A TRANSIENT ANALYSIS FOR THE CONVEX COMBINATION OF TWO ADAPTIVE FILTERS WITH TRANSFER OF COEFFICIENTS

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

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

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

ABSTRACT

This paper proposes an improved model for the transient of convex combinations of adaptive filters. A previous model, based on a first-order Taylor series approximation of the nonlinear functions that appear in convex combinations, tended to overestimate the variance of the auxiliary variable used to estimate the mixing parameter. In this paper, we apply a second-order Taylor approximation that improves these estimates, and obtains better agreement with simulations. In addition, we also extend the model to include a simple mechanism for the transfer of coefficients between the constituent filters, a procedure that greatly improves the convergence of the overall filter, and provide an expression to select the free parameter used in such a scheme.

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

1. INTRODUCTION

Over the last years, combinations of adaptive filters have been a topic of intense research in the signal processing community (see, e.g., [1–9]). The first combined scheme that attracted attention was the convex combination of adaptive filters 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 [3]. In such scheme, depicted in Fig. 1, the output of the overall filter is given by $y(n) = \lambda(n)y_1(n)+[1-\lambda(n)]y_2(n)$, where $\lambda(n) \in [0, 1]$ is the mixing parameter, $y_i(n)$, $i = 1, 2$, are the outputs of the transversal filters, i.e., $y_i(n) = \mathbf{u}^T(n)\mathbf{w}_i(n-1)$, $\mathbf{u}(n)$ and $\mathbf{w}_i(n-1) \in \mathbb{R}^M$ being, respectively, the common regressor vector, and the weight vectors of each component filter.

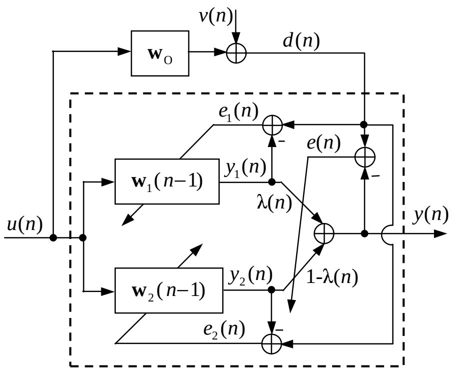

Fig. 1. Convex combination of two transversal adaptive filters.

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

In order to restrict the mixing parameter to the interval $[0, 1]$ and to reduce gradient noise when $\lambda \approx 0$ or $\lambda \approx 1$, a nonlinear transformation and an auxiliary variable $a(n)$ are used, i.e., $\lambda(n) = 1/[1+e^{-a(n-1)}]$, where $a(n)$ is updated to minimize the square of the overall error $e(n) = d(n) - y(n)$ [3]:

$$a(n) = a(n-1)+\mu_a\big[y_1(n)-y_2(n)\big]e(n)\lambda(n)\big[1-\lambda(n)\big]. \quad (1)$$

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

We assume that the two component filters are adapted to minimize their own quadratic errors using the least mean square (LMS) algorithm with different step sizes ($\mu_1 > \mu_2$), i.e.,

$$\mathbf{w}_i(n) = \mathbf{w}_i(n-1) + \mu_i e_i(n)\mathbf{u}(n), \quad i = 1, 2. \quad (2)$$

where $e_i(n) = d(n) - y_i(n)$.

The performance of the convex combination of two LMS algorithms can be improved if interaction between the component filters is allowed, as proposed in [2]. The convergence of the slow filter can be accelerated when an abrupt change appears by transferring a part of the filter $\mathbf{w}_1$ to $\mathbf{w}_2$. Thus, the modified adaption rule for $\mathbf{w}_2$ becomes

$$\mathbf{w}_2(n) = \alpha\left[\mathbf{w}_2(n-1) + \mu_2 e_2(n)\mathbf{u}(n)\right] + (1-\alpha)\mathbf{w}_1(n-1), \quad (3)$$

where α is a parameter close to 1. As observed in [2], an inconvenience of the new learning rule is that it increases the final misadjustment of the slow filter. To avoid this, the coefficient transfer must only be applied when $\lambda(n) \geq \beta$, where β is a threshold close to the maximum value that can be reached by $\lambda(n)$. Common choices in the literature are $a_+ = 4$ and $\beta = 0.98$.

