. 1999, pp. 174–179.

[34] M. T. M. Silva and M. D. Miranda, "Tracking issues of some blind equalization algorithms," *IEEE Signal Process. Lett.*, vol. 11, pp. 760–763, Sept. 2004.

[35] 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. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. II, Orlando, FL, 2002, pp. 1385–1388.

[36] V. H. Nascimento, M. T. M. Silva, L. A. Azpicueta-Ruiz, and J. Arenas-García, "On the tracking performance of combinations of least mean squares and recursive least squares adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Dallas, TX, 2010, pp. 3710–3713.

[37] J. E. Mazo, "On the independent theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963–993, May/Jun. 1979.

[38] B. D. O. Anderson, R. R. Bitmead, C. R. Johnson Jr., P. V. Kokotovic, R. L. Kosut, I. M. Y. Mareels, L. Praly, and B. D. Riedle, *Stability of adaptive systems: passivity and averaging analysis*. MIT Press, Cambridge, MA, 1986.

[39] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Companies, Inc., 1991.

[40] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 793–806, Aug 1986.