Improving the tracking capability of adaptive filters via convex combination

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

Abstract

As is well known, Hessian-based adaptive filters (such as the recursive-least squares algorithm – RLS, for supervised adaptive filtering, or the Shalvi-Weinstein algorithm – SWA, for blind equalization) converge much faster than gradient-based algorithms (such as the least-mean-squares algorithm – LMS, or the constant-modulus algorithm – CMA). However, when the problem is tracking a time-variant filter, the issue is not so clear-cut: there are environments for which each family presents better performance. Given this, we propose the use of a convex combination of algorithms of different families to obtain an algorithm with superior tracking capability. We show the potential of this combination and provide a unified theoretical model for the steady-state excess mean-square error for convex combinations of gradient- and Hessian-based algorithms, assuming a random-walk model for the parameter variations. The proposed model is valid for algorithms of the same or different families, and for supervised (LMS and RLS) or blind (CMA and SWA) algorithms.

Index Terms

Adaptive filters, adaptive equalizers, convex combination, tracking, least mean square methods, recursive estimation, unsupervised learning.

I. Introduction

When choosing an adaptive algorithm for a given application, one of the important points to be considered is the algorithm's ability to track variations in the statistics of the signals of interest. This is especially important in mobile communications [1], and in applications that demand long filters, such as acoustic echo cancellation [2].

Manuscript received July 2, 2007; revised January 9, 2008. This work was partly supported by CNPq under grant No. 303361/2004-2 and by FAPESP under grant No. 2004/15114-2. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jonathon Chambers.

M. T. M. Silva and V. H. Nascimento are with the Dept. of Electronic Systems Eng., University of São Paulo, São Paulo, SP, Brazil, e-mails: {magno, vitor}@lps.usp.br, ph. +55-11-3091-5606, fax: +55-11-3091-5718.

Digital Object Identifier 10.1109/TSP.2008.919105

There are two standard approaches to adaptive filtering: gradient-based, such as the least-mean-squares algorithm, LMS; and Hessian-based, such as the recursive least-squares algorithm, RLS. Of these two, the latter, given its use of estimates of the Hessian of the cost function being minimized, converges at a much faster rate than the former, as is well-known [3], [4]. However, Eweda showed in [5] that for the tracking of time-variant parameters, LMS may in fact outperform RLS, depending on the statistics of the regressor and desired signals. A similar behavior was observed in blind algorithms for channel equalization: the gradient-based constant-modulus algorithm (CMA) [6], [7] has a considerably slower convergence than the Hessian-based Shalvi-Weinstein algorithm (SWA) [8]. Again, as in the case of LMS and RLS, it was shown in [9] that the tracking capabilities of CMA and SWA depend heavily on the statistics of the input signals, and CMA may outperform SWA, depending on the environment.

In this paper we use the observations of Eweda and [9], together with the convex combination of adaptive filters proposed in [10] and further extended and analyzed respectively in [11] and [12], to take advantage of the different tracking capabilities of LMS and RLS (resp., CMA and SWA) to arrive at supervised (resp., blind) algorithms with superior tracking performance.

The idea of combining the outputs of several different independently-run adaptive algorithms to achieve better performance than that of a single filter is not new. It apparently was first proposed in [13], and latter improved in [14], [15]. Similar ideas have also been used in the information theory literature, see, e.g., [16]. The algorithms proposed in [13]–[15] are based on a Bayesian argument, and construct an overall (combined) filter through a linear combination of the outputs of several independent adaptive filters. The weights are the *a posteriori* probabilities that the underlying models used to describe each individual algorithm are "true". Since the weights add up to one, in a sense these first papers also proposed "convex" combinations of algorithms. The method of [12] is receiving more attention due to its relative simplicity and the proof that the combination is universal, i.e., the combined estimate is at least as good as the best of the component filters in steady-state, for stationary inputs.

Previous works on convex combinations of adaptive filters mostly restricted themselves to combinations of filters of the same families, i.e, two LMS [11], [12], [17], two RLS [11] or two CMA [18] filters with different step-sizes or forgetting factors. A combination of two filters based on different cost-functions (but both gradient-based) was proposed in [19], combining normalized LMS and normalized sign-LMS to obtain an algorithm with improved robustness without the slow convergence behavior of sign-LMS. Using a different combination rule, combinations of Kalman or RLS filters were proposed, using the different combination rule proposed in [14], [15] (but the combination rule also allows for the use of other algorithms, and for the use of filters with a different number of taps). It should be noted that theoretical models (approximations for the overall filter's steady-state excess mean-square error) are available in the literature only for the combination of two LMS algorithms [12], [17]. However, the possibility of extension of these models to different combinations of algorithms was already indicated in [17] and [12]. The proof in [12] that the combination is universal also applies to different choices of algorithms.

The present paper extends previous results in four ways: (1) proposing the combination of *supervised* algorithms of different families to take advantage of their different tracking capabilities; (2) extending this result also to *blind* algorithms of different families; (3) providing theoretical models (in a unified way) for the steady-state mean-square error for combinations of filters of the same or different families, assuming a random-walk model for the parameter variations; and (4) providing theoretical models for combinations of blind algorithms of the same or different families. To the best of our knowledge, all these are novel contributions. In particular, the models for the combinations of two RLS, two CMA or two SWA filters are also new results. For combinations of filters of different families, the results presented here are more accurate than those we published as conference papers, in [20] (supervised filters) and [21] (blind filters). We also extend these previous results both by providing a unified analysis, which is valid for combinations of filters of the same or of different families, and for supervised or blind algorithms. In this sense, the analysis provided here also recovers the results for the combination of two LMS filters, presented in [12]. Unlike this reference, here we use the traditional analysis method, where one computes a recursion for the autocorrelation matrix of the weight-error vector of a filter, as opposed to the feedback method of [3]. In passing, we should add that the analysis for blind adaptive filters using the traditional method that we present here is also novel, and it gives the same results obtained using the feedback method for CMA in [22]–[24] and for SWA in [9].

In the remainder of this section we provide a few examples to motivate the combination of filters of two different families, both for supervised (combination of one RLS with one LMS) and for blind algorithms (combination of one SWA with one CMA).

A. Introductory simulations

In the supervised case, we simulate the identification of a time-variant channel (Rayleigh fading channel) with 5 coefficients [3, p. 401]. The parameters λ, μ, μ_α and α^+, which control the adaptive filters and the combination algorithm, are described in Section II. Figure 1-(a) shows curves of one realization of squared *a priori* errors for RLS ($\lambda = 0.995$), LMS ($\mu = 0.01$), and their convex combination C-RLS-LMS ($\mu_\alpha = 400, \alpha^+ = 4$). To facilitate the visualization, the curves were filtered by a moving-average filter with 512 coefficients. The convex combination performs at least as well as the best of its components, outperforming slightly both of them in some situations. This behavior can be confirmed by the mixing parameter $\eta(n)$ shown in Figure 1-(b). When $\eta(n) \approx 1$, the combination performs close to RLS, when $\eta(n) \approx 0$, it is close to LMS, and when $0 < \eta(n) < 1$, the combination tends to be better than both independent filters.

For the same example, we show in Figure 2 a comparison between C-RLS-LMS, the convex combination of two LMSs (CLMS), and the robust variable step-size LMS (RVSS-LMS) of [25]. We observe that C-RLS-LMS presents better tracking performance than CLMS, and both are better than RVSS-LMS. Thus, the convex combination of one RLS with one LMS can be a better alternative for tracking performance.

In the blind equalization case, we consider a Rayleigh fading channel with fast variation (maximum Doppler spread $f_D = 80$ Hz) and 3 coefficients [3, p. 401]. Figure 3-(a) shows residual intersymbol interference (ISI) [8] curves for SWA ($\lambda = 0.999$), CMA ($\mu = 2 \times 10^{-3}$), and their convex combination ($\mu_\alpha = 15$, $\alpha^+ = 4$). The combination usually performs as the best of each component equalizer, being slightly better than both of them in some situations. In this example, the adaptation of the mixing parameter was not fast enough to switch between filters in a few brief occasions, most notably at the end of the simulation. This happens because the adaptation rule for the mixing parameter $\eta(n)$ needs some time to identify that a change is necessary. Figure 3-(b) shows the mixing parameter $\eta(n)$, which confirms this behavior. When $\eta(n) \approx 1$, the combination performs close to SWA, when $\eta(n) \approx 0$, it is close to CMA, and when $0 < \eta(n) < 1$, it can be better than both of its equalizers.

B. *Organization of the paper*

The paper is organized as follows. In the next section, we describe the convex combination of two adaptive filters, for both supervised (RLS and LMS) and blind (SWA and CMA) algorithms. In Section III, the tracking analyses are presented. Initially, in Section III-A, we summarize results for the tracking analysis of the LMS and RLS algorithms. Then, in Section III-B, we present the tracking analysis of CMA and SWA using the traditional method. Finally, in Section III-C, the tracking analysis of the considered convex combinations is provided. Comparisons between analytical and experimental results for the steady-state excess mean-square error are shown through simulations in Section IV. Section V provides a summary of the main conclusions of the paper.

II. Problem Formulation

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

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

where the subscript i is associated to the first ($i = 1$) or second ($i = 2$) filter of the combination, $\mathbf{w}_i(n)$ represents the length-M coefficient vector, ρ_i is a step-size, $\mathbf{M}_i(n)$ is a symmetric non-singular matrix, $\mathbf{u}(n)$ is the input regressor vector, and $e_i(n)$ is the estimation error. Many algorithms can be written as in (1), by proper choices of ρ_i, $\mathbf{M}_i(n)$, and $e_i(n)$. In this paper, in order to simplify the arguments, we assume that all quantities are real.

In supervised adaptive filtering,

$$e_i(n) = d(n) - y_i(n), \tag{2}$$

where $y_i(n) = \mathbf{u}^T(n)\mathbf{w}_i(n-1)$ is the output of the i^{th} transversal filter and $d(n)$ is the desired response. In this case, a linear regression model holds, that is,

$$d(n) = \mathbf{u}^T(n)\mathbf{w}_{\text{o}}(n-1) + v(n), \tag{3}$$

with $\mathbf{w}_\text{o}(n-1)$ being the time-variant optimal solution and $v(n)$ a zero-mean random process with variance $\sigma_v^2 = \text{E}\{v^2(n)\}$, uncorrelated with $\mathbf{u}(n)$ [3, Sec. 6.2.1]. Here, $\text{E}\{\cdot\}$ denotes the expectation operation and the sequences $\{u(n)\}$ and $\{v(n)\}$ are assumed stationary. We shall use the common assumption that $v(n)$ is independent of $\mathbf{u}(n)$ (not only uncorrelated) [3].

In blind equalization, algorithms based on the constant modulus cost function [6], [7] define $e_i(n)$ as

$$e_i(n) = \left[r^a - y_i^2(n)\right] y_i(n), \tag{4}$$

where $r^a = \text{E}\{a^4(n)\}/\text{E}\{a^2(n)\}$ and $\{a(n)\}$ represents the transmitted sequence. Due to the equivalence between the constant modulus and Shalvi-Weinstein cost functions shown in [26], CMA and SWA seek to optimize the same criterion. Thus, although SWA was originally derived in [8] through the minimization of the SW cost function using empirical cumulants, it can also be interpreted as a constant-modulus-based algorithm.