By relying on a Taylor series approximation of the nonlinearities, a theoretical model for the transient behavior of convex combinations was proposed recently in [9]. Although the problem is highly nonlinear and quite challenging, good agreement between model and theory was obtained, except for an overestimation of the variance of the auxiliary variable $a(n)$ in some instants, which in turn led to an overestimation of the overall excess mean-square error (EMSE) during the switching from the fast to the slow filter.

In this paper, we extend the transient analysis of [9], improving the model for the variance of the auxiliary variable $a(n)$ using second-order Taylor approximation and including the coefficient transfer procedure. We also obtain a theoretical expression for selecting parameter α of the weight transfer procedure. The theoretical models allow a better understanding of the influence of design parameters on performance, providing the designer with tools to correctly apply the algorithms. Although our analysis is valid for combinations of different kinds of adaptive filters, we particularize the results for the combination of two LMS filters with different step sizes. Furthermore, we assume that the auxiliary variable $a(n)$ is updated via (1). However, using the results of [9], our analysis can be easily extended to the normalized mixing parameter estimation algorithm of [5]. In order to simplify the arguments, we assume that all quantities are real.

2. TRANSIENT ANALYSIS

In the analysis, a linear regression model for the desired signal is assumed, i.e., $d(n) = \mathbf{u}^T(n)\mathbf{w}_o + v(n)$, $\mathbf{w}_o$ being the unknown optimum coefficient vector (Wiener solution) and $v(n)$ an i.i.d. (independent and identically distributed) and zero-mean random process with variance σ_v^2. In order to make the analysis more tractable, the sequences $\{u(n)\}$ and $\{v(n)\}$ are assumed stationary and we will use the common assumption that $v(n)$ is independent of $\mathbf{u}(n)$ (not only uncorrelated) [10, Sec. 6.2.1]. Defining the weight-error vectors $\widetilde{\mathbf{w}}_i(n) = \mathbf{w}_o - \mathbf{w}_i(n)$ and the *a priori* errors $e_{a,i}(n) = \mathbf{u}^T(n)\widetilde{\mathbf{w}}_i(n-1)$, we find that

$$e_i(n) = e_{a,i}(n) + v(n), \tag{4}$$

and similarly for the overall error, i.e., $e(n) = e_a(n) + v(n)$. An important consequence of this model is that $v(k)$ will be independent of all $\mathbf{w}_i(j)$, $\widetilde{\mathbf{w}}_i(j)$, and $e_{a,i}(k)$, $i = 1, 2$, $j < k$, for any particular time instant k [10, Lemma 6.2.1].

It is common in the literature to evaluate the EMSE as $\zeta_{ij}(n) \triangleq \mathrm{E}\{e_{a,i}(n)e_{a,j}(n)\} \approx \mathrm{Tr}(\mathbf{R}\mathbf{S}_{ij}(n-1))$, where $\mathrm{E}\{\cdot\}$ represents expectation, $\mathbf{R} \triangleq \mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}$, and

$$\mathbf{S}_{ij}(n) \triangleq \mathrm{E}\{\widetilde{\mathbf{w}}_i(n)\widetilde{\mathbf{w}}_j^T(n)\}, \quad i,j = 1,2 \tag{5}$$

is the covariance ($i = j$) or the cross-variance ($i \neq j$) matrix of the weight-error vectors. This approach is based on the independence assumption between the regressor vector $\mathbf{u}(n)$ and weight-error vectors $\widetilde{\mathbf{w}}_i(n-1)$, which is a widely accepted assumption [10, 11].

In the sequel, we obtain recursions for $\mathbf{S}_{12}(n)$ and $\mathbf{S}_{22}(n)$, assuming the coefficient transfer procedure. We should notice that the case without coefficient transfer can be obtained making $\alpha \leftarrow 1$. Subtracting both sides of (2) (with $i=1$) and (3) from $\mathbf{w}_o$, we obtain

$$\widetilde{\mathbf{w}}_1(n) = \widetilde{\mathbf{w}}_1(n-1) - \mu_1 e_1(n)\mathbf{u}(n) \quad \text{and} \tag{6}$$

$$\widetilde{\mathbf{w}}_2(n) = \alpha\left[\widetilde{\mathbf{w}}_2(n-1) - \mu_2 e_2(n)\mathbf{u}(n)\right] + (1-\alpha)\widetilde{\mathbf{w}}_1(n-1). \tag{7}$$

