A TRANSIENT ANALYSIS FOR THE CONVEX COMBINATION OF ADAPTIVE FILTERS

Vítor H. Nascimento†, *Magno T. M. Silva*†, *Renato Candido*†, *and Jerónimo Arenas-García*‡

† University of São Paulo, Brazil
{vitor, magno, renatocan}@lps.usp.br

‡ Univ. Carlos III de Madrid, Spain
jarenas@tsc.uc3m.es

ABSTRACT

Combination schemes are gaining attention as an interesting way to improve adaptive filter performance. In this paper we pay attention to a particular convex combination scheme with nonlinear adaptation that has recently been shown to be universal –i.e., to perform at least as the best component filter– in steady-state; however, no theoretical model for the transient has been provided yet. By relying on Taylor Series approximations of the nonlinearities, we propose a theoretical model for the transient behavior of such convex combinations. In particular, we provide expressions for the time evolution of the mean and the variance of the mixing parameter, as well as for the mean square overall filter convergence. The accuracy of the model is analyzed for the particular case of a combination of two LMS filters with different step sizes, explaining also how our results can help the designer to adjust the free parameters of the scheme.

***Index Terms*—** Adaptive filters, convex combination, transient analysis, LMS algorithm.

1. INTRODUCTION

Combinations of adaptive filters have recently attracted attention due to their ability to improve transient and steady-state performance of adaptive filters in stationary and non-stationary environments. The first algorithm to attract attention was [1], which proposed a convex combination of adaptive filters, and presented a model for the steady-state performance of combinations of two LMS (least-mean square) filters, as shown in Fig. 1. 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, restricted to the interval $[0,1]$ (thus the name *convex* combination), $y_i(n)$, $i=1,2$, are the outputs of the transversal filters, i.e., $y_i(n) = \mathbf{u}^T(n)\mathbf{w}_i(n)$, $\mathbf{u}(n), \mathbf{w}_i(n) \in \mathbb{R}^M$ are respectively the common regressor vector, and the weight vectors of each component filter. In order to restrict the mixing parameter to interval $[0,1]$ and to reduce gradient noise when $\eta \approx 0$ or $\eta \approx 1$, a nonlinear transformation and an auxiliary variable $a(n)$ are used. $\eta(n)$ is defined as

$$\eta(n) = \frac{1}{1+\mathrm{e}^{-a(n)}}, \tag{2}$$

where $a(n)$ is updated to minimize the square of the overall error $e(n) = d(n) - y(n)$:

$$a(n+1) = a(n) + \mu_a\big[y_1(n) - y_2(n)\big]e(n)\eta(n)\big[1-\eta(n)\big]. \tag{3}$$

In practice, $a(n)$ is restricted (by saturation of the above recursion) to an interval $[-a_+, a_+]$, since factor $\eta(n)\big[1-\eta(n)\big]$ in (3) would virtually stop adaptation if $a(n)$ were allowed to grow too much. A common choice in the literature is $a_+ = 4$.

The work of Nascimento, Silva and Candido was partly supported by by CNPq under Grants 303361/2004-2, 302633/2008-1, and 136050/2008-5; and FAPESP under Grants 2008/04828-5 and 2008/00773-1. The work of Arenas-García was partly supported by MEC project TEC2008-02473 and CAM project S-0505/TIC/0223.

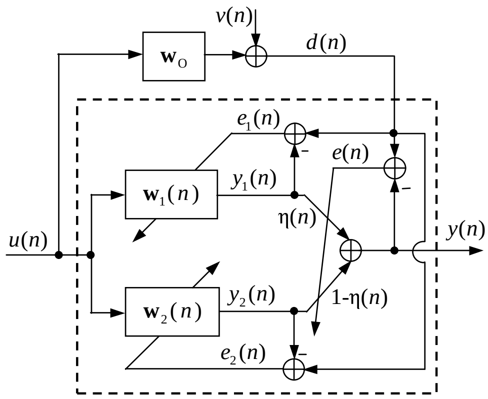

Fig. 1. Convex combination of two transversal adaptive filters.

