STEADY-STATE PERFORMANCE OF CONVEX COMBINATIONS OF ADAPTIVE FILTERS

J. Arenas-García, A. R. Figueiras-Vidal* †

Dept. of Signal Theory and Communications
Universidad Carlos III de Madrid
28911 Leganés, Spain

A. H. Sayed‡

Dept. of Electrical Engineering
University of California
Los Angeles, CA 90095, USA

ABSTRACT

Combination approaches can improve the performance of adaptive schemes. In this paper, we study the steady-state performance of an adaptive convex combination of transversal filters and show its universality in the sense that the combination performs, in steady-state, at least as well as its best component. We specialize the results to a convex combination of LMS filters using energy conservation arguments.

1. INTRODUCTION

Combination approaches can be used to achieve improved adaptive filter performance [1, 2, 3, 4]. In this paper we study the adaptive convex combination scheme of [5, 6], which obtains the output of the overall filter as – see Fig. 1:

$$y(n) = \lambda(n)y_1(n) + [1-\lambda(n)]y_2(n) \quad (1)$$

where $y_1(n)$ and $y_2(n)$ are the outputs of two transversal filters at time n (i.e., $y_i(n) = \mathbf{w}_i^T(n)\mathbf{u}(n)$, $i = 1, 2$, $\mathbf{w}_i^T(n)$ being the weight vectors characterizing the component filters and $\mathbf{u}(n)$ their common regressor vector) and $\lambda(n)$ is a mixing non-negative scalar parameter. The idea is that if $\lambda(n)$ is assigned appropriate values at each iteration, then the above combination will extract the best properties of filters $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$.

We consider the case in which both component filters are independently adapted, using their own design rules. Thus, for general transversal schemes we have

$$\mathbf{w}_i(n+1) = \mathbf{f}_i\left[\mathbf{w}_i(n), \mathbf{u}(n), d(n), \mathbf{p}_i(n)\right], \ i = 1, 2 \quad (2)$$

where $d(n)$ stands for the desired signal, $\mathbf{p}_i(n)$ is a state vector, and $\mathbf{f}_i[\cdot]$ refers to the adaptation function. For simplicity, we shall assume in the following that $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ are equal length filters, so that the overall filter can also be thought of as a transversal filter with weight vector

$$\mathbf{w}(n) = \lambda(n)\mathbf{w}_1(n) + [1-\lambda(n)]\mathbf{w}_2(n) \quad (3)$$

For the adaptation of the mixing parameter $\lambda(n)$ we shall use a stochastic gradient rule that minimizes the quadratic error of the overall filter, $e^2(n) = [d(n) - y(n)]^2$. However, instead of directly modifying $\lambda(n)$, we will adapt a variable $a(n)$ that defines $\lambda(n)$ via a sigmoidal function, i.e., $\lambda(n) = \text{sgm}(a(n)) = \left(1 + e^{-a(n)}\right)^{-1}$. The update equation for $a(n)$ is given by

$$\begin{aligned} a(n+1) &= a(n) - \frac{\mu_a}{2}\frac{\partial e^2(n)}{\partial a(n)} \quad (4) \\ &= a(n) + \mu_a e(n)\left[y_1(n) - y_2(n)\right]\lambda(n)\left[1-\lambda(n)\right] \end{aligned}$$

Note that the factor $\lambda(n)[1-\lambda(n)]$ reduces the adaptation speed, and consequently the gradient noise introduced by (4), near $\lambda(n) = 0$ or $\lambda(n) = 1$. However, the update for $a(n)$ could stop whenever $\lambda(n)$ is too close to these limits. To circumvent this difficulty, we restrict the values of $a(n)$ to lie inside an interval $[-a^+, a^+]$, which limits the permissible range of $\lambda(n)$ to $[1-\lambda^+, \lambda^+]$ where $\lambda^+ = \text{sgm}(a^+)$ is a constant close to 1.

In the rest of this paper we carry out a statistical analysis of the steady-state performance of the combination procedure (1)-(4); in particular, we will show that it is *nearly* universal [7], in the sense that it can perform as close as desired to its best component, and, in certain situations, better than any of them.