To obtain a recursion for $\mathbf{S}_{12}(n)$, we multiply (6) by the transpose of (7) and take expectations on both sides. Using (4), after some algebra, we get

$$\begin{aligned}
\mathrm{E}\{\widetilde{\mathbf{w}}_1(n)\widetilde{\mathbf{w}}_2^T(n)\} &\approx \alpha\mathrm{E}\{\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\} \\
&-\alpha\mu_2\mathrm{E}\{\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\} \\
&+(1-\alpha)\mathrm{E}\{\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_1^T(n-1)\} \\
&-\alpha\mu_1\mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\} \\
&+\alpha\mu_1\mu_2\mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_2^T(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\} \\
&+\alpha\mu_1\mu_2\sigma_v^2\mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\} \\
&-(1-\alpha)\mu_1\mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\widetilde{\mathbf{w}}_1(n-1)\widetilde{\mathbf{w}}_1^T(n-1)\}.
\end{aligned} \tag{8}$$

Assuming independence between $\mathbf{u}(n)$ and $\widetilde{\mathbf{w}}_i(n-1)$ and Gaussian input regressors, the following approximation holds [10–12]

$$\mathrm{E}\left\{\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{S}_{12}\mathbf{u}(n)\mathbf{u}^T(n)\right\} \approx 2\mathbf{R}\mathbf{S}_{12}\mathbf{R} + \mathrm{Tr}\left(\mathbf{R}\mathbf{S}_{12}\right)\mathbf{R}.$$

Using these assumptions, (8) can be simplified to

$$\begin{aligned}
\mathbf{S}_{12}(n) &\approx \alpha\mathbf{S}_{12}(n-1) - \alpha\mu_1\mathbf{R}\mathbf{S}_{12}(n-1) - \alpha\mu_2\mathbf{S}_{12}(n-1)\mathbf{R} \\
&+ \alpha\mu_1\mu_2\left[2\mathbf{R}\mathbf{S}_{12}(n-1)\mathbf{R} + \mathrm{Tr}(\mathbf{R}\mathbf{S}_{12}(n-1))\mathbf{R} + \sigma_v^2\mathbf{R}\right] \\
&- (1-\alpha)\mu_1\mathbf{R}\mathbf{S}_{11}(n-1) + (1-\alpha)\mathbf{S}_1(n-1).
\end{aligned} \tag{9}$$

Analogously, multiplying $\mathbf{w}_2(n)$ by its transpose in (3) and using the same previous assumptions, we obtain

$$\begin{aligned}
\mathbf{S}_{22}(n) &= \alpha^2\mathbf{S}_{22}(n-1) - \alpha^2\mu_2\left[\mathbf{S}_{22}(n-1)\mathbf{R} + \mathbf{R}\mathbf{S}_{22}(n-1)\right] \\
&+ \alpha^2\mu_2^2\left[2\mathbf{R}\mathbf{S}_{22}(n-1)\mathbf{R} + \mathrm{Tr}(\mathbf{R}\mathbf{S}_{22}(n-1))\mathbf{R} + \sigma_v^2\mathbf{R}\right] \\
&- \alpha(1-\alpha)\mu_2\left[\mathbf{R}\mathbf{S}_{12}(n-1) + \mathbf{S}_{12}(n-1)\mathbf{R}\right] \\
&+ 2\alpha(1-\alpha)\mathbf{S}_{12}(n-1).
\end{aligned} \tag{10}$$

We can now study the convergence rate of $\mathbf{w}_2$ as a function of α. For this purpose, we first determine a recursion for the diagonal of $\mathbf{Q}^T\mathbf{S}_{22}(n)\mathbf{Q}$, where $\mathbf{Q}$ is an orthogonal transformation that diagonalizes $\mathbf{R}$, i.e., $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, $\mathbf{Q}^T\mathbf{R}\mathbf{Q} = \mathbf{\Lambda}$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_M)$ and $\lambda_1, \lambda_2 \cdots \lambda_M$ are the eigenvalues of $\mathbf{R}$. Defining $\mathbf{s}_{ij}(n) = \mathrm{diag}(\mathbf{Q}^T\mathbf{S}_{ij}(n)\mathbf{Q})$, $i, j = 1, 2$ (the diagonal elements of $\mathbf{Q}^T\mathbf{S}_{ij}(n)\mathbf{Q}$) and $\boldsymbol{\ell} = \left[\begin{smallmatrix}\lambda_1 & \cdots & \lambda_M\end{smallmatrix}\right]^T$, (10) reduces to