Several variations and improvements on the original idea were later proposed, we give here only a short sample. Combinations of two RLS (recursive least-square) or CMA (constant modulus) algorithms were also proposed in other works, e.g. [2]. A combination approach to variable-length adaptive filtering was presented in [3]. The new method was also used in the solution of practical problems, as in [4]. More recently, a theoretical model for the combination of filters of different families, such as one LMS and one RLS, was proposed in [5]. All theses references provide models only for the steady-state mean-square error (MSE) of the combination, without models for the convergence of the mixing parameter to its optimum value ([1] does show that the combination converges to the desired solutions). The transient performance of the mixing parameter is an important issue, however. The combination will only perform well if $\eta(n)$ is able to choose, at each instant, the best combination of the component filters. A normalized algorithm for the estimation of the mixing parameter was proposed in [6] to improve the transient and tracking performance of $\eta(n)$, showing very good results. However, no theoretical model for the transient was provided.

A different combination approach was proposed more recently in [7]. In that paper, the mixing parameter is free to be negative or larger than one, so we have a more general *affine* combination, not necessarily convex. Two algorithms were proposed for the estimation of the mixing parameter, and their performance was investigated through simulations and comparisons with the optimum value. A theoretical model for the transient of the mixing parameters of a practical scheme for affine combinations was proposed in [8], which also extended the normalization idea of [6] to the affine combination algorithm. The adaptation rule for the mixing parameter in affine combinations is simpler than that for convex combinations, since there is no need for a nonlinear transformation to restrict $\eta(n)$ to interval $[0,1]$. However, convex combinations appear to be less sensitive to the choice of μ_a, and have a built-in mechanism to reduce the misadjustment due to the adaptation of the mixing parameter (the factor $\eta(1-\eta)$ in (3)).

In this paper we propose a theoretical model for the convex combination algorithm (2)–(3) that includes the transient behavior of $\eta(n)$. Large values of μ_a are necessary to guarantee that the combination will closely track the best filter at all times, but a too large value will increase the excess mean-square error of the combination. A theoretical model allows one to understand the influence of design parameters on performance, giving tools for the designer to correctly apply the algorithms.

2. APPROXIMATE UPDATE EQUATION

We assume for simplicity that all signals are real and that the input is stationary, although the model can be easily extended to non-stationary signals, as was done in [5], and also to complex signals. Since the desired signal $d(n)$ and regressor vector are assumed stationary, they are related through $d(n) = \mathbf{w}_o^T\mathbf{u}(n) + v(n)$, where $\mathbf{w}_o$ is the unknown optimum coefficient vector (Wiener solution) and $v(n)$ has zero-mean, with variance σ_o^2, and is uncorrelated to $\mathbf{u}(n)$. The regressor sequence $\{\mathbf{u}(n)\}$ is assumed zero-mean, stationary with covariance matrix $\boldsymbol{R}$. The component filters compute estimates $\mathbf{w}_i(n)$, $i = 1,2$ and errors $e_i(n) = d(n)-y_i(n)$, $y_i(n) = \mathbf{w}_i^T(n)\mathbf{u}(n)$ through an adaptive algorithm such as LMS (the model below can be applied to many different algorithms). Define also the error vectors $\tilde{\mathbf{w}}_i(n) = \mathbf{w}_o - \mathbf{w}_i(n)$, and the *a-priori* errors $e_{a,i}(n) = \tilde{\mathbf{w}}_i^T(n)\mathbf{u}(n)$, $i = 1,2$.

Under these conditions,

$$y_1(n) - y_2(n) = \big[\mathbf{w}_1(n) - \mathbf{w}_2(n)\big]^T\mathbf{u}(n) = e_{a,2}(n) - e_{a,1}(n),$$

and (to simplify notation we omit the time index n in some variables in the remaining of the text)

$$e(n) = \eta e_1(n) + (1-\eta)e_2(n) = \eta e_{a,1} + (1-\eta)e_{a,2} + v(n). \quad (4)$$

The recursion for $a(n)$ then reads (with $\eta = \eta(a) = 1/(1+\mathrm{e}^{-a(n)})$)

$$\begin{aligned} a(n+1) = a(n) + \mu_a\big[(1-\eta)e_{a,2}^2 + (2\eta-1)e_{a,1}e_{a,2} - \\ - \eta e_{a,1}^2 + (e_{a,2}-e_{a,1})v(n)\big]\eta(1-\eta). \end{aligned} \quad (5)$$