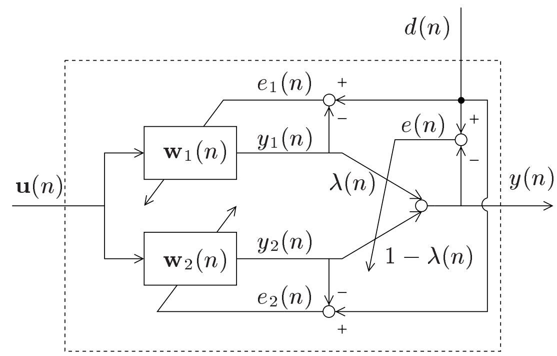

Fig. 1. Adaptive convex combination of two transversal filters. Each component is adapted using its own rules and errors, while the mixing parameter, $\lambda(n)$, is selected to minimize the quadratic error of the overall filter.

2. STEADY-STATE PERFORMANCE

We adopt the following assumptions, which are usually realistic for many applications:

- $d(n)$ and $\mathbf{u}(n)$ are related via $d(n) = \mathbf{w}_0^T\mathbf{u}(n) + e_0(n)$, for some unknown weight vector $\mathbf{w}_0$ and where $e_0(n)$ is an independent and identically distributed (i.i.d.) noise, independent of $\mathbf{u}(m)$ for any n and m, and with variance σ_0^2.

*The work of J. Arenas-García and A. R. Figueiras-Vidal was partly supported by CICYT grant TIC2002-03713.

†This work was performed while the first author was visiting the UCLA Adaptive Systems Laboratory. His work was also supported by the Spanish Ministry of Education, Culture and Sports under the FPU program.

‡The work of A. H. Sayed was supported by NSF grants CCF-0208573 and ECS-0401188.

- The initial conditions $\mathbf{w}_1(0)$, $\mathbf{w}_2(0)$ and $a(0)$ are independent of $\{\mathbf{u}(n), d(n), e_0(n)\}$ for all n.
- $E\{\mathbf{u}(n)\} = \mathbf{0}$, $E\{\mathbf{u}(n)\mathbf{u}^T(n)\} = \mathbf{R}$, $E\{d(n)\} = 0$, and $E\{e_0(n)\} = 0$.

It is also convenient to introduce some notation and additional variables:

- Weight error vectors: $\boldsymbol{\varepsilon}_i(n) = \mathbf{w}_0 - \mathbf{w}_i(n),\ i = 1, 2$ for the component filters, and $\boldsymbol{\varepsilon}(n) = \mathbf{w}_0 - \mathbf{w}(n)$ for their combination.
- A-priori errors: $e_{a,i}(n) = \boldsymbol{\varepsilon}_i^T(n)\mathbf{u}(n), i = 1, 2$, and $e_a(n) = \boldsymbol{\varepsilon}^T(n)\mathbf{u}(n)$.
- A-posteriori errors: $e_{p,i}(n) = \boldsymbol{\varepsilon}_i^T(n+1)\mathbf{u}(n), i = 1, 2$, and $e_p(n) = \boldsymbol{\varepsilon}^T(n+1)\mathbf{u}(n)$.
- Output error: $e(n) = d(n) - y(n)$.

To measure filter performance it is customary to use the excess mean-square error (EMSE), which is defined as the steady-state excess over the minimum achievable error (σ_0^2). When analyzing steady-state operation, we are mainly interested in the limiting value as n goes to ∞. It can be easily seen that $e(n) = e_a(n)+e_0(n)$ and the EMSE of the filters (isolated and combined) can be calculated as:

$$J_{ex,i}(\infty) = \lim_{n\to\infty} E\{e_{a,i}^2(n)\},\ i = 1, 2 \tag{5}$$

$$J_{ex}(\infty) = \lim_{n\to\infty} E\{e_a^2(n)\} \tag{6}$$

2.1. Steady-state EMSE of the combination

To obtain an expression that relates the EMSE of the overall filter to those of its components we subtract both terms of (1) from $d(n)$ and find that

$$\begin{aligned} e(n) &= \lambda(n)e_1(n) + [1 - \lambda(n)]e_2(n) \\ &= \lambda(n)e_{a,1}(n) + [1 - \lambda(n)]e_{a,2}(n) + e_0(n) \end{aligned} \tag{7}$$