The supervised LMS and RLS algorithms and the blind CMA and SWA employ the step-sizes ρ_i, the estimation errors $e_i(n)$, and the matrices $\mathbf{M}_i(n)$ as in Table I. In this table, $\mathbf{I}$ is the $M \times M$ identity matrix, $0 \ll \lambda_i < 1$ is a forgetting factor, and

$$\bar{\gamma} \triangleq 3\text{E}\{a^2(n)\} - r^a. \tag{5}$$

For RLS and SWA, $\mathbf{M}_i(n) = \widehat{\mathbf{R}}_i^{-1}(n)$ is obtained via the matrix inversion lemma [3, 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., $\mathbf{R} \triangleq \text{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}$. These matrices are related via

$$\text{E}\big\{\widehat{\mathbf{R}}_i(n)\big\} = \frac{\mathbf{R}}{(1-\lambda_i)}. \tag{6}$$

Although we use the same notation for the LMS and CMA step-sizes in Table I, the step-size intervals which ensure the convergence and stability of such algorithms are different. For the LMS algorithm, this step-size interval is well-known in the literature [3], [4], whereas for CMA, the derivation of this interval remains an open problem.

The convex combination of two adaptive filters proposed in [11], [12], and [18] is depicted in Figure 4. Figure 4-(a) considers supervised filtering and can be used for different applications, such as system identification, adaptive equalization, echo or noise cancelation, etc. [3], [4]. Figure 4-(b) shows a simplified communications system with a convex 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., [22], [27]–[29] 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. For real data, perfect equalization is achieved when the overall channel-equalizer impulse response is of the form $[\,0\,\cdots\,0\;\delta\;0\,\cdots\,0\,]^T$, where $\delta = \pm 1$. In this case, the equalizer reaches the so-called zero-forcing solution and $y(n) = \delta a(n - \tau_d)$. The two possibilities $\delta = 1$ or $\delta = -1$ occur due the fact that constant-modulus-based algorithms do not solve phase ambiguities introduced by the channel. Such ambiguities can be corrected by using differential modulation [1]. Since both solutions give equally good results, we assume in the next section that the algorithm is initialized so that it converges to the case of $\delta = 1$. Since we are studying its steady-state performance, this does not imply a restriction in the applicability of our results.

In both schemes, the output of the overall filter is given by $y(n) = \eta(n)y_1(n) + [1 - \eta(n)]y_2(n)$. The mixing parameter $\eta(n)$ is modified via an auxiliary variable $\alpha(n-1)$ and a sigmoidal function [11], [12], that is,

$$\eta(n) = \mathrm{sgm}[\alpha(n-1)] = \frac{1}{1 + e^{-\alpha(n-1)}}, \tag{7}$$

with $\alpha(n)$ being updated as

$$\alpha(n) = \alpha(n-1) + \mu_\alpha e_\alpha(n)\eta(n)[1 - \eta(n)], \tag{8}$$

where

$$e_\alpha(n) = e(n)[y_1(n) - y_2(n)] \tag{9}$$

and $e(n) = \eta(n)e_1(n) + [1 - \eta(n)]e_2(n)$. Eq. (8) was obtained in [11], [12] and [18], using a stochastic gradient rule to minimize the instantaneous MSE cost function for the supervised case and the constant modulus cost function for the blind case. The auxiliary variable $\alpha(n)$ is used to keep $\eta(n)$ in the interval $[0, 1]$. A drawback of this scheme is that $\alpha(n)$ stops updating whenever $\eta(n)$ is close to 0 or 1. To avoid this, [12], [18] suggest that $\alpha(n)$ be restricted (by simple saturation) to lie inside a symmetric interval $[-\alpha^+,\ \alpha^+]$. Thus, a minimum level of adaptation is always guaranteed [12].

Following [12], in all our simulations we restrict $\alpha(n)$ to the above interval, but use a modified mixing variable $\eta_u(n)$, as described below:

$$\eta_u(n) = \begin{cases} \eta(n), & \text{if } |\alpha(n)| < \alpha^+ - \epsilon, \\ 1, & \text{if } \alpha(n) \geq \alpha^+ - \epsilon, \\ 0, & \text{if } \alpha(n) \leq -\alpha^+ + \epsilon, \end{cases}$$

where ϵ is a small constant. This modification tends to improve the performance of the overall algorithm when one of the component filters performs substantially better than the other [12]. Note that the adaptation rule for $\alpha(n)$ still uses the original, unmodified $\eta(n)$, but the overall output is now given by

$$y_u(n) = \eta_u(n)y_1(n) + [1 - \eta_u(n)]y_2(n).$$

III. Tracking analysis

We assume that in a nonstationary environment, the variation in the optimal solution $\mathbf{w}_{\rm o}$ follows a random-walk model [3, p. 359], that is,

$$\mathbf{w}_{\rm o}(n) = \mathbf{w}_{\rm o}(n-1) + \mathbf{q}(n). \tag{10}$$

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 $\{\mathbf{w}_{\rm o}(-1), \mathbf{w}(-1), \alpha(-1)\}$ and of $\{\mathbf{u}(l)\}$ for all l [3, 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 (see assumption A3 below).

One measure of the filter performance is given by the excess mean-square error (EMSE), defined as

$$\zeta \triangleq \lim_{n\to\infty} {\rm E}\{e_a^2(n)\}, \tag{11}$$

where

$$e_a(n) = \mathbf{u}^T(n)\widetilde{\mathbf{w}}(n-1), \tag{12}$$

and

$$\widetilde{\mathbf{w}}(n-1) = \mathbf{w}_{\rm o}(n-1) - \mathbf{w}(n-1). \tag{13}$$

The *a priori* error $e_a(n)$ of the overall scheme can be written as a function of the *a priori* errors of the component filters, i.e.,

$$e_a(n) = \eta_u(n)e_{a,1}(n) + [1-\eta_u(n)]e_{a,2}(n), \tag{14}$$

where $e_{a,i}(n) = \mathbf{u}^T(n)\widetilde{\mathbf{w}}_i(n-1)$ and $\widetilde{\mathbf{w}}_i(n-1) = \mathbf{w}_{\rm o}(n-1) - \mathbf{w}_i(n-1),\ i = 1,2$. It is common in the literature to evaluate the EMSE as

$$\zeta = \lim_{n\to\infty} {\rm Tr}\big(\mathbf{R}\mathbf{S}(n-1)\big), \tag{15}$$

where ${\rm Tr}(\mathbf{A})$ stands for the trace of matrix $\mathbf{A}$ and

$$\mathbf{S}(n-1) \triangleq {\rm E}\{\widetilde{\mathbf{w}}(n-1)\widetilde{\mathbf{w}}^T(n-1)\}. \tag{16}$$

This approach is based on the independence assumption between the regressor vector $\mathbf{u}(n)$ and weight-error vector $\widetilde{\mathbf{w}}(n-1)$. This condition is a part of the widely used independence assumptions in adaptive filter theory [4]. It was shown in [30], for instance, that for LMS-type algorithms and for sufficiently small step-sizes, the results obtained from such independence assumptions tend to match reasonably well the real filter performance. Furthermore, [31] argues that it is not necessary for $\widetilde{\mathbf{w}}(n-1)$ to be independent of $\mathbf{u}(n)$, but only of $\mathbf{u}(n)\mathbf{u}^T(n)$, which represents a weaker assumption, since the outer product does not contain phase information about $\mathbf{u}(n)$.

The main focus of the analysis that follows is the behavior of the algorithm in steady-state, i.e., after initial convergence of the coefficients. Although the optimum weights are time-variant, under the model assumed for their variation, it is well-known that the EMSE of an adaptive filter approaches a steady-state value, see, e.g., [3].

It was proved in [12] that the performance of convex combinations of adaptive filters is at least as good as that of the best of its components in stationary environments. In this work, we will make the assumption that the adaptation of $\eta(n)$ is fast enough, so that, after initial convergence, the overall algorithm will follow the best component filter at every instant. The simulations presented in Section I-A shows that, even when the optimum coefficients change quite rapidly, this assumption is rarely violated. Note that the same assumption was implicitly used in [12].

A. Tracking analysis of supervised algorithms

There have been several works in the literature on the steady-state tracking performance of supervised adaptive algorithms (see, e.g., [3]–[5], [24] and the references therein). For sufficiently small μ and $(1-\lambda)$, analytical expressions for the EMSE of LMS and RLS algorithms are given respectively by

$$\zeta^{\text{LMS}} = \frac{\mu\sigma_v^2 \text{Tr}(\mathbf{R}) + \mu^{-1}\text{Tr}(\mathbf{Q})}{2} \tag{17}$$

and

$$\zeta^{\text{RLS}} = \frac{\sigma_v^2(1-\lambda)M + (1-\lambda)^{-1}\text{Tr}(\mathbf{QR})}{2}. \tag{18}$$

The ratio between the minimum value of ζ for these algorithms is given as

$$\frac{\zeta_{\text{min}}^{\text{RLS}}}{\zeta_{\text{min}}^{\text{LMS}}} = \sqrt{\frac{M\text{Tr}(\mathbf{QR})}{\text{Tr}(\mathbf{R})\text{Tr}(\mathbf{Q})}}. \tag{19}$$

This ratio, obtained in [5], allows us to compare the tracking performance of LMS and RLS. Clearly, the results of such comparison depend on the environment, i.e., there are situations where RLS has superior tracking capability compared to LMS, and vice-versa [5]. This is highlighted considering three different choices for matrix $\mathbf{Q}$ [3]–[5]:

(i) $\mathbf{Q}$ is a multiple of $\mathbf{I}$: the performance of LMS is similar to that of RLS;

(ii) $\mathbf{Q}$ is a multiple of $\mathbf{R}$: LMS is superior; and

(iii) $\mathbf{Q}$ is a multiple of $\mathbf{R}^{-1}$: RLS is superior.

B. Tracking analysis of blind equalization algorithms

Analytical expressions for the EMSE of blind equalization algorithms have been computed in the literature (see, e.g., [22]–[24], [32], and [9]). Using Lyapunov stability and averaging analysis, an approximate expression for the EMSE of CMA was obtained in [32]. Later, [22] and [23] focused on the CMA steady-state performance, using the feedback analysis. Considering still the feedback method, [9] analyzed the tracking of constant-modulus-based algorithms (including SWA) in a unified manner. In the sequel, we present an alternative analysis using the traditional method, i.e., we compute the EMSE of CMA and SWA via (15). In the remainder of this section, we suppress the subscript i, since we are interested in the analysis of each algorithm individually.

The steady-state analysis of blind algorithms of the form (1) is based on the following assumptions:

A1. $\mathrm{E}\{a^k(n)\} = 0, \quad k = 2m+1, \; m \in \mathbb{N}$, and $\bar{\gamma} > 0$, i.e., $a(n)$ is sub-Gaussian, and the constellation is symmetric, as is the case for most constellations used in digital communications [3], [8].

A2. The regressor vector $\mathbf{u}(n)$ and the weight-error vector $\widetilde{\mathbf{w}}(n-1)$ are independent in the steady-state. As mentioned before, this independence assumption is commonly used with good results for the analysis of adaptive algorithms [4].

A3. The signal-to-noise ratio at the input is high, so that $a(n-\tau_d) \approx \mathbf{u}^T(n)\mathbf{w}_{\mathrm{o}}(n-1)$, i.e., the optimum filter achieves perfect equalization. However, due to channel variation and gradient noise, the equalizer weight vector $\mathbf{w}(n-1)$ is not equal to $\mathbf{w}_{\mathrm{o}}(n-1)$, even in steady-state. Using (13), (12), and the above approximation, we have $y(n) = \mathbf{u}^T(n)\mathbf{w}(n-1) = \mathbf{u}^T(n)[\mathbf{w}_{\mathrm{o}}(n-1) - \widetilde{\mathbf{w}}(n-1)]$, that is,

$$y(n) \approx a(n-\tau_d) - e_a(n). \tag{20}$$

This approximation was also used in the CMA steady-state analyses of [22, Sec. III-A] and [23]. As in the cited references, we assume that the filter parameters and initial condition are chosen so that the combined channel-equalizer response converges to $[\,0\,\cdots\,0\;\delta\;0\,\cdots\,0\,]^T$, with $\delta = 1$. A similar analysis holds for the other optimum solution (convergence to a zero-forcing solution with $\delta = -1$).