We can rewrite this expression in a more convenient form defining

$$f_1(a) \triangleq -\eta^2(1-\eta), \qquad f_{12}(a) \triangleq \eta(2\eta-1)(1-\eta), \quad (6a)$$

$$f_2(a) \triangleq \eta(1-\eta)^2, \qquad f_v(a) \triangleq \eta(1-\eta), \quad \text{so that} \quad (6b)$$

$$\begin{aligned} a(n+1) = a(n) + \mu_a\big[f_1 e_{a,1}^2 + f_{12}e_{a,1}e_{a,2} + \\ + f_2 e_{a,2}^2 + (e_{a,2}-e_{a,1})f_v v(n)\big]. \end{aligned} \quad (7)$$

We will now find an approximate recursion for the expected value of $a(n)$. Since the distribution of $a(n)$ is unknown, we cannot compute exactly expected values involving the nonlinear functions (6). We therefore expand the nonlinear functions as Taylor series, around the expected value $\bar{a}(n) \triangleq \mathrm{E}\{a(n)\}$:

$$f_i(a) \approx f_i(\bar{a}) + \frac{\mathrm{d}\, f_i}{\mathrm{d}\, a}(\bar{a})\big(a-\bar{a}\big), \ i = 1,2, \quad (8)$$

and similarly for f_{12} and f_v. The derivatives can be evaluated as

$$g_1 \triangleq \frac{\mathrm{d}\, f_1}{\mathrm{d}\,\eta}\eta' = -\big[2\eta(1-\eta) - \eta^2\big]\eta', \quad (9a)$$

$$g_{12} \triangleq \frac{\mathrm{d}\, f_{12}}{\mathrm{d}\,\eta}\eta' = -(1-6\eta+6\eta^2)\eta', \quad (9b)$$

$$g_2 \triangleq \frac{\mathrm{d}\, f_2}{\mathrm{d}\,\eta}\eta' = (1-\eta)(1-3\eta)\eta', \quad (9c)$$

$$g_v \triangleq \frac{\mathrm{d}\, f_v}{\mathrm{d}\,\eta}\eta' = \big(1-2\eta\big)\eta', \quad (9d)$$

$$\eta' \triangleq \frac{\mathrm{d}\,\eta}{\mathrm{d}\, a} = \frac{\mathrm{e}^{-a}}{(1+\mathrm{e}^{-a})^2}. \quad (9e)$$

Denoting $\bar{f}_i(n) = f_i(\bar{a}(n))$, $\bar{g}_i(n) = g_i(\bar{a}(n))$, $i = 1,2$ (similarly for $\bar{f}_{12}$, $\bar{f}_v$, $\bar{g}_{12}$ and $\bar{g}_v$), the approximate recursion for $a(n)$ becomes

$$\begin{aligned} a(n+1) \approx a(n) + \mu_a\Big[\big(\bar{f}_1 + (a-\bar{a})\bar{g}_1\big)e_{a,1}^2 + \\ + \big(\bar{f}_{12} + (a-\bar{a})\bar{g}_{12}\big)e_{a,1}e_{a,2} + \big(\bar{f}_2 + (a-\bar{a})\bar{g}_2\big)e_{a,2}^2 + \\ + \big(\bar{f}_v + (a-\bar{a})\bar{g}_v\big)(e_{a,2}-e_{a,1})v(n)\Big]. \end{aligned} \quad (10)$$

We use this recursion in the remainder of this section to study the transient of the convex combination algorithm.

2.1. Convergence in the mean

We now take expected values on (10), using the following assumptions:

A-1. The noise $v(n)$ has zero mean with variance σ_0^2, and is independent, identically distributed (iid) and independent of $\{\mathbf{u}(n)\}$.

A-2. The auxiliary variable $a(n)$ varies slowly enough for the conditional expected value $\mathrm{E}\{e_{a,i}^k(n)e_{a,j}^\ell a(n)|a(n)\}$ to be approximately equal to $\mathrm{E}\{e_{a,i}^k(n)e_{a,j}^\ell(n)\}a(n)$, where $i,j = 1,2$ and $k,\ell = 0\dots4$, $k+\ell \le 4$.

The first assumption is common in adaptive filter analysis, and usually leads to good results [9]. The second follows from observations: simulations show that $a(n)$ converges slowly compared to variations in the input $\mathbf{u}(n)$ (and thus to variations on the *a-priori* errors) even for the large values of step-size μ_a usually employed.