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

$$e_a(n) = \lambda(n)e_{a,1}(n) + [1 - \lambda(n)]e_{a,2}(n) \tag{8}$$

Now, taking the limit of $E\{e_a^2(n)\}$ as n goes to ∞ we have

$$\begin{aligned} J_{ex}(\infty) = \lim_{n\to\infty} E\{&\lambda^2(n)e_{a,1}^2(n) + [1 - \lambda(n)]^2 e_{a,2}^2(n) \\ &+2\lambda(n)[1 - \lambda(n)]e_{a,1}(n)e_{a,2}(n)\} \end{aligned} \tag{9}$$

The appearance of a cross-expectation term between the mixing parameter $\lambda(n)$ and the a-priori errors of the component filters in (9), together with the fact that the equation governing the adaption of $a(n)$ is nonlinear in $\lambda(n)$, makes the exact evaluation of (9) a difficult task. However, there exist two cases in which $J_{ex}(\infty)$ could be evaluated in a simpler manner:

- If $a(n) \to a^+$ as $n \to \infty$ a.s. (almost surely), then $\lambda(n) \to \lambda^+$ a.s. and (9) could be simplified to

  $$\begin{aligned} J_{ex}(\infty) = {\lambda^+}^2 J_{ex,1}(\infty) &+ (1 - \lambda^+)^2 J_{ex,2}(\infty) \\ &+ 2\lambda^+(1 - \lambda^+)J_{ex,12}(\infty) \end{aligned}$$

  where we have defined the cross-EMSE of the two filters as $J_{ex,12}(\infty) = \lim_{n\to\infty} E\{e_{a,1}(n)e_{a,2}(n)\}$. Then, using the fact that λ^+ is close to one, we conclude that

  $$\boxed{J_{ex}(\infty) \approx J_{ex,1}(\infty)} \tag{10}$$

  where the approximation is as accurate as desired by increasing the value of a^+.
- Similarly, if $a(n) \to -a^+$ as $n \to \infty$ a.s., we conclude that

  $$\begin{aligned} J_{ex}(\infty) = (1 - \lambda^+)^2 J_{ex,1}(\infty) &+ {\lambda^+}^2 J_{ex,2}(\infty) \\ &+ 2(1 - \lambda^+)\lambda^+ J_{ex,12}(\infty)J_{ex,2}(\infty) \end{aligned} \tag{11}$$

  i.e.,

  $$\boxed{J_{ex}(\infty) \approx J_{ex,2}(\infty)} \tag{12}$$

So let us examine the steady-state behavior of $a(n)$. Taking expectations of both sides of (4), we arrive at

$$\begin{aligned} &E\{a(n + 1)\} \approx \\ &[E\{a(n)\} + \mu_a E\{\lambda(n)[1 - \lambda(n)]e(n)[y_1(n) - y_2(n)]\}]_{-a^+}^{a^+} \end{aligned} \tag{13}$$

where the square brackets denote truncation to the indicated values, as explained after (4). Note that this is only an approximation since we have switched the order of the expectation and the truncation operators on the right-hand side. The approximation seems reasonable because the likelihood of $a(n)$ before truncation being much higher than a^+ or much lower than $-a^+$ is small. To see this, note that the closer $a(n)$ is to the limits, the smaller the magnitude of the update factor $\lambda(n)[1 - \lambda(n)]$.

Now, introducing (7) into (13), and using the relation $y_1(n) - y_2(n) = e_{a,2}(n) - e_{a,1}(n)$, we obtain an expression relating the adaptation of the mixing parameter to the a-priori errors of the component filters:

$$\begin{aligned} E\{a(n + 1)\} = [E\{a\} &+ \mu_a E\{\lambda(1 - \lambda)e_0(e_{a,2} - e_{a,1})\} \\ &+ \mu_a E\{\lambda(1 - \lambda)^2(e_{a,2}^2 - e_{a,1}e_{a,2})\} \\ &+\mu_a E\{\lambda^2(1 - \lambda)(e_{a,1}e_{a,2} - e_{a,1}^2)\}]_{-a^+}^{a^+} \end{aligned} \tag{14}$$

where, for compactness, we have omitted the time index n in the right hand side of the expression.

In (14), the expectation term depending on $e_0(n)$ vanishes as a consequence of $e_0(n)$ being independent of the other factors inside the expectation and $E\{e_0(n)\} = 0$. Regarding the expectations in the second and third lines, they can be simplified by using the following reasonable assumption.