Using A3, (4) can be rewritten as

$$e(n) = \gamma(n)e_a(n) + e_a^3(n) - 3e_a^2(n)a(n-\tau_d) + \beta(n), \tag{21}$$

where

$$\gamma(n) = 3a^2(n-\tau_d) - r^a \tag{22}$$

and

$$\beta(n) = -a^3(n-\tau_d) + a(n-\tau_d)r^a. \tag{23}$$

If $e_a^2(n)$ is reasonably small in steady-state, terms depending on higher-order combinations of $e_a(n)$ can be disregarded in (21), which leads to the approximation

$$e(n) \approx \gamma(n)e_a(n) + \beta(n). \tag{24}$$

From (23) and A1, $\beta(n)$ is an i.i.d. random variable, which satisfies $\mathrm{E}\{\beta(n)\} = 0$ and

$$\sigma_\beta^2 = \mathrm{E}\{\beta^2(n)\} = \mathrm{E}\{a^6(n) - (r^a)^2 a^2(n)\}. \tag{25}$$

To proceed, we also assume that

A4. $\beta(n)$ and $\gamma(n)$ are independent of $\widetilde{\mathbf{w}}(n-1)$ in steady-state. This essentially requires that the weigh-error vector be insensitive, in steady-state, to the actually transmitted symbols $\{a(n)\}$.

Subtracting both sides of (1) from $\mathbf{w}_{\rm o}(n)$, using the model (10) and the approximation (24), we arrive at

$$\begin{aligned}\widetilde{\mathbf{w}}(n) =&[\mathbf{I}-\rho\gamma(n)\mathbf{M}(n)\mathbf{u}(n)\mathbf{u}^T(n)]\widetilde{\mathbf{w}}(n-1)\\ &-\rho\beta(n)\mathbf{M}(n)\mathbf{u}(n)+\mathbf{q}(n).\end{aligned} \tag{26}$$

Remark: Note that (26) also holds for supervised filters, only in that case we would have $\gamma(n)\equiv 1$, and the measurement noise $v(n)$ instead of $\beta(n)$. In addition, $\beta(n)$ is identically zero for constellations which do have constant modulus, so the variability in the modulus of $a(n)$ (as measured by $\beta(n)$) plays the role of measurement noise for the class of blind algorithms considered here.

Multiplying (26) by its transpose, taking the expectations of both sides, and using the fact that $\mathbf{q}(n)$ is independent of the initial conditions and of $\mathbf{u}(n)$, we obtain

$$\begin{aligned}\mathrm{E}\{\widetilde{\mathbf{w}}(n)\widetilde{\mathbf{w}}^T(n)\} \approx\ & \mathrm{E}\{\widetilde{\mathbf{w}}(n-1)\widetilde{\mathbf{w}}^T(n-1)\}\\
&-\overbrace{\rho\mathrm{E}\{\gamma(n)\widetilde{\mathbf{w}}(n-1)\widetilde{\mathbf{w}}^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}(n)\}}^{\mathcal{A}}\\
&-\overbrace{\rho\mathrm{E}\{\gamma(n)\mathbf{M}(n)\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}(n-1)\widetilde{\mathbf{w}}^T(n-1)\}}^{\mathcal{B}}\\
&+\overbrace{\rho^2\mathrm{E}\{\gamma^2(n)\mathbf{M}(n)\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}(n-1)}^{\mathcal{C}}\\
&\qquad\times\widetilde{\mathbf{w}}^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}(n)\}\\
&+\overbrace{\rho^2\mathrm{E}\{\beta^2(n)\mathbf{M}(n)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}(n)\}}^{\mathcal{D}}\\
&-\overbrace{\rho\mathrm{E}\{\beta(n)\widetilde{\mathbf{w}}(n-1)\mathbf{u}^T(n)\mathbf{M}(n)\}}^{\mathcal{E}}\\
&-\overbrace{\rho\mathrm{E}\{\beta(n)\mathbf{M}(n)\mathbf{u}(n)\widetilde{\mathbf{w}}^T(n-1)\}}^{\mathcal{F}}\\
&+\overbrace{\rho^2\mathrm{E}\{\gamma(n)\beta(n)\mathbf{M}(n)\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}(n-1)\mathbf{u}^T(n)\mathbf{M}(n)\}}^{\mathcal{G}}\\
&+\overbrace{\rho^2\mathrm{E}\{\gamma(n)\beta(n)\mathbf{M}(n)\mathbf{u}(n)\widetilde{\mathbf{w}}^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}(n)\}}^{\mathcal{H}}\\
&+\mathrm{E}\{\mathbf{q}(n)\mathbf{q}^T(n)\}.\end{aligned} \tag{27}$$

To simplify (27), we remark that:

R1. When $\mathbf{M}(n)=\widehat{\mathbf{R}}^{-1}(n)$ appears inside the expectations of (27), we simply replace it by its mean. Using (6), we have $\mathrm{E}\{\mathbf{M}(n)\}\approx(1-\lambda)\mathbf{R}^{-1}$. For $\lambda\approx 1$, this is a reasonable steady-state approximation [3, Sec. 6.9.2].

R2. For channels with long impulse response, the following approximations are reasonable

$$\mathrm{E}\{\gamma(n)u^2(n)\} \approx \bar{\gamma}\,\mathrm{E}\{u^2(n)\} \tag{28}$$

and

$$\mathrm{E}\{\beta^2(n)u^2(n)\} \approx \sigma_\beta^2\mathrm{E}\{u^2(n)\}, \tag{29}$$

with $\bar{\gamma}$ and σ_β^2 defined in (5) and (25), respectively.

R3. In steady-state, the weight-error vector has zero mean, i.e., $\mathrm{E}\{\widetilde{\mathbf{w}}(n)\} = \mathbf{0}$.

Remarks R2 and R3 are justified in detail in Appendix A.

Now, using assumptions A1-A4 and remarks R1-R3, we can evaluate the terms $\boldsymbol{\mathcal{A}}$-$\boldsymbol{\mathcal{H}}$ of (27):

$\boldsymbol{\mathcal{A}}$- Using A2, A4, R1, (28), and (16), $\boldsymbol{\mathcal{A}}$ can be approximated by

$$\begin{aligned}\boldsymbol{\mathcal{A}} &= \rho\mathrm{E}\{\mathrm{E}\{\gamma(n)\widetilde{\mathbf{w}}(n-1)\widetilde{\mathbf{w}}^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}(n)|\mathbf{u}(n)\}\}\\ &\approx \rho\mathrm{E}\{\gamma(n)\mathrm{E}\{\widetilde{\mathbf{w}}(n-1)\widetilde{\mathbf{w}}^T(n-1)\}\mathbf{u}(n)\mathbf{u}^T(n)\}\mathrm{E}\{\mathbf{M}(n)\}\\ &\approx \rho\,\bar{\gamma}\,\mathbf{S}(n-1)\,\mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}\mathrm{E}\{\mathbf{M}(n)\}\\ &= \rho\,\bar{\gamma}\,\mathbf{S}(n-1)\,\mathbf{R}\,\mathrm{E}\{\mathbf{M}(n)\}.\end{aligned} \tag{30}$$

$\boldsymbol{\mathcal{B}}$- Analogously, we obtain for $\boldsymbol{\mathcal{B}}$

$$\boldsymbol{\mathcal{B}} \approx \rho\,\bar{\gamma}\,\mathrm{E}\{\mathbf{M}(n)\}\,\mathbf{R}\,\mathbf{S}(n-1). \tag{31}$$

$\boldsymbol{\mathcal{C}}$- For CMA, since $\rho = \mu$ and $\mathbf{M}(n) = \mathbf{I}$, $\boldsymbol{\mathcal{C}}$ reduces to

$$\boldsymbol{\mathcal{C}}^{\text{CMA}} \approx \mu^2\mathrm{E}\{\gamma^2(n)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{S}(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\}. \tag{32}$$

Analogously for SWA, replacing $\mathbf{M}(n) = \widehat{\mathbf{R}}^{-1}(n)$ by its mean and $\rho = (\bar{\gamma})^{-1}$, we get

$$\begin{aligned}\boldsymbol{\mathcal{C}}^{\text{SWA}} \approx\,& (1-\lambda)^2(\bar{\gamma})^{-2}\mathrm{E}\{\gamma^2(n)\mathbf{R}^{-1}\mathbf{u}(n)\mathbf{u}^T(n)\\ &\times \mathbf{S}(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{R}^{-1}\}.\end{aligned} \tag{33}$$

From (30)-(33), we observe that the four first terms of the right-hand side of (27) are linear in $\mathbf{S}(n-1)$. Thus, assuming that the CMA step-size μ is sufficiently small and the SWA forgetting-factor λ is sufficiently close to 1, the term $\boldsymbol{\mathcal{C}}$ for both algorithms can be neglected with respect to the three first terms on the right-hand side of (27).

$\boldsymbol{\mathcal{D}}$- Using R1 and (29), we get

$$\boldsymbol{\mathcal{D}} \approx \rho^2\sigma_\beta^2\,\mathrm{E}\{\mathbf{M}(n)\}\,\mathbf{R}\,\mathrm{E}\{\mathbf{M}(n)\}. \tag{34}$$

$\boldsymbol{\mathcal{E}}$- The term $\boldsymbol{\mathcal{E}}$ can be rewritten as

$$\begin{aligned}\boldsymbol{\mathcal{E}} &= \rho \mathrm{E}\{\mathrm{E}\{\beta(n)\widetilde{\mathbf{w}}(n-1)\mathbf{u}^T(n)\mathbf{M}(n)|\mathbf{u}(n)\}\}\\ &\approx \rho \mathrm{E}\{\beta(n)\mathrm{E}\{\widetilde{\mathbf{w}}(n-1)\}\mathbf{u}^T(n)\mathbf{M}(n)\}.\end{aligned} \tag{35}$$

Since $\mathrm{E}\{\widetilde{\mathbf{w}}(n-1)\} = \mathbf{0}$ in steady-state, the term $\boldsymbol{\mathcal{E}}$ is a null matrix with dimensions $M \times M$, i.e., $\boldsymbol{\mathcal{E}} = \mathbf{0}_{M\times M}$.

Using the same arguments, we can show that the terms $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{G}}$, and $\boldsymbol{\mathcal{H}}$ are also null matrices in steady-state.