Under these assumptions, we have $\mathrm{E}\{(a-\bar{a})\bar{g}_i e_{a,i}e_{a,j}\} \approx 0$, $i = 1,2$ (similarly for the terms involving $\bar{g}_{12}$ and $\bar{g}_v$}), so

$$\bar{a}(n+1) \approx \bar{a}(n) + \mu_a\big[\bar{f}_1\xi_1(n) + \bar{f}_{12}\xi_{12}(n) + \bar{f}_2\xi_2(n)\big], \quad (11)$$

where $\xi_1(n) = \mathrm{E}\{e_{a,1}^2(n)\}$, $\xi_{12}(n) = \mathrm{E}\{e_{a,1}(n)e_{a,2}(n)\}$, $\xi_2(n) = \mathrm{E}\{e_{a,2}^2(n)\}$. $\xi_1(n)$ and $\xi_2(n)$ are the mean-square errors of the component filters, which are readily available in the literature (see, e.g., [9,10]. The cross-term $\xi_{12}(n)$ is also available, see [1,5] (this last reference gives models for combinations of several different adaptive filtering algorithms). As in (3), we restrict $\bar{a}(n+1)$ to the interval $[-a_+, a_+]$.

Define $\bar{\eta} = \eta(\bar{a}(n))$, $\bar{\eta}' = \eta'(\bar{a}(n))$. Applying an approximation similar to the used in (7) to the overall *a-priori* error $e_a(n) = \eta e_{a,1}(n) + (1-\eta)e_{a,2}(n)$, we obtain

$$e_a(n) \approx \Big[\bar{\eta} + (a-\bar{a})\bar{\eta}'\Big]\big[e_{a,1}(n) - e_{a,2}(n)\big] + e_{a,2}(n), \quad (12)$$

so, using Assumptions A-1 and A-2, we have $\mathrm{E}\{e_a(n)\} \approx 0$.

2.2. Mean-square analysis

Using (3) and (12) we can obtain a model for the excess mean-square error (EMSE) of the combination. Squaring (12), taking the expected value and using Assumptions A-1 and A-2, we obtain

$$\begin{aligned}\mathrm{E}\{e_a^2(n)\} \approx \left[\bar{\eta}^2 + \sigma_a^2\bar{\eta}'^2\right]\left[\xi_1(n) - 2\xi_{12}(n) + \xi_2(n)\right] + \\ + 2\bar{\eta}\left[\xi_{12}(n) - \xi_2(n)\right] + \xi_2(n),\end{aligned} \tag{13}$$

where $\sigma_a^2(n) = \mathrm{E}\{\left[a(n) - \bar{a}(n)\right]^2\}$. Despite the approximations, this model gives accurate results, as will be shown in Sec. 3.

We now find an approximation for $\sigma_a^2(n)$, by squaring (10), taking the expected value, and subtracting the square of (11). During the computation of the square of (10), we find third- and fourth-order powers of $e_{a,1}^k(n)e_{a,2}^\ell(n)$, with $k+\ell = 3$ or 4. In order to evaluate their means, we need a third assumption, also common in the literature, and which gives good results mainly when the length of the component filters is large:

A-3. The *a-priori* errors $e_{a,1}(n)$ and $e_{a,2}(n)$ are jointly Gaussian, which implies [11]

$$\mathrm{E}\{e_{a,i}^4(n)\} = 3\xi_i^2(n), \quad i = 1,2 \tag{14}$$

$$\mathrm{E}\{e_{a,1}^k(n)e_{a,2}^\ell(n)\} = 0, \quad \text{if } k+\ell = 3, \tag{15}$$

$$\mathrm{E}\{e_{a,1}^k e_{a,2}^\ell\} = \begin{cases} 3\xi_1(n)\xi_{12}(n), & \text{if } k=3, \ell=1, \\ 3\xi_{12}(n)\xi_2(n), & \text{if } k=1, \ell=3, \\ 2\xi_{12}^2(n) + \xi_1(n)\xi_2(n), & \text{if } k=\ell=2. \end{cases} \tag{16}$$