Assumption. *In steady-state, $\lambda(n)$ is independent of the a priori errors of both component filters.*

Thus taking the limit as $n \to \infty$, Eq. (14) becomes

$$\begin{aligned} E\{a(n + 1)\} = \big[E\{a(n)\} &+ \mu_a E\{\lambda(n)[1 - \lambda(n)]^2\}\Delta J_2 \\ &-\mu_a E\{\lambda^2(n)[1 - \lambda(n)]\}\Delta J_1\big]_{-a^+}^{a^+};\ n \to \infty \end{aligned} \tag{15}$$

where we have defined

$$\Delta J_i = J_{ex,i}(\infty) - J_{ex,12}(\infty),\ i = 1, 2 \tag{16}$$

which measures the difference between the individual EMSEs and the cross-EMSE.

Eq. (15) shows that the limiting value of $E\{a(n)\}$ depends on the values of $\Delta J_i,\ i = 1, 2$. It is useful to distinguish among three different situations:[1]

[1]From the definition of $J_{ex,12}(\infty)$, and from Cauchy-Schwartz inequality, it can be seen that the magnitude of $J_{ex,12}(\infty)$ cannot be simultaneously higher than the EMSEs of the two component filters.

1. $J_{ex,1}(\infty) \le J_{ex,12}(\infty) \le J_{ex,2}(\infty)$. In this case, we have $\Delta J_1 \le 0$ and $\Delta J_2 \ge 0$. Furthermore, both $E\{\lambda(n)[1-\lambda(n)]^2\}$ and $E\{\lambda^2(n)[1-\lambda(n)]\}$ are lower bounded by $\lambda^+(1-\lambda^+)^2$. It follows that we can assume

$$E\{a(n+1)\} \ge [E\{a(n)\} + C]_{-a^+}^{a^+}\,; \quad n \to \infty \qquad (17)$$

with C being a positive constant. Therefore, the only valid stationary point for (15) is $E\{a(\infty)\} = a^+$. But since $a(n) \in [-a^+, a^+]$, this suggests that $a(\infty) = a^+$ a.s. As we have already explained, when this occurs we have $J_{ex}(\infty) \approx J_{ex,1}(\infty)$, and the combination performs like its best component filter.

2. $J_{ex,1}(\infty) \ge J_{ex,12}(\infty) \ge J_{ex,2}(\infty)$. Now, we have $\Delta J_1 \ge 0$ and $\Delta J_2 \le 0$, allowing us to write

$$E\{a(n+1)\} \le [E\{a(n)\} - C]_{-a^+}^{a^+}\,; \quad n \to \infty \qquad (18)$$

for a positive constant C. Applying parallel arguments to those in the previous case, we conclude that $a(\infty) = -a^+$ and $J_{ex}(\infty) \approx J_{ex,2}(\infty)$. Again, the behavior of the overall filter is as good as its best element.

3. $J_{ex,12}(\infty) < J_{ex,i}(\infty),\ i = 1, 2$. When the cross-EMSE is lower than the EMSE of the two individual filters, we have $\Delta J_i > 0, i = 1, 2$, and a stationary point of (15) may be approximately characterized by the condition

$$E\{\lambda(n)[1-\lambda(n)]^2\}\Delta J_2 = E\{\lambda^2(n)[1-\lambda(n)]\}\Delta J_1; \quad n \to \infty \qquad (19)$$

It is difficult to derive from the above relation an expression for $\bar{\lambda}(\infty) = \lim_{n\to\infty} E\{\lambda(n)\}$. Further understanding about the performance of the system can be obtained by assuming (only for this third case) that the variance of $\lambda(n)$ is small for $n \to \infty$. Proceeding in this way, it is immediate to obtain:

$$[1 - \bar{\lambda}(\infty)]\Delta J_2 = \bar{\lambda}(\infty)\Delta J_1 \qquad (20)$$

from which we can set

$$\bar{\lambda}(\infty) = \left[\frac{\Delta J_2}{\Delta J_1 + \Delta J_2}\right]_{1-\lambda^+}^{\lambda^+} \qquad (21)$$

so that

$$\lambda^+ \ge \bar{\lambda}(\infty) > 0.5; \quad \text{if } J_{ex,1}(\infty) < J_{ex,2}(\infty)$$
$$0.5 > \bar{\lambda}(\infty) \ge 1 - \lambda^+; \quad \text{if } J_{ex,1}(\infty) > J_{ex,2}(\infty)$$