$$\mathbf{s}_{22}(n) = \alpha^2\mathbf{A}_2\mathbf{s}_{22}(n-1) - 2\alpha(1-\alpha)\mu_2\mathbf{\Lambda}\mathbf{s}_{12}(n-1) + \alpha^2\mu_2^2\sigma_v^2\boldsymbol{\ell}, \tag{11}$$

where $\mathbf{A}_2 = \left[\mathbf{I} - 2\mu_2\mathbf{\Lambda} + \mu_2^2\left(2\mathbf{\Lambda}^2 + \boldsymbol{\ell}\boldsymbol{\ell}^T\right)\right]$. Making $\alpha \to 1$ and changing the index 2 by 1 in (11), we can obtain a similar expression for $\mathbf{s}_{11}(n)$, i.e.,

$$\mathbf{s}_{11}(n) = \mathbf{A}_1\mathbf{s}_{11}(n-1) + \mu_1^2\sigma_v^2\boldsymbol{\ell}. \tag{12}$$

When the transfer of coefficients occurs, $\mathbf{w}_2$ should follow $\mathbf{w}_1$ closely. The slowest mode in (12) has a convergence rate of $\lambda_{\max}(\mathbf{A}_1)$, where $\lambda_{\max}(\mathbf{A})$ is the largest eigenvalue of matrix $\mathbf{A}$. We choose α so that the slowest mode in (11) converges at least as fast as that, i.e.,

$$\alpha \leq \sqrt{\lambda_{\max}(\mathbf{A}_1)/\lambda_{\max}(\mathbf{A}_2)}. \tag{13}$$

For sufficiently small step-sizes μ_1 and μ_2, (13) reduces to

$$\alpha \leq \sqrt{\frac{1 - 2\mu_1\lambda_{\min}(\mathbf{R})}{1 - 2\mu_2\lambda_{\min}(\mathbf{R})}}. \tag{14}$$

We now derive a model for the mixing parameter $\lambda(n)$. Noting that $y_1(n) - y_2(n) = e_{a,2}(n) - e_{a,1}(n)$, we can rewrite

$$e(n) = \lambda(n)e_{a,1}(n) + [1 - \lambda(n)]e_{a,2}(n) + v(n). \tag{15}$$

To simplify notation, the time index is omitted in some variables. The recursion for $a(n)$ then reads (with $\lambda = \lambda(a(n-1)) = 1/\left[1+\mathrm{e}^{-a(n-1)}\right]$)

$$\begin{aligned}
a(n) = a(n-1) + \mu_a\big[&-\lambda e_{a,1}^2 + (1-\lambda)e_{a,2}^2 \\
&+ (2\lambda - 1)e_{a,1}e_{a,2} + (e_{a,2} - e_{a,1})v\big]\lambda(1-\lambda).
\end{aligned} \tag{16}$$

We can rewrite this expression in a more convenient form defining

$$\varepsilon_1 \triangleq e_{a,1}^2, \qquad \varepsilon_2 \triangleq e_{a,2}^2, \tag{17a}$$

$$\varepsilon_3 \triangleq e_{a,1}e_{a,2}, \qquad \varepsilon_4 \triangleq (e_{a,2} - e_{a,1})v(n), \tag{17b}$$

$$f_1(a) \triangleq -\lambda^2(1-\lambda), \qquad f_2(a) \triangleq \lambda(1-\lambda)^2, \tag{17c}$$

$$f_3(a) \triangleq \lambda(2\lambda - 1)(1-\lambda), \qquad f_4(a) \triangleq \lambda(1-\lambda), \tag{17d}$$

so that

$$a(n) = a(n-1) + \mu_a\sum_{k=1}^{4} f_k\varepsilon_k. \tag{18}$$

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 exact expected values involving the nonlinear functions (17c) and (17d). We therefore expand these nonlinear functions as a Taylor series, around the expected value $\bar{a}(n-1) \triangleq \mathrm{E}\{a(n-1)\}$, i.e.,

$$f_k(a) \approx f_k(\bar{a}) + \frac{\mathrm{d}f_k}{\mathrm{d}a}(\bar{a})\left(a - \bar{a}\right), \quad k = 1, \cdots, 4 \tag{19}$$

Although these first-order approximations suffice for modeling the transient of $\bar{a}(n)$, for the mean-square analysis second-order expansions produce more accurate results. Therefore, Table 1 includes both first- and second-order derivatives of f_k (denoted as g_k and h_k, respectively), for $k = 1, \cdots, 4$. We noticed in simulations that this mixed model is more accurate than a complete second-order model.