Using Assumptions A-1–A-3, the recursion for $\sigma_a^2(n)$ becomes

$$\sigma_a^2(n+1) = \left(1 + 2\mu_a G_1 + \mu_a^2 G_2\right)\sigma_a^2(n) + \mu_a^2 F, \tag{17}$$

where

$$\begin{aligned} F &= 2\bar{f}_2^2\xi_2^2 + \bar{f}_{12}^2(\xi_{12}^2 + \xi_1\xi_2) + 2\bar{f}_1^2\xi_1^2 + 4\bar{f}_{12}\bar{f}_2\xi_{12}\xi_2 + \\ &\quad + 4\bar{f}_1\bar{f}_2\xi_{12}^2 + 4\bar{f}_1\bar{f}_{12}\xi_1\xi_{12} + \bar{f}_v^2(\xi_1 - 2\xi_{12} + \xi_2)\sigma_0^2, \\ G_1 &= \bar{g}_1\xi_1 + \bar{g}_{12}\xi_{12} + \bar{g}_2\xi_2, \\ G_2 &= 3\bar{g}_1^2\xi_1^2 + (\bar{g}_{12}^2 + 2\bar{g}_1\bar{g}_2)(2\xi_{12}^2 + \xi_1\xi_2) + 6\bar{g}_1\bar{g}_{12}\xi_1\xi_{12} + \\ &\quad + 6\bar{g}_{12}\bar{g}_2\xi_{12}\xi_2 + 3\bar{g}_2^2\xi_2^2 + \bar{g}_v^2(\xi_1 - 2\xi_{12} + \xi_2)\sigma_0^2. \end{aligned}$$

Note that, since $a(n) \in [-a_+, a_+]$, the maximum possible value of σ_a^2 is a_+^2, so we restrict $\sigma_a^2(n+1)$ in (17) to the interval $[0, a_+^2]$.

2.3. Normalized mixing parameter estimation

The models may be easily modified for the normalized mixing parameter estimation algorithm of [6]. In this case, (3) is modified to

$$a(n+1) = a(n) + \frac{\mu_a}{p(n)+\delta}\left(y_1(n) - y_2(n)\right)e(n)\eta(1-\eta), \tag{18}$$

where δ is a regularization parameter (a small positive number), and $p(n)$ is an estimate of the power of the regressor for the mixing parameter estimation, obtained using a first-order low-pass filter with $0 \ll \lambda_a < 1$:

$$\begin{aligned} p(n+1) &= \lambda_a p(n) + (1-\lambda_a)\left(y_1(n) - y_2(n)\right)^2 = \\ &= \lambda_a p(n) + (1-\lambda_a)\left(e_{a,2}(n) - e_{a,1}(n)\right)^2. \end{aligned} \tag{19}$$

In general, it is better to initialize the mixing parameter to a_+ (so the combination will start following the fast filter). During the first iterations, if the fast and slow filters are given the same initial condition, $p(n)$ would be small, which in turn would lead to a small equivalent step-size and a large variance for $a(n)$. Since this is not desirable, it is better to initialize $p(n)$ with a relatively large value.

The average $\bar{p}(n) \triangleq \mathrm{E}\{p(n)\}$ is given by

$$\bar{p}(n+1) = \lambda_a\bar{p}(n) + (1-\lambda_a)\left(\xi_1(n) - 2\xi_{12}(n) + \xi_2(n)\right),$$

using the initial condition $\bar{p}(0) = p(0)$. In practice one chooses λ_a close to 1 (e.g., 0.9). This implies that $p(n)$ varies slowly enough for the simple approximation below to give good results:

$$\frac{\mu_a}{p(n)+\delta} \approx \frac{\mu_a}{\bar{p}(n)+\delta} \triangleq \tilde{\mu}_a(n). \tag{20}$$

Our model for the normalized mixing parameter algorithm is therefore obtained simply by using $\tilde{\mu}_a(n)$ instead of μ_a in (11) and (17).

In the next section we compare the performance of our models with ensemble-average learning curves.