For $\bar{\lambda}(\infty) = \lambda^+$ and $\bar{\lambda}(\infty) = 1 - \lambda^+$ we already know that the performance of the combination is that of its best component. However, for intermediate values of $\bar{\lambda}(\infty)$ in (21), the overall filter need not converge towards the best component. This behavior of the mixing parameter does not imply that the combination is suboptimal in this third case. In fact, the behavior may be superior to both component filters. Indeed, it can be verified that the value for $\bar{\lambda}(\infty)$ in (21) (neglecting the truncation) is the one that minimizes (9) under the assumption of zero variance for $\lambda(\infty)$:

$$J_{ex}(\infty) = \bar{\lambda}^2(\infty)J_{ex,1}(\infty) + [1-\bar{\lambda}(\infty)]^2 J_{ex,2}(\infty) + 2\bar{\lambda}(\infty)[1-\bar{\lambda}(\infty)]J_{ex,12}(\infty)$$

Introducing (21) into this expression, we get after some algebra

$$J_{ex}(\infty) = J_{ex,12}(\infty) + \frac{\Delta J_1 \Delta J_2}{\Delta J_1 + \Delta J_2} \qquad (22)$$

so that, since $\bar{\lambda}(\infty) < 1$, the following bounds hold:

$$J_{ex}(\infty) = J_{ex,12}(\infty) + \bar{\lambda}(\infty)\Delta J_1 < J_{ex,1}(\infty)$$
$$J_{ex}(\infty) = J_{ex,12}(\infty) + [1-\bar{\lambda}(\infty)]\Delta J_2 < J_{ex,2}(\infty)$$

i.e.,

$$\boxed{J_{ex}(\infty) < \min[J_{ex,1}(\infty), J_{ex,2}(\infty)]} \qquad (23)$$

In summary, the above three cases allow us to conclude that the combination procedure (1)-(4) is *nearly* universal, i.e., its steady-state performance is as close as desired to its best component filter (for a sufficiently high a^+). Furthermore, when certain conditions are met, the combination outperforms both components.

The above steady-state analysis of the combination scheme (1) applies to general adaptive filters (2). The analysis did not assume any particular form for the update function $\mathbf{f}_i[\cdot]$. To study the overall filter performance for a particular update of $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ it is enough to derive expressions for the associated EMSEs and cross-EMSE. We will do so in the following section for a convex combination of LMS filters operating in a stationary scenario; it will turn out that the third scenario described above does not arise in this case; though it occurs for nonstationary environments and for some other filter combinations.

3. COMBINATION OF LMS FILTERS

In this section we study the stationary performance of an adaptive convex combination of two LMS filters (CLMS), which only differ in their step-sizes. Designing criteria for hard switching the step-size of an LMS filter (in a variable step-size implementation) is generally challenging; in this sense, CLMS could be thought of as an effective method for (softly) discriminating between the best of the μ_1 and μ_2 step-sizes.

Without loss of generality, we will assume that $\mu_1 > \mu_2$ so that the first filter adapts faster. Using the energy conservation approach of [8, Ch. 6], and assuming that at steady-state operation $\|\mathbf{u}(n)\|^2$ is independent of $e_{a,i}(n)$, it is known that the EMSEs of the LMS components are given by [8, Eq. (6.5.11)]

$$J_{ex,i}(\infty) = \frac{\mu_i \sigma_0^2 \mathrm{Tr}(\mathbf{R})}{2 - \mu_i \mathrm{Tr}(\mathbf{R})}; \quad \mu_i < \frac{2}{\mathrm{Tr}(\mathbf{R})} \qquad (24)$$

which, in passing, we note that it is an increasing function of μ_i over $\mu_i < 2/\mathrm{Tr}(\mathbf{R})$.