Denoting $\bar{f}_k = f_k[\bar{a}(n-1)]$ and $\bar{g}_k = g_k[\bar{a}(n-1)]$, the approximate recursion for $a(n)$ becomes

$$a(n) \approx a(n-1) + \mu_a\sum_{k=1}^{4}\left(\bar{f}_k + (a - \bar{a})\bar{g}_k\right)\varepsilon_k. \tag{20}$$

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

Table 1. First and second order derivatives of f_k, $k = 1, \cdots, 4$.

k	$g_k \triangleq \mathrm{d}f_k/\mathrm{d}a$	$h_k \triangleq \mathrm{d}^2 f_k/\mathrm{d}a^2$
1	$-3\lambda^4+5\lambda^3-2\lambda^2$	$12\lambda^5-27\lambda^4+19\lambda^3-4\lambda^2$
2	$-3\lambda^4+7\lambda^3-5\lambda^2+\lambda$	$12\lambda^5-33\lambda^4+31\lambda^3-11\lambda^2+\lambda$
3	$6\lambda^4-12\lambda^3+7\lambda^2-\lambda$	$-24\lambda^5+60\lambda^4-50\lambda^3+15\lambda^2-\lambda$
4	$2\lambda^3-3\lambda^2+\lambda$	$-6\lambda^4+12\lambda^3-7\lambda^2+\lambda$

2.1. Convergence in the mean

We now take expected values on (20), using (4) and assuming that

A1. The auxiliary variable $a(n)$ varies slowly enough for the conditional expected value $\mathrm{E}\{e^{\ell}_{a,i}(n)e^{m}_{a,j}a(n)|a(n)\}$ to be approximately equal to $\mathrm{E}\{e^{\ell}_{a,i}(n)e^{m}_{a,j}(n)\}a(n)$, where $i, j = 1, 2$ and $\ell, m = 0 \dots 4$, $m + \ell \leq 4$.

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. A consequence of A1 is that $\mathrm{E}\{(a-\bar{a})\bar{g}_k\varepsilon_k\} \approx 0$, $k = 1, \cdots, 4$, so

$$\bar{a}(n) \approx \bar{a}(n-1) + \mu_a\left[\bar{f}_1\zeta_{11}(n) + \bar{f}_2\zeta_{22}(n) + \bar{f}_3\zeta_{12}(n)\right]. \quad (21)$$

As in (1), we restrict $\bar{a}(n+1)$ to the interval $[-a_+, a_+]$.

To complete the first-order analysis, we should notice that the mean of the overall *a priori* error is zero. Defining $\bar{\lambda} = \lambda(\bar{a}(n-1))$, $\bar{\lambda}' = \frac{\mathrm{d}\lambda}{\mathrm{d}a}[\bar{a}(n-1)]$, and applying an approximation similar to that used in (20) to the overall *a priori* error $e_a(n) = \lambda[e_{a,1}(n) - e_{a,2}(n)] - e_{a,2}(n)$, we obtain

$$e_a(n) \approx \left[\bar{\lambda} + (a-\bar{a})\bar{\lambda}'\right]\left[e_{a,1}(n) - e_{a,2}(n)\right] + e_{a,2}(n), \quad (22)$$

so, using model (4) and Assumption A1, we have $\mathrm{E}\{e_a(n)\} \approx 0$.