From the previous results, (27) reduces to

$$\begin{aligned}\mathbf{S}(n) \approx \mathbf{S}(n-1) &- \rho\bar{\gamma}\mathbf{S}(n-1)\,\mathbf{R}\,\mathrm{E}\{\mathbf{M}(n)\}\\ &- \rho\bar{\gamma}\mathrm{E}\{\mathbf{M}(n)\}\,\mathbf{R}\,\mathbf{S}(n-1)\\ &+ \rho^2\sigma_\beta^2\,\mathrm{E}\{\mathbf{M}(n)\}\,\mathbf{R}\,\mathrm{E}\{\mathbf{M}(n)\} + \mathbf{Q}.\end{aligned} \tag{36}$$

Following a similar analysis for LMS [4], it can be shown that this recursion is stable for sufficiently small ρ — however, (36) cannot be used to find a range of ρ that guarantees stability, due to the approximations made, particularly the discarding of (33) (we intend to pursue this matter elsewhere). For small ρ, when $n \to \infty$, we obtain for CMA

$$\mu\bar{\gamma}\,[\mathbf{S}(\infty)\mathbf{R} + \mathbf{R}\mathbf{S}(\infty)] \approx \mu^2\sigma_\beta^2\,\mathbf{R} + \mathbf{Q}. \tag{37}$$

Taking the trace of both sides of (37), we arrive at

$$\zeta^{\text{CMA}} = \mathrm{Tr}\big(\mathbf{R}\mathbf{S}(\infty)\big) \approx \frac{\mu\sigma_\beta^2\,\mathrm{Tr}(\mathbf{R}) + \mu^{-1}\mathrm{Tr}(\mathbf{Q})}{2\bar{\gamma}}. \tag{38}$$

Analogously for SWA, we have

$$2(1-\lambda)\mathbf{S}(\infty) \approx (\bar{\gamma})^{-2}\sigma_\beta^2\,(1-\lambda)^2\mathbf{R}^{-1} + \mathbf{Q}. \tag{39}$$

Multiplying both sides of (39) by $\mathbf{R}$, and taking the trace, we obtain

$$\zeta^{\text{SWA}} = \mathrm{Tr}\big(\mathbf{R}\mathbf{S}(\infty)\big) \approx \frac{\sigma_\beta^2\,(1-\lambda)M}{2\bar{\gamma}^2} + \frac{\mathrm{Tr}(\mathbf{Q}\mathbf{R})}{2(1-\lambda)}. \tag{40}$$

These results coincide with those of [23] and [9], obtained with the feedback analysis for sufficiently small μ and $(1-\lambda)$. Furthermore, as shown in [9], the ratio between the minimum value of ζ of SWA and CMA is equal to (19). This allows a direct extension to the blind context of the results comparing the tracking performance of the RLS and LMS algorithms.

In Table II, the analytical expressions for the EMSEs of the supervised and blind algorithms considered here are summarized for convenient reference. Comparing these expressions, one can observe that the EMSE of LMS and RLS can also be obtained respectively from the EMSE of CMA and SWA, making $\sigma_\beta^2 \leftarrow \sigma_v^2$ and $\bar{\gamma} \leftarrow 1$. Thus, there is an evident equivalence between LMS and CMA and between RLS and SWA, which reinforces the link between blind equalization algorithms and classical adaptive filtering of [33].

C. Tracking analysis of convex combinations

Using the same arguments of [12, Sec. III], it is possible to show that the convex combinations of algorithms of the form (1) are universal in the mean-square error sense. Thus, considering, for example, the convex combination of one RLS and one LMS, when RLS outperforms LMS in the steady-state, the behavior of the overall filter will be close to that of RLS and $\zeta \approx \zeta^{\rm RLS}$. On the other hand, when LMS is superior, $\zeta \approx \zeta^{\rm LMS}$. Moreover, there are situations where the combination will outperform both of them. In this case, the EMSE of the overall filter will be close to

$$\zeta \approx \zeta_{12} + \frac{\Delta\zeta_1 \Delta\zeta_2}{\Delta\zeta_1 + \Delta\zeta_2}, \tag{41}$$

where ζ_{12} is the cross-EMSE , defined as

$$\zeta_{12} \triangleq \lim_{n\to\infty} {\rm E}\{e_{a,1}(n)e_{a,2}(n)\}, \tag{42}$$

and $\Delta\zeta_i = \zeta_i - \zeta_{12},\ \ i = 1, 2$. Eq. (41) was obtained in [12, Eq. (33)] for the combination of two LMS filters with different step-sizes (μ_1-LMS and μ_2-LMS). However, it is also valid for the convex combination of other algorithms of the form (1), as for the combination of two RLSs with different forgetting factors (λ_1-RLS and λ_2-RLS), one RLS and one LMS (λ_1-RLS and μ_2-LMS), two CMAs (μ_1-CMA and μ_2-CMA), two SWAs (λ_1-SWA and λ_2-SWA), and one SWA and one CMA (λ_1-SWA and μ_2-CMA). The EMSE of the overall filter is the minimum of the values calculated by the expressions of each component filter and (41).

Thus, the tracking analysis of convex combinations of the algorithms of the form (1) depends on the results of Table II, and on the analytical expressions of the cross-EMSE. Using independence assumptions, such expressions can be obtained through the evaluation of

$$\zeta_{12} = \lim_{n\to\infty} {\rm Tr}\big(\mathbf{R}\mathbf{S}_{12}(n-1)\big), \tag{43}$$

where

$$\mathbf{S}_{12}(n-1) \triangleq {\rm E}\{\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\}. \tag{44}$$

Although we are more interested in convex combinations of algorithms with different tracking capabilities as LMS with RLS or CMA with SWA, we also obtain analytical expressions for ζ_{12}, considering the combinations of two RLSs, two CMAs, and two SWAs. Although our method also applies to the combination of two LMS filters, this case was already analyzed in [12] using the feedback method.

Using the linear regression model of (3) for the desired response $d(n)$, the error $e_i(n),\ i = 1, 2$ defined in (2) can be rewritten as a function of the *a priori* error $e_{a,i}(n)$ and of the disturbance $v(n)$, i.e.,

$$e_i(n) = e_{a,i}(n) + v(n). \tag{45}$$

To make the performance analysis of the supervised convex combinations more tractable, we assume that

A5. the sequence $\{v(n)\}$ is independent of $\{\mathbf{u}(l)\}$ for all n and l. This assumption is widely used in the analysis of adaptive algorithms [3, p. 284], [12, Sec. II].

Comparing (45) to (24), the errors for supervised or blind filters satisfy

$$e_i(n) = \kappa(n)e_{a,i}(n) + \varphi(n), \tag{46}$$

where $\kappa(n) = 1$ 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, $\mathrm{E}\{\varphi(n)\} = 0$, and $\varphi(n)$ and $\kappa(n)$ are assumed to be independent of $\widetilde{\mathbf{w}}(n-1)$ in steady-state. Denoting $\bar{\kappa} = \mathrm{E}\{\kappa(n)\}$ and $\sigma_\varphi^2 = \mathrm{E}\{\varphi^2(n)\}$, R2 and A5 imply that

$$\mathrm{E}\{\kappa(n)u^2(n)\} \approx \bar{\kappa}\mathrm{E}\{u^2(n)\}, \tag{47}$$

and

$$\mathrm{E}\{\varphi^2(n)u^2(n)\} \approx \sigma_\varphi^2\mathrm{E}\{u^2(n)\}. \tag{48}$$

Subtracting both sides of (1) from $\mathbf{w}_\mathrm{o}(n)$, using the model (10) and replacing $e_i(n)$ by (46), we arrive at

$$\begin{aligned}\widetilde{\mathbf{w}}_i(n) =& [\mathbf{I} - \rho_i\kappa(n)\mathbf{M}_i(n)\mathbf{u}(n)\mathbf{u}^T(n)]\widetilde{\mathbf{w}}_i(n-1)\\ &- \rho_i\varphi(n)\mathbf{M}_i(n)\mathbf{u}(n) + \mathbf{q}(n).\end{aligned} \tag{49}$$

In order to obtain ζ_{12}, we multiply (49) with $i = 1$ by its transpose with $i = 2$ and take the expectations of both sides. Then, assuming that $\mathbf{q}(n)$ is independent of the initial conditions and of $\mathbf{u}(n)$, after some algebra, we get

$$\begin{aligned}\mathrm{E}\{\widetilde{\mathbf{w}}_1(n)\widetilde{\mathbf{w}}_2^T(n)\} &= \mathrm{E}\{\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\}\\ &-\rho_2\mathrm{E}\{\kappa(n)\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}_2(n)\}\\ &-\rho_1\mathrm{E}\{\kappa(n)\mathbf{M}_1(n)\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\}\\ &+\rho_1\rho_2\mathrm{E}\{\kappa^2(n)\mathbf{M}_1(n)\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}_1(n-1)\\ &\qquad\times\widetilde{\mathbf{w}}_2^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}_2(n)\}\\ &+\rho_1\rho_2\mathrm{E}\{\varphi^2(n)\mathbf{M}_1(n)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}_2(n)\}\\ &-\rho_2\mathrm{E}\{\varphi(n)\widetilde{\mathbf{w}}_1(n-1)\mathbf{u}^T(n)\mathbf{M}_2(n)\}\\ &-\rho_1\mathrm{E}\{\varphi(n)\mathbf{M}_1(n)\mathbf{u}(n)\widetilde{\mathbf{w}}_2^T(n-1)\}\\ &+\rho^2\mathrm{E}\{\kappa(n)\varphi(n)\mathbf{M}_1(n)\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}_1(n-1)\mathbf{u}^T(n)\mathbf{M}_2(n)\}\\ &+\rho^2\mathrm{E}\{\kappa(n)\varphi(n)\mathbf{M}_1(n)\mathbf{u}(n)\widetilde{\mathbf{w}}_2^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{M}_2(n)\}\\ &+\mathrm{E}\{\mathbf{q}(n)\mathbf{q}^T(n)\}.\end{aligned} \tag{50}$$

Using the same arguments as in Section III-B, i.e., A1-A4/A5, (47), (48), R1, and R3, (50) can be simplified to

$$\begin{aligned}\mathbf{S}_{12}(n) &= \mathbf{S}_{12}(n-1)\\ &\quad - \rho_2\bar{\kappa}\,\mathbf{S}_{12}(n-1)\,\mathbf{R}\,\mathrm{E}\{\mathbf{M}_2(n)\}\\ &\quad - \rho_1\bar{\kappa}\,\mathrm{E}\{\mathbf{M}_1(n)\}\,\mathbf{R}\,\mathbf{S}_{12}(n-1)\\ &\quad + \rho_1\rho_2\sigma_\varphi^2\,\mathrm{E}\{\mathbf{M}_1(n)\}\,\mathbf{R}\,\mathrm{E}\{\mathbf{M}_2(n)\} + \mathbf{Q}.\end{aligned} \tag{51}$$

When $n \to \infty$, for the combination of μ_1-LMS with μ_2-LMS or μ_1-CMA with μ_2-CMA, we obtain

$$\bar{\kappa}\left[\mu_2\mathbf{S}_{12}(\infty)\mathbf{R} + \mu_1\mathbf{R}\mathbf{S}_{12}(\infty)\right] = \mu_1\mu_2\,\sigma_\varphi^2\,\mathbf{R} + \mathbf{Q}. \tag{52}$$

Taking the trace of both sides (52), we arrive at

$$\zeta_{12} = \mathrm{Tr}\big(\mathbf{R}\mathbf{S}_{12}(\infty)\big) = \frac{\mu_1\mu_2\mathrm{Tr}(\mathbf{R})\,\sigma_\varphi^2 + \mathrm{Tr}(\mathbf{Q})}{\bar{\kappa}\,(\mu_1+\mu_2)}. \tag{53}$$

Similar expressions were also obtained using the feedback analysis in [12] for the combination of two LMS filters and in [21] for the combination of two CMAs.