To derive an expression for $J_{ex,12}(\infty)$ we will proceed from [8, Eq. (6.3.7)] which relates the weight, a-priori and a-posteriori errors of a general class of adaptive filters:

$$\boldsymbol{\varepsilon}_i(n+1) + \frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2} e_{a,i}(n) = \boldsymbol{\varepsilon}_i(n) + \frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2} e_{p,i}(n) \quad (25)$$

Multiplying the transpose of (25) by (25) itself for $i = 1$ and 2, respectively, we arrive at the following generalized energy conservation relation:

$$\boldsymbol{\varepsilon}_1^T(n+1)\boldsymbol{\varepsilon}_2(n+1) + \frac{e_{a,1}(n)e_{a,2}(n)}{\|\mathbf{u}(n)\|^2} = \boldsymbol{\varepsilon}_1^T(n)\boldsymbol{\varepsilon}_2(n) + \frac{e_{p,1}(n)e_{p,2}(n)}{\|\mathbf{u}(n)\|^2} \qquad (26)$$

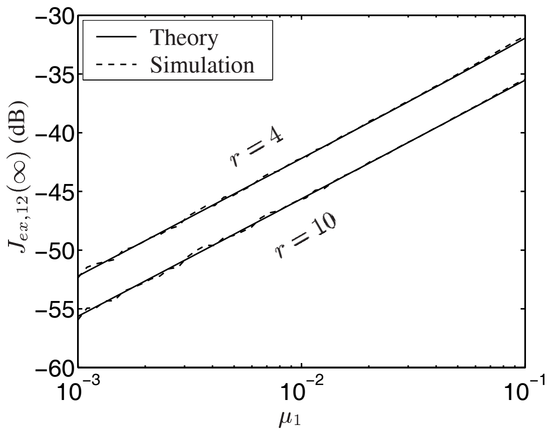

Fig. 2. Steady-state theoretical and estimated cross-EMSE of two LMS filters for different values of μ_1 and $\mu_2 = \mu_1/r$.

Taking expectations, and using the fact that in steady-state

$$E\{\boldsymbol{\varepsilon}_1^T(n+1)\boldsymbol{\varepsilon}_2(n+1)\} = E\{\boldsymbol{\varepsilon}_1^T(n)\boldsymbol{\varepsilon}_2(n)\};n\to\infty$$

we get

$$E\left\{\frac{e_{a,1}(n)e_{a,2}(n)}{\|\mathbf{u}(n)\|^2}\right\} = E\left\{\frac{e_{p,1}(n)e_{p,2}(n)}{\|\mathbf{u}(n)\|^2}\right\};n\to\infty \quad (27)$$

The above expression holds for any filters $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$. When these are restricted to the LMS case, it can be shown that their a priori and a-posteriori errors are related via [8, Eq. (6.3.3)]

$$e_{p,i}(n) = e_{a,i}(n) - \mu_i\|\mathbf{u}(n)\|^2 e_i(n), \;\; i=1,2 \quad (28)$$

Introducing (28) into (27), and multiplying terms we get

$$\begin{aligned}\mu_1\mu_2 E\{\|\mathbf{u}(n)\|^2\, e_1(n)e_2(n)\} &= \mu_1 E\{e_{a,2}(n)e_1(n)\}\\ &+ \mu_2 E\{e_{a,1}(n)e_2(n)\};\; n\to\infty \quad (29)\end{aligned}$$

Now, using $e_i(n) = e_{a,i}(n) + e_0(n)$, and using the independence of $e_0(n)$ with respect to the other variables, we get

$$J_{ex,12}(\infty) = \frac{\mu_{12}}{2}\left[E\{\|\mathbf{u}(n)\|^2 e_{a,1}(n)e_{a,2}(n)\} + \sigma_0^2\mathrm{Tr}(\mathbf{R})\right]$$

as n goes to ∞, where we have defined

$$\mu_{12} = (2\mu_1\mu_2)/(\mu_1+\mu_2) \quad (30)$$

Finally, and using again that $\|\mathbf{u}(n)\|^2$ is independent of $e_{a,i}(n)$ in steady-state, the cross-EMSE of the filters is given by

$$J_{ex,12}(\infty) = \frac{\mu_{12}\sigma_0^2\mathrm{Tr}(\mathbf{R})}{2-\mu_{12}\mathrm{Tr}(\mathbf{R})} \quad (31)$$

The similarity that exist between (31) and (24), together with the fact that $\mu_1 \ge \mu_{12} \ge \mu_2$, allows us to conclude that $J_{ex,1}(\infty) \ge J_{ex,12}(\infty) \ge J_{ex,2}(\infty)$, so we are in the second of the situations considered in Section 2. Consequently, the mean-square error of the CLMS filter is approximately equal to that of the μ_2-LMS filter: $J_{ex}(\infty) \approx J_{ex,2}(\infty)$.