2.2. Mean-square analysis

Using (1) and (22) we can obtain a model for the EMSE of the combination. Squaring (22), taking the expected value, and using model (4) and Assumption A1, we obtain

$$\mathrm{E}\{e_a^2(n)\} \approx \left[\bar{\lambda}^2+\sigma_a^2(n-1)\bar{\lambda}'^2\right]\left[\zeta_{11}(n)-2\zeta_{12}(n)+\zeta_{22}(n)\right] + 2\bar{\lambda}\left[\zeta_{12}(n) - \zeta_{22}(n)\right] + \zeta_{22}(n). \quad (23)$$

We now find a recursion for $\sigma_a^2(n) = \mathrm{E}\{a^2(n)\} - \bar{a}^2(n)$. Squaring (18), we arrive at

$$a^2(n) = a^2(n-1) + \mu_a^2\sum_{k=1}^{4}\sum_{\ell=1}^{4} f_k f_\ell \varepsilon_k\varepsilon_\ell + 2\mu_a a(n-1)\sum_{k=1}^{4} f_k\varepsilon_k. \quad (24)$$

Using first-order Taylor series as in (19), the function $f_k f_\ell$ can be approximated around $\bar{a}(n-1)$ by

$$f_k f_\ell \approx \bar{f}_k\bar{f}_\ell + \left(\bar{g}_k\bar{f}_\ell + \bar{f}_k\bar{g}_\ell\right)(a-\bar{a}). \quad (25)$$

To obtain a more accurate model for the variance of $a(n)$ than that of [9], we approximate the function $a(n-1)f_k$ by a second-order Taylor series around $\bar{a}(n-1)$, i.e.,

$$af_k \approx \bar{a}\bar{f}_k + (\bar{f}_k + \bar{a}\bar{g}_k)(a-\bar{a}) + 0.5(2\bar{g}_k + \bar{a}\bar{h}_k)(a-\bar{a})^2, \quad (26)$$

where $\bar{h}_k$ are the second-order derivatives of f_k shown in Table 1 for $k = 1, \cdots, 4$, and calculated at $\bar{\lambda} = \lambda(\bar{a}(n-1))$.

Thus, using (25), (26), and A1, taking expectations on both sides of (24), noting that $\mathrm{E}\{a-\bar{a}\} = 0$ and $\mathrm{E}\{(a-\bar{a})^2\} = \sigma_a^2$, we obtain

$$\mathrm{E}\{a^2(n)\} \approx \mathrm{E}\{a^2(n-1)\} + \mu_a^2\sum_{k=1}^{4}\sum_{\ell=1}^{4}\bar{f}_k\bar{f}_\ell\mathrm{E}\{\varepsilon_k\varepsilon_\ell\} + 2\mu_a\sum_{k=1}^{4}\left[\bar{a}(n-1)\bar{f}_k + \left(\bar{g}_k + 0.5\bar{a}\bar{h}_k\right)\sigma_a^2(n-1)\right]\mathrm{E}\{\varepsilon_k\}. \quad (27)$$

Now, note that $\mathrm{E}\{\varepsilon_4\} = \mathrm{E}\{\varepsilon_4\varepsilon_k\} = 0$ for $k = 1, 2, 3$. To complete the computation of (27), third- and fourth-order powers of $e^k_{a,1}(n)e^\ell_{a,2}(n)$, with $k + \ell = 3$ or 4 [represented by $\mathrm{E}\{\varepsilon_k\varepsilon_\ell\}$ in (27)] should be evaluated. For this purpose, we need another assumption, also common in the literature, and which gives good results mainly for long adaptive filters:

A2. The *a-priori* errors $e_{a,1}(n)$ and $e_{a,2}(n)$ are jointly Gaussian, which implies [12]

$$\mathrm{E}\{e^4_{a,i}(n)\} = 3\zeta^2_{ii}(n), \quad i = 1, 2 \quad (28)$$

$$\mathrm{E}\{e^k_{a,1}(n)e^\ell_{a,2}(n)\} = 0, \quad \text{if } k + \ell = 3, \quad (29)$$

$$\mathrm{E}\{e^k_{a,1}e^\ell_{a,2}\} = \begin{cases} 3\zeta_{11}(n)\zeta_{12}(n), & \text{if } k=3, \ell=1, \\ 3\zeta_{12}(n)\zeta_{22}(n), & \text{if } k=1, \ell=3, \\ 2\zeta^2_{12}(n)+\zeta_{11}(n)\zeta_{22}(n), & \text{if } k=\ell=2. \end{cases} \quad (30)$$