For the combination of λ_1-RLS with λ_2-RLS or λ_1-SWA with λ_2-SWA, replacing $\mathbf{M}_i(n) = \widehat{\mathbf{R}}_i^{-1}(n),\ i = 1, 2$ by their means, we obtain in the steady-state

$$\begin{aligned}&\big[(1-\lambda_1) + (1-\lambda_2)\big]\mathbf{S}_{12}(\infty)\\ &\qquad = \rho_1\rho_2(1-\lambda_1)(1-\lambda_2)\sigma_\varphi^2\,\mathbf{R}^{-1} + \mathbf{Q}.\end{aligned} \tag{54}$$

Multiplying both sides of (54) by $\mathbf{R}$ and taking the trace, we arrive at

$$\zeta_{12} = \frac{\rho_1\rho_2(1-\lambda_1)(1-\lambda_2)M\,\sigma_\varphi^2 + \mathrm{Tr}(\mathbf{Q}\mathbf{R})}{(1-\lambda_1) + (1-\lambda_2)}. \tag{55}$$

Finally, for the combination of λ_1-RLS with μ_2-LMS or λ_1-SWA with μ_2-CMA, we have in the steady-state

$$\mathbf{S}_{12}(\infty)\boldsymbol{\Gamma} = \rho_1\,\mu_2\,(1-\lambda_1)\,\sigma_\varphi^2\,\mathbf{I} + \mathbf{Q}. \tag{56}$$

where

$$\boldsymbol{\Gamma} = (1-\lambda_1)\mathbf{I} + \bar{\kappa}\,\mu_2\mathbf{R}. \tag{57}$$

Post-multiplying both sides of (56) by $\boldsymbol{\Sigma} \triangleq \boldsymbol{\Gamma}^{-1}\mathbf{R}$ and taking the trace, we obtain

$$\zeta_{12} = \rho_1\mu_2(1-\lambda_1)\sigma_\varphi^2\,\mathrm{Tr}\big(\boldsymbol{\Sigma}\big) + \mathrm{Tr}\big(\mathbf{Q}\boldsymbol{\Sigma}\big). \tag{58}$$

The results of (53), (55) and (58) are summarized in Table III for all the convex combinations of adaptive algorithms considered here.

IV. Simulation results

To verify the behavior of the proposed scheme and the validity of the tracking analysis in the supervised case, we consider a system identification application. The initial optimal solution is formed with $M = 5$ independent random values between 0 and 1, and is given by

$$\mathbf{w}_\mathrm{o}^T(0) = \begin{bmatrix} 0.5349 & 0.9527 & -0.9620 & -0.0158 & -0.1254 \end{bmatrix}.$$

The regressor $\mathbf{u}(n)$ is obtained from a process $u(n)$ as $\mathbf{u}^T(n) = [\, u(n) \; u(n{-}1) \; \cdots \; u(n{-}4) \,]$, where $u(n)$ is generated with a first-order autoregressive model, whose transfer function is $\sqrt{1-b^2}/(1-bz^{-1})$. 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.

Figure 5 shows the EMSE and $\mathrm{E}\{\eta(n)\}$ estimated from the ensemble-average of 500 independent runs for λ_1-RLS ($\lambda_1 = 0.92$), μ_2-LMS ($\mu_2 = 0.04$), and their convex combination ($\mu_\alpha = 100$, $\alpha^+ = 4$). To facilitate the visualization, the EMSE curves were filtered by a moving-average filter with 32 coefficients. At every 10^5 iterations, the nonstationary environment, represented by the matrix $\mathbf{Q}$, is changed. During the first 10^5 iterations, $\mathbf{Q} = 10^{-6}\mathbf{R}$, LMS presents better tracking performance than that of RLS and the combination performs close to LMS with $\mathrm{E}\{\eta(n)\} \approx 0$. When the matrix $\mathbf{Q}$ becomes equal to $10^{-6}\mathbf{R}^{-1}$, this behavior changes: although RLS is slightly better than LMS, the combination performs better than both of them and $\mathrm{E}\{\eta(n)\} \approx 0.42$. For $\mathbf{Q} = 2 \times 10^{-5}\mathbf{R}$, the combination switches back to LMS and $\mathrm{E}\{\eta(n)\} \approx 0$. Finally, for $\mathbf{Q} = 2 \times 10^{-5}\mathbf{R}^{-1}$, the performance of RLS is better and the combination behaves close to it, with $\mathrm{E}\{\eta(n)\} \approx 0.92$. The dashed lines in Figure 5-(a) show the predicted values of ζ for the combination, which are in a good agreement with the experimental results. Note also that the small disagreement between our model and the simulations for $n > 3 \times 10^5$ is due to an imprecision in the model for RLS.

Figure 6 shows the EMSE for different values of c^2, considering theoretical and experimental results for λ_1-RLS ($\lambda_1 = 0.92$), μ_2-LMS ($\mu_2 = 0.04$), and their convex combination ($\mu_\alpha = 100$, $\alpha^+ = 4$). We assume ensemble-average of 100 independent runs and three different nonstationary environments: $\mathbf{Q} = \mathrm{c}^2\mathbf{I}$, $\mathbf{Q} = \mathrm{c}^2\mathbf{R}$, and $\mathbf{Q} = \mathrm{c}^2\mathbf{R}^{-1}$. Good agreement between analysis and simulation can be observed for all kinds of considered nonstationary environments. Note that the small disagreement between our model and the simulations observed in Figure 6 for large c^2 is due to an imprecision in the model for RLS: this can be seen by comparing the theoretical and simulation curves for RLS alone.

In Figure 7, we show the EMSE for different values of c^2, considering theoretical and experimental results for λ_1-RLS ($\lambda_1 = 0.92$), λ_2-RLS ($\lambda_2 = 0.995$), and their convex combination ($\mu_\alpha = 100$, $\alpha^+ = 4$). For the simulations, each point is an ensemble-average of 100 independent runs with $\mathbf{Q} = \mathrm{c}^2\mathbf{R}$. Again, we observe good agreement between theoretical and experimental EMSE. Such agreement also occurs for other kinds of nonstationary environments, with $\mathbf{Q} = \mathrm{c}^2\mathbf{I}$ or $\mathbf{Q} = \mathrm{c}^2\mathbf{R}^{-1}$.

In the blind equalization case, we assume 4-PAM (pulse amplitude modulation) with statistics $\mathrm{E}\{a^6(n)\} = 365$, $\mathrm{E}\{a^2(n)\} = 5$, $r^a = 8.2$ and channel coefficients[1] $[0.1, 0.3, 1, -0.1, 0.5, 0.2]$ [22]. In the combinations, each component filter has $M = 4$ coefficients as a $T/2$-fractionally spaced equalizer and is initialized with only one non-null element in the second position.

Figure 8 shows the EMSE for different values of c^2, considering theoretical and experimental results for λ_1-SWA ($\lambda_1 = 0.99$), μ_2-CMA ($\mu_2 = 10^{-4}$), and their convex combination ($\mu_\alpha = 0.1$, $\alpha^+ = 4$). Again, each point is an ensemble-average of 100 independent runs and we present results for three different nonstationary environments: $\mathbf{Q} = \mathrm{c}^2\mathbf{I}$, $\mathbf{Q} = \mathrm{c}^2\mathbf{R}$, and $\mathbf{Q} = \mathrm{c}^2\mathbf{R}^{-\mathbf{1}}$. Although the agreement between analysis and simulation is not as good as in the supervised case, the predicted values model reasonably well the overall behavior of the algorithms, independently of the nonstationary environment. Note that a difference of a few dB is common in models for blind algorithms, due to the strong assumptions necessary for the analysis.

Figures 9 and 10 show the EMSE for different values of c^2, considering respectively theoretical and experimental results for μ_1-CMA, μ_2-CMA, and their convex combination with $\mathbf{Q} = \mathrm{c}^2\mathbf{R}$, and for λ_1-SWA, λ_2-SWA, and their convex combination, with $\mathbf{Q} = \mathrm{c}^2\mathbf{R}^{-1}$. We assume $\mu_1 = 10^{-3}$, $\mu_2 = 10^{-4}$, $\lambda_1 = 0.99$, $\lambda_2 = 0.999$, $\mu_\alpha = 0.1$, $\alpha^+ = 4$, and ensemble-average of 100 independent runs. The agreement between analysis and simulation for the combination of two SWAs is better than that for the combination of two CMAs. This occurs because the steady-state model of SWA shows a better agreement with experimental results than that of CMA, as shown in the figures 9 and 10. This behavior also happens for other kinds of nonstationary environments.

V. CONCLUSION

In this paper we proposed the convex combination of filters of different families (gradient-based and Hessian-based) to achieve an overall filter with superior tracking performance. In addition, we presented a unified model for the convex combination of several different adaptive algorithms, both supervised and unsupervised (blind). Our models for the combination of two RLSs, two CMAs, two SWAs, one LMS with one RLS, and one CMA with one SWA are novel, and show good agreement with simulations.

[1]Although we assume channels with long impulse response to justify R2, our model agrees well with simulations even for a rather short channel such as this one. Good agreement was also observed with longer channels.

Appendix A

Remarks R2 and R3

Considering an FIR channel with impulse response $[\,h_0, h_1, \cdots, h_{L-1}\,]$, its output in a noise-free environment is given by

$$u(n) = \sum_{\ell=0}^{L-1} h_\ell \, a(n-\ell).$$

In the analysis, we use $\mathrm{E}\{\gamma(n)u^2(n)\} \approx \bar{\gamma}\,\mathrm{E}\{u^2(n)\}$. From (5), this approximation reduces to say that

$$\mathrm{E}\{a^2(n-\tau_d)u^2(n)\} \approx \mathrm{E}\{a^2(n-\tau_d)\}\mathrm{E}\{u^2(n)\}. \tag{59}$$

We check the validity of this approximation, evaluating both sides of (59). Evaluating the term on the right-hand side of (59), using the fact that $\{a(n)\}$ is i.i.d., we arrive at

$$\begin{aligned}
&\mathrm{E}\{a^2(n-\tau_d)\mathrm{E}\{u^2(n)\}\\
&= \mathrm{E}\{a^2(n-\tau_d)\}\mathrm{E}\left\{\sum_{\ell=0}^{L-1}\sum_{m=0}^{L-1} h_\ell h_m a(n-\ell)a(n-m)\right\}\\
&= \mathrm{E}\{a^2(n)\}^2 \sum_{k=0}^{L-1} h_k^2.
\end{aligned} \tag{60}$$

Analogously for the term on the left-hand side of (59), we obtain

$$\begin{aligned}
&\mathrm{E}\{a^2(n-\tau_d)u^2(n)\}\\
&= \mathrm{E}\left\{a^2(n-\tau_d)\sum_{\ell=0}^{L-1}\sum_{m=0}^{L-1} h_\ell h_m a(n-\ell)a(n-m)\right\}\\
&= h_{\tau_d}^2 \mathrm{E}\{a^4(n)\} + \mathrm{E}\{a^2(n)\}^2 \sum_{k=0,\, k\neq\tau_d}^{L-1} h_k^2.
\end{aligned} \tag{61}$$

Replacing $\mathrm{E}\{a^4(n)\}$ by $(1+\varepsilon_4)\mathrm{E}\{a^2(n)\}^2$ in (61), we get

$$\mathrm{E}\{a^2(n-\tau_d)u^2(n)\} = \mathrm{E}\{a^2(n)\}^2 \left(\varepsilon_4\, h_{\tau_d}^2 + \sum_{k=0}^{L-1} h_k^2\right). \tag{62}$$

The value of ε_4 is nonnegative and depends on the constellation. For example, $\varepsilon_4 = 0$ for the constant-modulus constellation of 2-PAM, $\varepsilon_4 = 0.64$ for 4-PAM, and $\varepsilon_4 = 0.73$ for 6-PAM. Assuming that L is sufficiently large such that $\varepsilon\, h_{\tau_d}^2 \ll \sum_{k=0}^{L-1} h_k^2$, (62) can be approximated by (60). In this case, (59) and (28) are satisfied.