Figure 2 plots expression (31) for different values of μ_1 and $\mu_2 = \mu_1/r$. The settings of the example are $M = 3$, $\mathbf{w}_0^T = [0.8212, 0.1620, -0.5897]$, and $e_0(n)$ and $u(n)$ are independent zero-mean Gaussian processes, with variances $\sigma_0^2 = 0.01$ and $\sigma_u^2 = 1$, respectively. Each estimated cross-EMSE was calculated by averaging $e_{a,1}(n)e_{a,2}(n)$ for 20000 iterations after the convergence of both LMS filters, and over 25 independent runs.

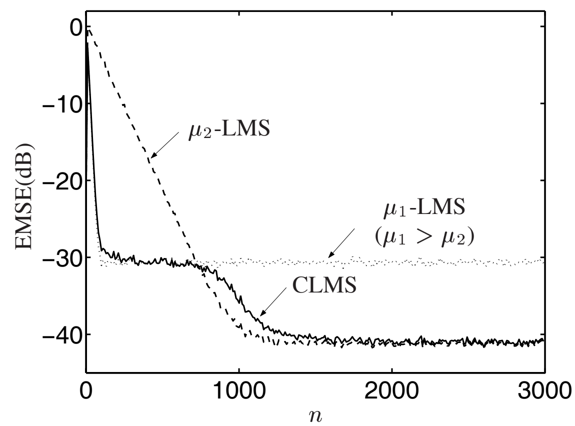

Fig. 3. Transient and steady-state performance of an adaptive combination of two LMS filters ($\mu_1 = 0.05$ and $\mu_2 = 0.005$) with parameters $\mu_a = 200$ and $a^+ = 4$.

In Figure 3, we have depicted the transient performance of the adaptive combination of two LMS filters with $\mu_1 = 0.05$ and $\mu_2 = 0.005$. The parameters of the combination have been set to $\mu_a = 200$ and $a^+ = 4$, while the weight vectors have been initialized with zeros and the mixing parameter to $\lambda(0) = 0.5$. All displayed results have been averaged over 1000 independent runs. Initially, CLMS retains the faster convergence of the μ_1-LMS. However, as predicted by our analysis, its steady-state performance is equivalent to that of the LMS with step-size μ_2.

4. CONCLUSION

Combination approaches can help improve adaptive filters performance. In this paper we analyzed the behavior of one such approaches, showing that it performs as close as desired to the best of its components, and, possibly, better than any of them.

5. REFERENCES

[1] P. Anderson, “Adaptive forgetting in recursive identification through multiple models,” *Intl. J. Control*, vol. 42, pp. 1175–1193, 1985.

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

[3] S. S. Kozat and A. C. Singer, “Multi-stage adaptive signal processing algorithms,” in *Proc. IEEE Sensor Array and Multichannel Signal Proc. Workshop*, Cambridge, MA, 2000, pp. 380–384.

[4] V. Vovk, “Competitive on-line statistics,” *Intl. Stat. Rev.*, vol. 69, pp. 213–248, 2001.

[5] J. Arenas-García, V. Gómez-Verdejo, M. Martínez-Ramón, and A. R. Figueiras-Vidal, “Separate-variable adaptive combination of LMS adaptive filters for plant identification,” in *Proc. of the 13th IEEE Intl. Workshop on Neural Networks for Signal Proc.*, Toulouse, France, 2003, pp. 239–248.

[6] J. Arenas-García, M. Martínez-Ramón, V. Gómez-Verdejo, and A. R. Figueiras-Vidal, “Multiple plant identifier via adaptive LMS convex combination,” in *Proc. IEEE Intl. Symp. on Intel. Signal Proc.*, Budapest, Hungary, 2003, pp. 137–142.

[7] N. Merhav and M. Feder, “Universal prediction,” *IEEE Trans. on Inf. Theory*, vol. 44, pp. 2124–2147, 1998.

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