Finally, subtracting the square of (21) from (27) and using A2, the recursion for $\sigma_a^2(n)$ becomes

$$\begin{aligned}\sigma_a^2(n) = \sigma_a^2(n-1)&\Bigg[1+\sum_{k=1}^{2}\Big(\bar{g}_k+0.5\bar{a}(n-1)\bar{h}_k\Big)\zeta_{kk} \\ &+\Big(\bar{g}_3+0.5\bar{a}(n-1)\bar{h}_3\Big)\zeta_{12}\Bigg] + \mu_a^2\big[2\left(\bar{f}_1^2\zeta_{11}^2+\bar{f}_2^2\zeta_{22}^2\right) \\ &+\bar{f}_3^2\left(\zeta_{12}^2+\zeta_{11}\zeta_{22}\right)+\bar{f}_4^2\sigma_v^2(\zeta_{11}-2\zeta_{12}+\zeta_{22}) \\ &+4\big(\bar{f}_1\bar{f}_2\zeta_{12}^2+\bar{f}_1\bar{f}_3\zeta_{11}\zeta_{12}+f_2f_3\zeta_{22}\zeta_{12}\big)\big]. \end{aligned} \quad (31)$$

Since $a(n) \in [-a_+, a_+]$ and $\sigma_a^2(n) = E\{a^2(n)\} - \bar{a}^2(n)$, we truncate at each iteration the variance $\sigma_a^2(n)$ to the interval $[0, a_+^2 - \bar{a}^2(n)]$.

When compared to the recursion for $\sigma_a^2(n)$ derived in [9, Eq.(17)], (31) gives more accurate results, as will be shown in Sec. 3.

2.3. Transient model considering the transfer of coefficients

Expanding the mixing parameter $\lambda(n)$ as a first-order Taylor series around $\bar{a}(n-1)$ yields the approximation $\sigma_\lambda^2(n) \approx \left[\bar{\lambda}'(n)\right]^2\sigma_a^2(n-1)$ for its variance. When the transfer of coefficients occurs, $\lambda(n) \geq \beta \approx 1$ and its derivative $\bar{\lambda}'(n)$ is small. Hence, the variance $\sigma_\lambda^2(n)$ will be also small. Thus, the transient model for the coefficient transfer procedure is easily obtained if instead of $\lambda(n) \geq \beta$, we use the approximate condition $\bar{\lambda}(n) \geq \beta$ to switch from the model with $\alpha = 1$ to the model with $\alpha \neq 1$.

3. SIMULATION RESULTS

In this section, we carry out simulation work to validate the accurateness of the new second-order model for the transient of the convex combination, both for the cases with and without transfer of coefficients. A vector $\mathbf{w}_\mathrm{o}$ of length $M = 7$ is generated randomly before each series of experiments, and normalized to have unitary norm. The noise $v(n)$ is i.i.d. Gaussian with variance $\sigma_v^2 = 10^{-2}$, and regressors $\mathbf{u}(n)$ are 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 noise with variance 1. All simulated curves have been obtained by averaging 1000 independent runs.

Fig. 2 compares the transient analysis from [9] and the extended second-order model proposed in this paper when no weight transfer is applied. We illustrate both a case with white input (with settings $\lambda_u = 0$, $\mu_1 = 0.05$, $\mu_2 = 0.005$ and $\mu_a = 100$), and another one with colored regressors ($\lambda_u = 0.7$, $\mu_1 = 0.1$, $\mu_2 = 0.01$ and $\mu_a = 100$). As we can see, the new model provides a more accurate approximation