To show the validity of $\mathrm{E}\{\beta^2(n)u^2(n)\} \approx \sigma_\beta^2 \mathrm{E}\{u^2(n)\}$, we use a similar argument. Recalling that $\{a(n)\}$ is an i.i.d. sequence, we get

$$\begin{aligned}
\mathrm{E}\{\beta^2(n)u^2(n)\} = {}& h_{\tau_d}^2 \mathrm{E}\{a^8(n) - 2r^a a^6(n) + (r^a)^2 a^4(n)\}\\
&+ \mathrm{E}\{\beta^2(n)\}\mathrm{E}\{a^2(n)\} \sum_{k=0,\, k\neq\tau_d}^{L-1} h_k^2.
\end{aligned} \tag{63}$$

Replacing $\mathrm{E}\{a^k(n)\} = (1+\varepsilon_k)\mathrm{E}\{a^{k-2}(n)\}\mathrm{E}\{a^2(n)\}$ in (63), we arrive at

$$\mathrm{E}\{\beta^2(n)u^2(n)\} = h_{\tau_d}^2 \mathrm{E}\{a^2(n)\}\mathrm{E}\Big\{\varepsilon_8 a^6(n) - \varepsilon_6(2r^a)a^4(n) + \varepsilon_4(r^a)^2a^2(n)\Big\} + \mathrm{E}\{\beta^2(n)\}\mathrm{E}\{a^2(n)\}\sum_{k=0}^{L-1} h_k^2. \quad (64)$$

Again, ε_i, $i = 4, 6, 8$ are nonnegative and depend on the constellation. If the channel is such that the first term of the right-hand side of (64) can be disregarded with respect to the second term, (29) of R2 will be satisfied.

To show that $\mathrm{E}\{\widetilde{\mathbf{w}}(n)\} = \mathbf{0}$ in the steady-state, we remark that

$$\begin{aligned}\mathrm{E}\{\beta(n)u(n)\} &= \mathrm{E}\Big\{[-a^3(n-\tau_d) + r^a a(n-\tau_d)] \cdot \sum_{\ell=0}^{L-1} h_\ell\, a(n-\ell)\Big\} \\ &= h_{\tau_d}\left[-\mathrm{E}\{a^4(n)\} + r^a\mathrm{E}\{a^2(n)\}\right] = 0. \end{aligned} \quad (65)$$

This is an exact relation. For $\mathbf{M}(n) \neq \mathbf{I}$, we use R1 to approximate $\mathrm{E}\{\beta(n)\mathbf{M}(n)\mathbf{u}(n)\} = \mathrm{E}\{\mathbf{M}(n)\}\mathrm{E}\{\beta(n)\mathbf{u}(n)\}$. Then, taking the expectations of both sides of (26) and using A2, we obtain

$$\begin{aligned}\mathrm{E}\{\widetilde{\mathbf{w}}(n)\} \approx{}& \mathrm{E}\{\widetilde{\mathbf{w}}(n-1)\} \\ &- \rho\mathrm{E}\{\mathbf{M}(n)\}\mathrm{E}\{\gamma(n)\mathbf{u}(n)\mathbf{u}^T(n)\}\mathrm{E}\{\widetilde{\mathbf{w}}(n-1)\} \\ &- \rho\mathrm{E}\{\mathbf{M}(n)\}\mathrm{E}\{\beta(n)\mathbf{u}(n)\} + \mathrm{E}\{\mathbf{q}(n)\} = \\ &= [\mathbf{I} - \rho\mathrm{E}\{\mathbf{M}(n)\}\mathrm{E}\{\gamma(n)\mathbf{u}(n)\mathbf{u}^T(n)\}]\,\mathrm{E}\{\widetilde{\mathbf{w}}(n-1)\}, \end{aligned} \quad (66)$$

from $\mathrm{E}\{\mathbf{q}(n)\} = \mathbf{0}$ and using (65). Since $\mathrm{E}\{\gamma(n)u^2(n)\} \neq 0$, we get $\mathrm{E}\{\widetilde{\mathbf{w}}(n)\} \to \mathbf{0}$ for sufficiently small ρ.

REFERENCES

[1] J. Proakis, *Digital Communications*, McGraw-Hill, NY, 4th edition, 2001.

[2] C. Breining et al., "Acoustic echo control – an application of very-high-order adaptive filters," *Signal Processing Magazine*, pp. 42–69, Jul. 1999.

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

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

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

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

[7] J. R. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-28, pp. 334–358, Apr. 1983.

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

[9] M. T. M. Silva and M. D. Miranda, “Tracking issues of some blind equalization algorithms,” *IEEE Signal Processing Letters*, vol. 11, pp. 760–763, Sep. 2004.

[10] M. Martínez-Ramón, J. Arenas-García, A. Návia-Vazquez, and A. R. Figueiras-Vidal, “An adaptive combination of adaptive filters for plant identification,” in *Proc. of 14th International Conference on Digital Signal Processing (DSP’2002)*. IEEE, 2002, vol. 2, pp. 1195–1198.

[11] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, “Plant identification via adaptive combination of transversal filters,” *Signal Processing*, vol. 86, pp. 2430–2438, Sep. 2006.

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

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

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

[15] M. Niedźwiecki, “Multiple-model approach to finite memory adaptive filtering,” *IEEE Transactions on Signal Processing*, vol. 40, pp. 470–473, Feb. 1992.

[16] S. S. Kozat, A. C. Singer, and G. Zeitler, “Universal piecewise linear prediction via context trees,” *IEEE Transactions on Signal Processing*, vol. 55, pp. 3730–3745, Jul. 2007.

[17] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, “Tracking properties of a convex combination of two adaptive filters,” in *Proc. of IEEE 13th Workshop on Statistical Signal Processing (SSP’05)*. IEEE, 2005, pp. 109–114.

[18] J. Arenas-García and A. R. Figueiras-Vidal, “Improved blind equalization via adaptive combination of constant modulus algorithms,” in *Proc. of ICASSP’06*. IEEE, 2006, vol. III, pp. 756–759.

[19] J. Arenas-García and A. R. Figueiras-Vidal, “Adaptive combination of normalized filters for robust system identification,” *Electronics Letters*, vol. 41, no. 15, pp. 874–875, Jul. 2005.

[20] M. T. M. Silva and V. H. Nascimento, “Convex combination of adaptive filters with different tracking capabilities,” in *Proc. of ICASSP’07*. IEEE, 2007, vol. III, pp. 925–928.

[21] M. T. M. Silva and V. H. Nascimento, “Convex combination of blind adaptive equalizers with different tracking capabilities,” in *Proc. of ICASSP’07*. IEEE, 2007, vol. III, pp. 457–460.

[22] J. Mai and A. H. Sayed, “A feedback approach to the steady-state performance of fractionally spaced blind adaptive equalizers,” *IEEE Transactions on Signal Processing*, vol. 48, pp. 80–91, Jan. 2000.

[23] N. R. Yousef and A. H. Sayed, “A feedback analysis of the tracking performance of blind adaptive equalization algorithms,” in *Proceedings of IEEE Conference on Decision and Control*, Dec. 1999, vol. 1, pp. 174–179.

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

[25] T. Aboulnasr and K. Mayyas, “A robust variable step-size LMS-type algorithm: analysis and simulations,” *IEEE Transactions on Signal Processing*, vol. 45, pp. 631–639, Mar. 1997.

[26] P. A Regalia, “On the equivalence between the Godard and Shalvi-Weinstein schemes of blind equalization,” *Signal Processing*, vol. 73, pp. 185–190, 1999.

[27] J. R. Treichler, I. Fijalkow, and C. R. Johnson Jr., “Fractionally spaced equalizers,” *IEEE Signal Processing Magazine*, vol. 13, pp. 65–81, May 1996.

[28] Y. Li and Z. Ding, “Global convergence of fractionally spaced Godard (CMA) adaptive equalizers,” *IEEE Transactions on Signal Processing*, vol. 44, pp. 818–826, Apr. 1996.

[29] 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.

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

[31] J. Minkoff, “Comment: on the unnecessary assumption of statistical independence between reference signal and filter weights in feedforward adaptive systems,” *IEEE Transactions on Signal Processing*, vol. 49, pp. 1109, May 2001.

[32] I. Fijalkow, C. E. Manlove, and C. R. Johnson, “Adaptive fractionally space blind CMA equalization: Excess MSE,” *IEEE Transactions on Signal Processing*, vol. 46, pp. 227–231, Jan. 1998.

[33] C. B. Papadias and D. T. M. Slock, “Normalized sliding window constant modulus and decision-direct algorithms: a link between blind equalization and classical adaptive filtering,” *IEEE Transactions on Signal Processing*, vol. 45, pp. 231–235, Jan. 1997.

Magno T. M. Silva (M'05) was born in São Sebastião do Paraíso, Brazil, in 1975. He received his Bachelor's degree in 1999, his Master's degree in 2001, and his Doctorate in 2005, all in Electrical Engineering from Escola Politécnica, University of São Paulo, Brazil. From February 2005 to July 2006 he was an Assistant Professor at Mackenzie Presbyterian University, Brazil. He is currently an Assistant Professor in the Department of Electronic Systems Engineering, Escola Politécnica, University of São Paulo. His research interests are in linear and nonlinear adaptive filtering.

Vítor H. Nascimento (M'90) was born in São Paulo, Brazil. He obtained the B.S. and M.S. degrees in Electrical Engineering from the University of São Paulo, in 1989 and 1992, respectively, and the Ph.D. degree from the University of California, Los Angeles, in 1999. From 1990 to 1994 he was a Lecturer at the Univ. of São Paulo, and in 1999 he joined the faculty at the same school, where he is now an Associate Professor. In 2002, he received the IEEE SPS 2002 Best Paper Award. He served as an Associate Editor for Signal Processing Letters from 2003 to 2005 and is currently an Associate Editor for Transactions on Signal Processing and the Eurasip Journal on Advances in Signal Processing, and a member of the IEEE-SPS Signal Processing Theory and Methods Technical Committee. His research interests include adaptive filtering theory and applications, linear and nonlinear estimation, and applied linear algebra.

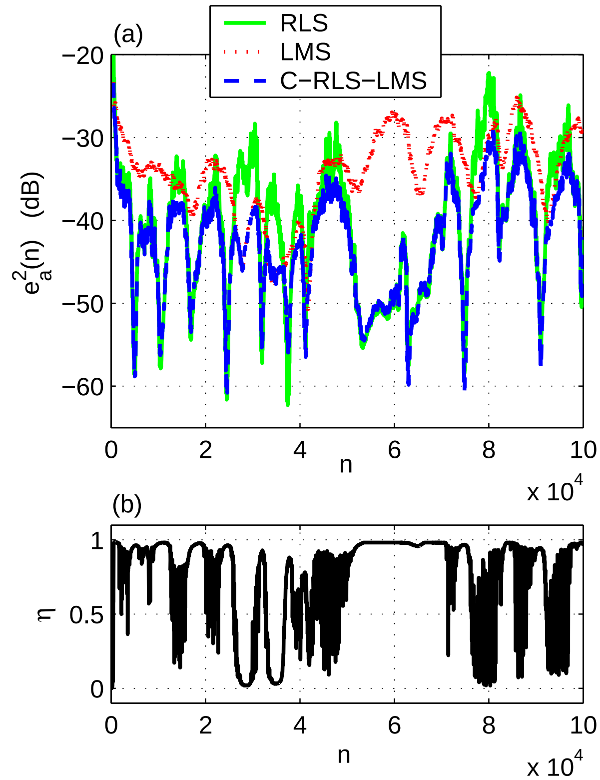


Fig. 1. Time-variant channel identification. (a) Squared *a priori* errors for RLS ($\lambda = 0.995$), LMS ($\mu = 0.01$), and their convex combination C-RLS-LMS ($\mu_\alpha = 400, \alpha^+ = 4$); (b) Evolution of the mixing parameter. Input: correlated Gaussian noise, AR-1 model with pole at 0.99; measurement noise with $\sigma_v^2 = 10^{-5}$; Rayleigh fading channel (5 coefficients, symbol period $T = 0.8\mu$s, maximum Doppler spread $f_D = 50$ Hz).