3. SIMULATIONS

In the simulations below we use two LMS filters to estimate a vector $\mathbf{w}_o$, generated randomly before each series of simulations (using a zero-mean Gaussian distribution with variance 1, without correlation between the coefficients). Vector $\mathbf{w}_o$ is not changed during the computation of each ensemble-average learning curve. The noise $v(n)$ is white Gaussian, with variance 10^{-3}, and the regressor vector $\mathbf{u}(n)$ is generated from a stationary sequence $\{u(n)\}$ passing through a tap-delay line, where

$$u(n+1) = \lambda_u u(n) + \sqrt{1-\lambda_u^2}\,\epsilon(n),$$

where $\epsilon(n)$ is a white Gaussian sequence with unit variance (as is the variance of $u(n)$). In the EMSE plots below, we filter the EMSE curves with a 5-tap moving-average filter to further smooth the graphs, using Matlab's filtfilt function to eliminate delay.

We first test the unnormalized recursion (3). The LMS filters estimate $M = 7$ coefficients, using $\mu_1 = 0.08$ (fast filter) and $\mu_2 = 0.008$ (slow filter), with white regressors ($\lambda_u = 0$) with $\sigma_u^2 = 1$. The mixing parameter is computed using $\mu_a = 1,000$ (chosen so that the combination will track correctly the best filter at all times). This is a challenging situation, since the combination algorithm needs a very large μ_a in order to switch to the slow filter at the correct point. This large μ_a in turn makes the variance of $a(n)$ very large at the beginning of the simulation (since ξ_1 and ξ_2 are both quite large there). Fig. 3 shows the EMSE for the component and combination and the theoretical models obtained taking or not into account σ_a^2 in (13).

We should remark that the estimation of rapidly growing variables is a difficult problem: a small error in the rate of growth will quickly lead to a large estimation error. However, although the model for the variance of σ_a^2 significantly differs from the observed in the simulations for small n, the model for the variance of η (and for the combination EMSE) is still quite good, in part because the derivative of η is small. The variance σ_a^2 is large also during the transition from the fast to the slow filter (the derivative of η with respect to a is at its largest value here). During the transition, again the predicted variance overestimates the simulation, but now (given the larger value of the derivative of η) this also affects the estimate of σ_η^2, and causes an overshoot in the estimate of the combination EMSE. However, we note that the EMSE model using $\sigma_a^2 = 0$ in (13) remains close to the ensemble-average curve. In all, we see that the model with $\sigma_a^2 = 0$ predicts well the combination EMSE even for such a large step-size, confirming that the variance of η does not have a large effect in the overall error variance.

In the second example we use the normalized mixing parameter estimation (20) with $\mu_a = 1$, $M = 25$ coefficients, and $\lambda_x = 0.8$. Fig. 3 shows that the normalized step-size chooses a small equivalent step-size $\tilde{\mu}_a$ at the beginning of the adaptation, thus stabilizing the adaptation of $a(n)$, while reverting to a large step-size at the end, in order for the algorithm to be able to switch from the fast to the slow filter at the correct moment. The agreement between theory and model is quite good, the only exception being an overestimation of the variance of $a(n)$ during the switch from one component filter to the other (but this effect is not as pronounced as in the unnormalized case.)

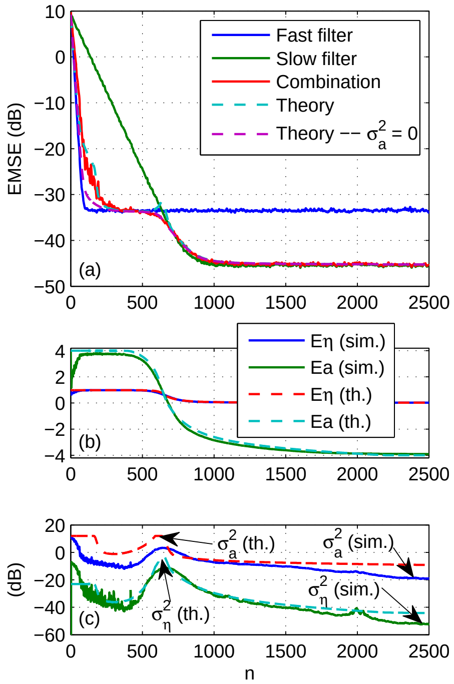

Fig. 2. (a) Theoretical and experimental EMSE for the combination of two LMS filters with $\lambda_x = 0$, $\mu_1 = 0.08$, $\mu_2 = 0.008$, and $\mu_a = 1{,}000$; (b) Ensemble-average of $a(n)$ and $\eta(n)$, and theoretical models $\bar{a}$ and $\bar{\eta}$; (c) Ensemble-average and model for $\sigma_a^2(n)$ and $\sigma_\eta^2(n)$. Average of 500 realizations.