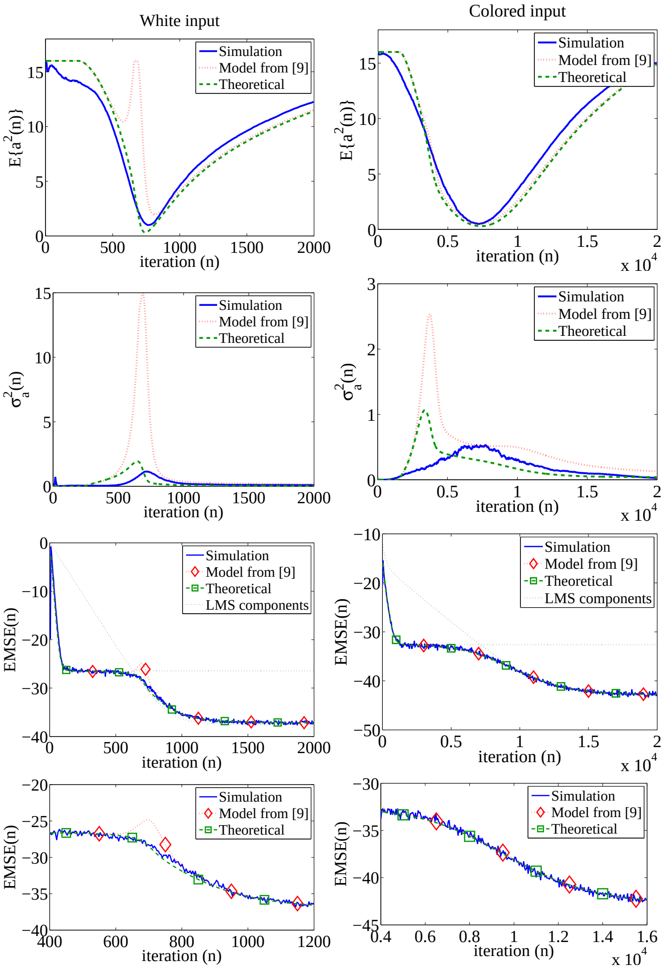

Fig. 2. Comparison of the transient model from [9] and the extended model proposed in this paper. The two first rows illustrate the estimations of $\mathrm{E}\{a^2(n)\}$, $\sigma_a^2(n)$, while the two bottom rows depict EMSE evolution and a detail of the switching between components.

of $\mathrm{E}\{a^2(n)\}$, and tends to correct the overstimation of the variance observed for the model from [9], both for white and colored inputs. This more accurate estimation of the variance results also in a better prediction of the EMSE of the combination, what is specially clear for white input, where the old model predicted a bump during the transient which does not appear in the simulations. We should also mention that qualitatively similar results are also observed for other values of μ_a.

Fig. 3 represents the performance of the combination when the transfer of coefficients mechanism is applied. We consider the previous scenario with white input, and use different values of parameter α. The upper pannel shows a good agreement between the theoretical and simulated curves, and illustrate the faster convergence provided by the weight transfer mechanism. The bottom pannel displays a detail of the EMSE evolution for the overall filter during the switching between components, for different values of α. In the light of the results, we can conclude that (14) allows an appropriate selection of the constant, in terms of the speed of convergence of the overall filter.

4. CONCLUSIONS

This work presented a new analytical model for the transient of the convex combination of two adaptive filters. Including the coefficient

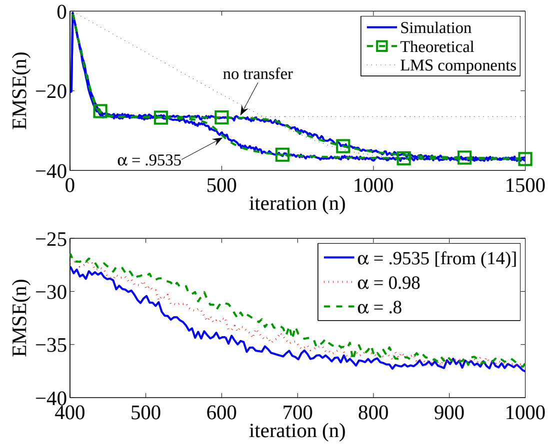

Fig. 3. Theoretical and simulated EMSEs for the combination of filters with weight transfer.

transfer procedure, recursions were derived for the mean and variance of the auxiliary variable $a(n)$, as well as for the EMSE of the combined filter. In special, the model for the variance of $a(n)$, which is based on a second-order Taylor series, is more accurate than our previous model as shown by simulations. The new model avoids an overestimation of the variance during the transition from the fast to the slow filters, what results in a more accurate EMSE prediction.

5. REFERENCES

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

[2] 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, Sept. 2006.

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

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

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

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

[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, pp. 1853–1864, May 2008.

[8] A. T. Erdogan, S. S. Kozat, and A. C. Singer, “Comparison of convex combination and affine combination of adaptive filters,” in *Proc. IEEE ICASSP*, Taipei, Taiwan, 2009, pp. 3089–3092.

[9] V. H. Nascimento, M. T. Silva, R. Candido, and J. Arenas-García, “A transient analysis for the convex combination of adaptive filters,” in *Proc. IEEE Workshop on Statistical Signal Process.*, Cardiff, UK, 2009, pp. 53–56.

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

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

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