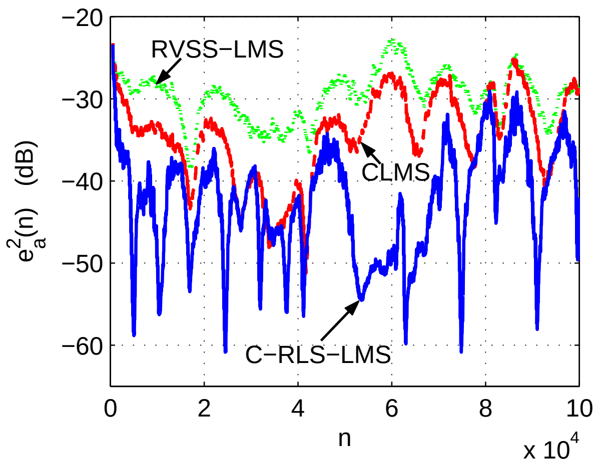


Fig. 2. Same example of Fig. 1. Squared *a priori* errors for C-RLS-LMS, CLMS ($\mu_1 = 0.01, \mu_2 = 0.001, \mu_\alpha = 400, \alpha^+ = 4$), RVSS-LMS ($\mu_{\min} = \mu_2, \mu_{\max} = \mu_1, \mu_{\text{init}} = (\mu_1 + \mu_2)/2, \beta = 0.97$).

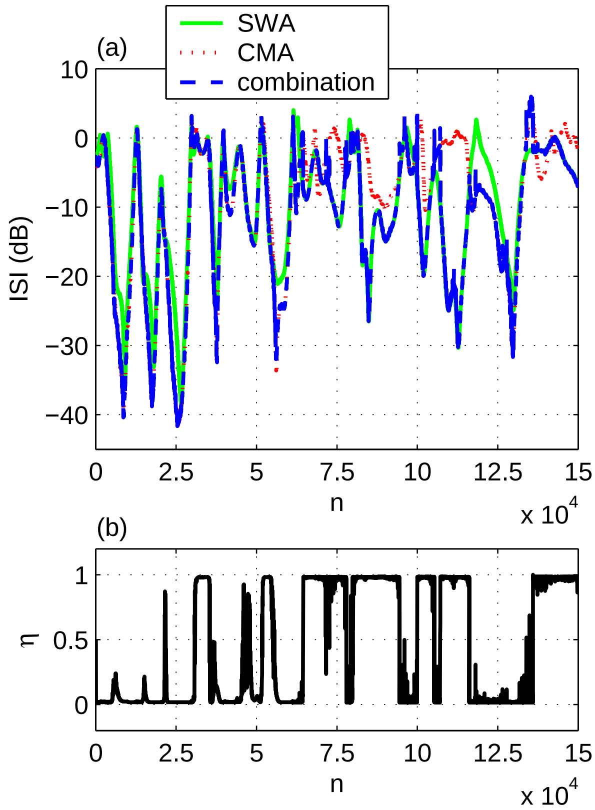
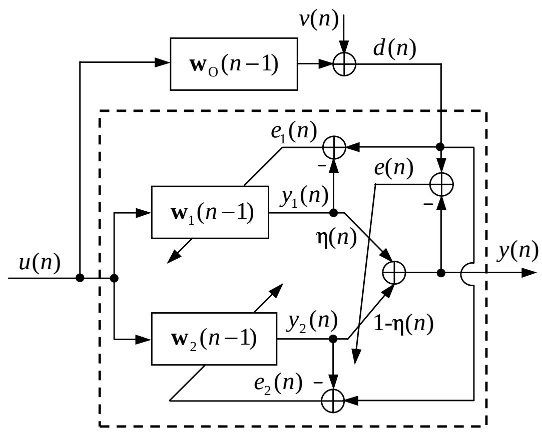
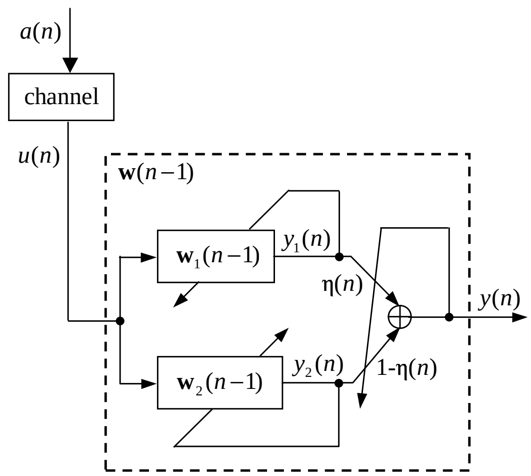


Fig. 3. Blind equalization. Residual intersymbol interference curves for CMA ($\mu = 2 \times 10^{-3}$), SWA ($\lambda = 0.999$), and their convex combination ($\mu_\alpha = 15$, $\alpha^+ = 4$); 2-PAM (pulse amplitude modulation); Rayleigh fading channel (3 coefficients, symbol period $T = 0.8\mu$s, maximum Doppler spread $f_D = 80$ Hz); SNR=30 dB; baud-rate equalizer with 11 coefficients.



(a)



(b)

Fig. 4. Adaptive convex combination of two transversal filters for (a) supervised filtering and (b) blind equalization.

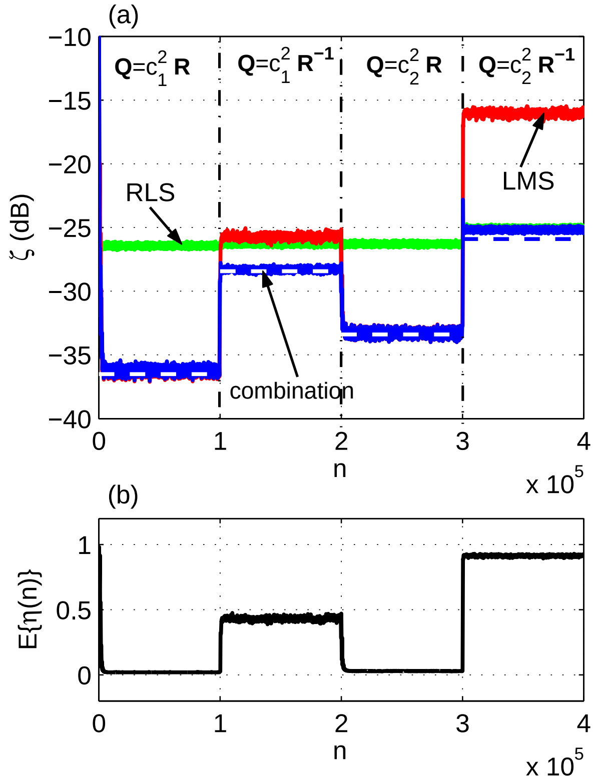


Fig. 5. (a) EMSE for λ_1-RLS, μ_2-LMS, and their convex combination; (b) ensemble-average of $\eta(n)$; $\lambda_1 = 0.92$, $\mu_2 = 0.04$, $\mu_\alpha = 100$, $\alpha^+ = 4$, $\mathrm{c}_1^2 = 2 \times 10^{-6}$, $\mathrm{c}_2^2 = 2 \times 10^{-5}$, $b = 0.8$; mean of 500 independent runs. In (a), the dashed lines represent the predicted values of ζ for the convex combination.

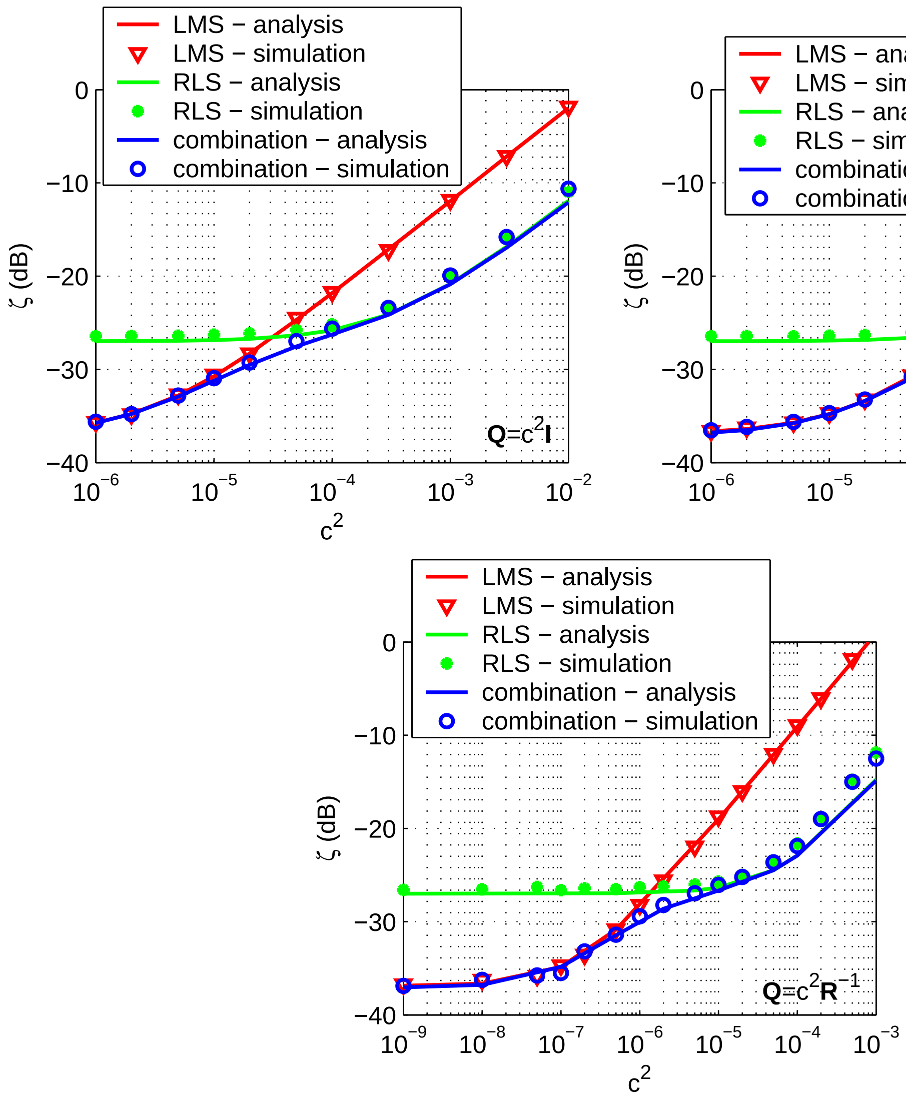


Fig. 6. EMSE for different values of c^2 considering theoretical and experimental results for the convex combination and theoretical results for LMS and RLS, with $\lambda_1 = 0.92$, $\epsilon = 12.5$, $\mu_2 = 0.04$, $\mu_\alpha = 100$, $\alpha^+ = 4$, $b = 0.8$; mean of 100 independent runs.

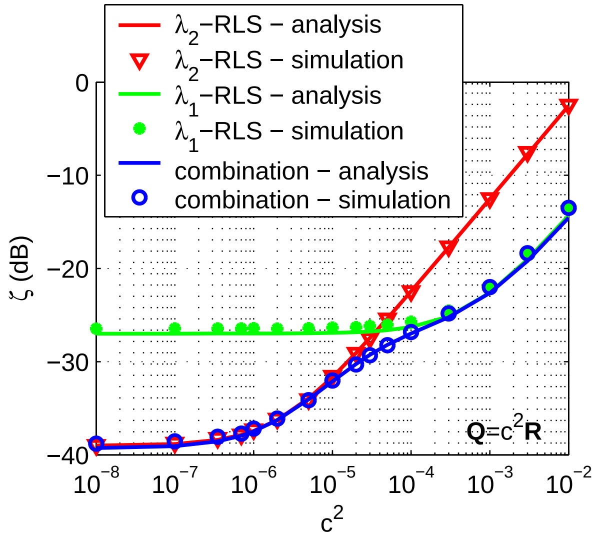


Fig. 7. EMSE for different values of c^2 considering theoretical and experimental results for the convex combination and theoretical results for λ_2-RLS and λ_1-RLS, with $\lambda_1 = 0.92$, $\lambda_2 = 0.995$, $\epsilon = 12.5$, $\mu_\alpha = 100$, $\alpha^+ = 4$, $b = 0.8$; mean of 100 independent runs.