4. CONCLUSIONS

We proposed a model for the transient of convex combinations of adaptive filters, illustrating our results with combinations of two LMS filters. Although the problem is highly nonlinear and quite challenging, we obtained quite good agreement between model and theory, except for an overestimation of the variance of the auxiliary variable $a(n)$ in some instants.

We should note that we chose to present simulations for some of the most difficult situations, in particular using a low noise variance, so the combination would need a large μ_a to switch from the fast to the slow filter at the right moment. In milder situations for which smaller step-sizes may be used, the agreement between theory and simulations is considerably better. However, even in the difficult situations presented here, the models follow quite closely the simulations.

5. REFERENCES

[1] 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, no. 3, pp. 1078–1090, Mar. 2006.

[2] J. Arenas-Garcia and A. R. Figueiras-Vidal, "Improved blind equalization via adaptive combination of constant modulus algorithms," in *Proc., ICASSP 2006*, Toulouse, France, pp. III–756–III–759.

[3] Y. Zhang and J. Chambers, "Convex combination of adaptive filters for a variable tap-length LMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 628–631, Oct. 2006.

[4] C. Lopes, E. Satorius, and A. H. Sayed, "Adaptive carrier tracking for direct-to-earth mars communications," in *Conf. Rec. 40th Asilomar Conf. on Sign., Syst. and Comp.*, Pacific Grove, CA, USA, Oct. 2006.

[5] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3137–3149, Jul. 2008.

[6] L. Azpicueta-Ruiz, A. Figueiras-Vidal, and J. Arenas-Garcia, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc., ICASSP 2008*, Las Vegas, NV, USA, pp. 3301–3304.

[7] 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, no. 5, pp. 1853–1864, May 2008.

[8] R. Candido, M. T. M. Silva, and V. H. Nascimento, "Affine combinations of adaptive filters," in *Conf. Rec. of the 42nd Asilomar Conf. on Sign., Syst. & Comp.*, 2008.

[9] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley, 2003.

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

[11] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. McGraw-Hill, 1984.

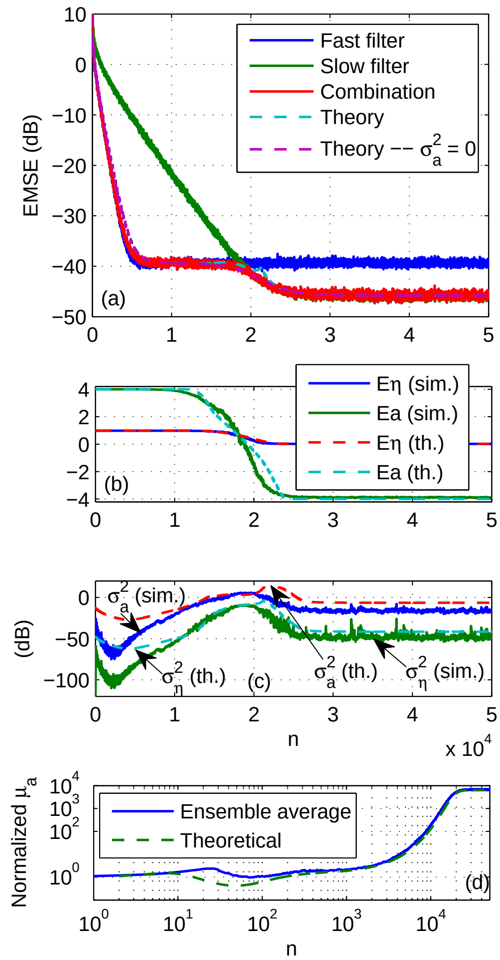

Fig. 3. (a) Theoretical and experimental EMSE for the combination of two LMS filters with $\lambda_u = 0.8$, $\mu_1 = 0.008$, $\mu_2 = 0.002$, and $\tilde{\mu}_a = 1{,}000$; (b) Ensemble-average of $a(n)$ and $\eta(n)$, and theoretical models $\bar{a}$ and $\bar{\eta}$; (c) Ensemble-average and model for $\sigma_a^2(n)$ and $\sigma_\eta^2(n)$; (d) Ensemble-average and model for $\tilde{\mu}_a$. Average of 50 realizations.