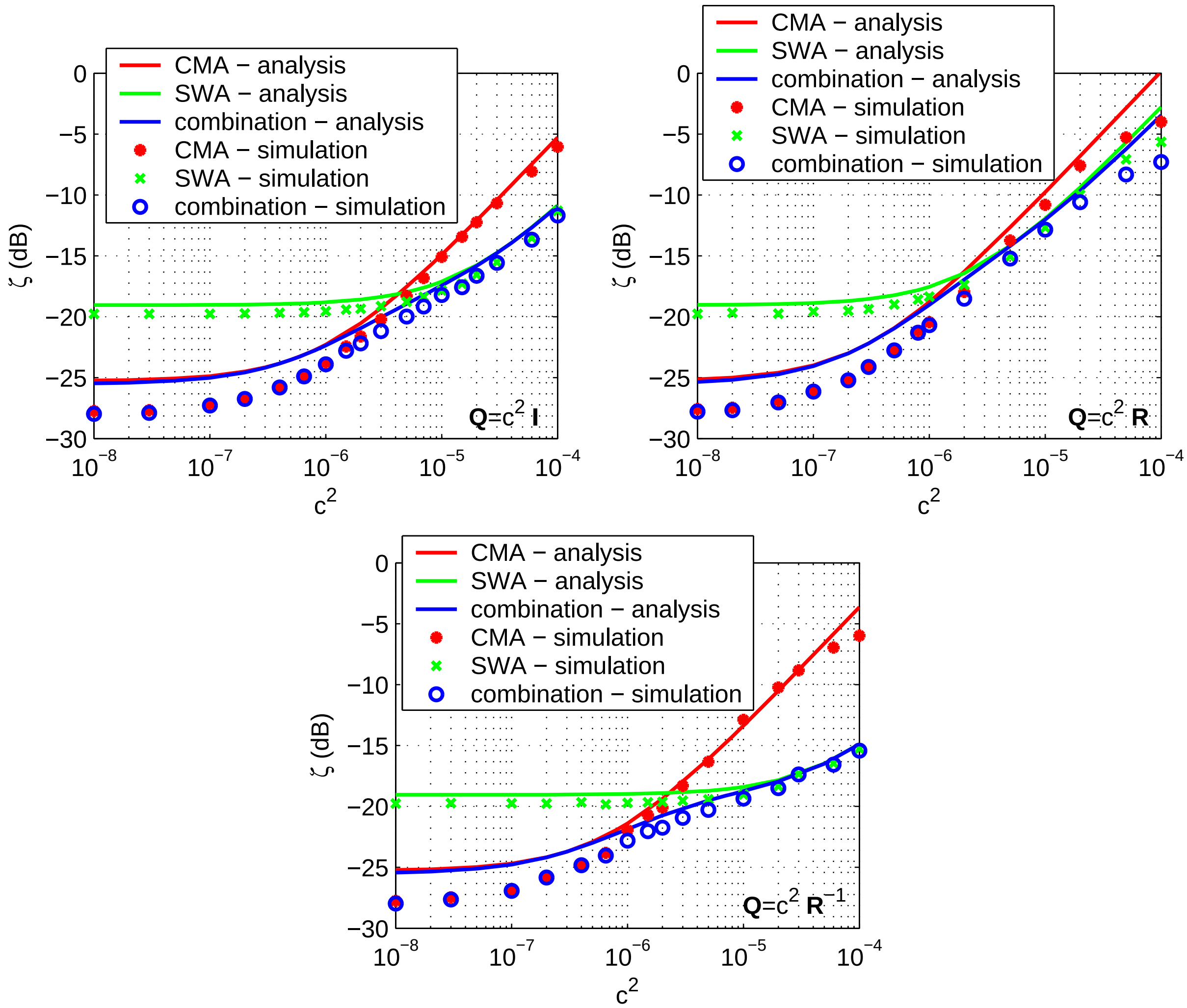


Fig. 8. EMSE for different values of c^2, considering theoretical and experimental results for CMA, SWA, and their convex combination, with $\lambda_1 = 0.99$, $\mu_2 = 10^{-4}$, $\mu_\alpha = 0.1$, $\alpha^+ = 4$; $M = 4$; 4-PAM; channel $[0.1,\ 0.3,\ 1,\ -0.1,\ 0.5,\ 0.2]$; $T/2$- fractionally-spaced equalizer; mean of 100 independent runs.

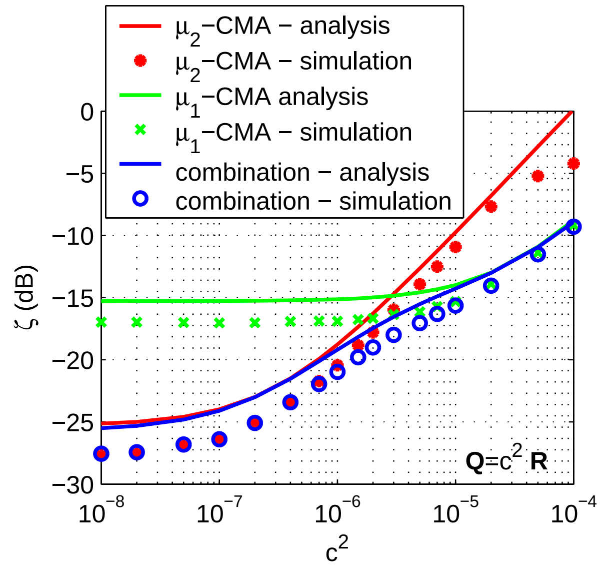


Fig. 9. EMSE for different values of c^2, considering theoretical and experimental results for μ_1-CMA, μ_2-CMA, and their convex combination, with $\mu_1 = 10^{-3}$, $\mu_2 = 10^{-4}$, $\mu_\alpha = 0.1$, $\alpha^+ = 4$; $M = 4$; 4-PAM; channel $[0.1,\, 0.3,\, 1,\, -0.1,\, 0.5,\, 0.2]$; $T/2$- fractionally-spaced equalizer; mean of 100 independent runs.

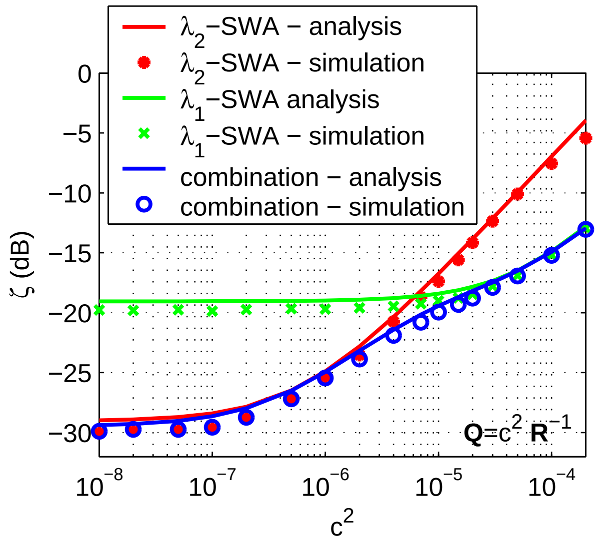


Fig. 10. EMSE for different values of c^2, considering theoretical and experimental results for λ_1-SWA, λ_2-SWA, and their convex combination, with $\lambda_1 = 0.99$, $\lambda_2 = 0.999$, $\mu_\alpha = 0.1$, $\alpha^+ = 4$; $M = 4$; 4-PAM; channel $[0.1,\, 0.3,\, 1,\, -0.1,\, 0.5,\, 0.2]$; $T/2$- fractionally-spaced equalizer; mean of 100 independent runs.

TABLE I

PARAMETERS OF THE CONSIDERED ALGORITHMS.

Alg.	ρ_i	$e_i(n)$	$\mathbf{M}_i^{-1}(n)$
LMS	μ_i	$d(n) - y_i(n)$	$\mathbf{I}$
CMA	μ_i	$[r^a - y_i^2(n)]y_i(n)$	
RLS	1	$d(n) - y_i(n)$	$\widehat{\mathbf{R}}_i(n) = \sum_{l=1}^{n} \lambda_i^{n-l}\mathbf{u}(l)\mathbf{u}^T(l)$
SWA	$1/\bar{\gamma}$	$[r^a - y_i^2(n)]y_i(n)$	

TABLE II

ANALYTICAL EXPRESSIONS FOR EMSE OF THE CONSIDERED ALGORITHMS.

Alg.	ζ
LMS	$\frac{\mu\sigma_v^2\mathrm{Tr}(\mathbf{R})+\mu^{-1}\mathrm{Tr}(\mathbf{Q})}{2}$
RLS	$\frac{\sigma_v^2(1-\lambda)M+(1-\lambda)^{-1}\mathrm{Tr}(\mathbf{QR})}{2}$
CMA	$\frac{\mu\sigma_\beta^2\mathrm{Tr}(\mathbf{R})+\mu^{-1}\mathrm{Tr}(\mathbf{Q})}{2\bar{\gamma}}$
SWA	$\frac{(\bar{\gamma})^{-1}\sigma_\beta^2(1-\lambda)M+\bar{\gamma}(1-\lambda)^{-1}\mathrm{Tr}(\mathbf{QR})}{2\bar{\gamma}}$

TABLE III

ANALYTICAL EXPRESSIONS FOR CROSS-EMSE OF THE CONSIDERED COMBINATIONS.

Combination	ζ_{12}
μ_1-LMS and μ_2-LMS	$\dfrac{\mu_1\mu_2\mathrm{Tr}(\mathbf{R})\sigma_v^2 + \mathrm{Tr}(\mathbf{Q})}{\mu_1+\mu_2}$
λ_1-RLS and λ_2-RLS	$\dfrac{(1-\lambda_1)(1-\lambda_2)M\sigma_v^2 + \mathrm{Tr}(\mathbf{QR})}{(1-\lambda_1)+(1-\lambda_2)}$
λ_1-RLS and μ_2-LMS	$\mu_2(1-\lambda_1)\,\sigma_v^2\,\mathrm{Tr}\big(\mathbf{\Sigma}\big) + \mathrm{Tr}\big(\mathbf{Q\Sigma}\big)$
μ_1-CMA and μ_2-CMA	$\dfrac{\mu_1\mu_2\mathrm{Tr}(\mathbf{R})\sigma_\beta^2 + \mathrm{Tr}(\mathbf{Q})}{\bar{\gamma}(\mu_1+\mu_2)}$
λ_1-SWA and λ_2-SWA	$\dfrac{(\bar{\gamma})^{-2}(1-\lambda_1)(1-\lambda_2)M\sigma_\beta^2 + \mathrm{Tr}(\mathbf{QR})}{(1-\lambda_1)+(1-\lambda_2)}$
λ_1-SWA and μ_2-CMA	$(\bar{\gamma})^{-1}\mu_2(1-\lambda_1)\sigma_\beta^2\mathrm{Tr}\big(\mathbf{\Sigma}\big)+\mathrm{Tr}\big(\mathbf{Q\Sigma}\big)$