TRACKING PROPERTIES OF A CONVEX COMBINATION OF TWO 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 provide a useful way to improve adaptive filter performance. In this paper, we study the tracking performance of an adaptive convex combination of two transversal filters. The individual filters are independently adapted using their own error signals, while the combination is adapted to minimize the error of the overall structure. We show the universality of the approach with respect to the component filters, i.e., that the overall filter is able to track changes at least as well as the best component filter. Using energy conservation arguments, we then specialize the results to a combination of two LMS filters.

1. INTRODUCTION

Combination approaches can be used to achieve improved adaptive filter performance [1, 2, 3, 4]. In this paper we shall study the following adaptive convex combination scheme, which obtains the output of the overall filter as – see Fig. 1 [5, 6, 7]:

$$\boxed{y(n) = \lambda(n)y_1(n) + [1 - \lambda(n)]y_2(n)} \tag{1}$$

where $y_1(n)$ and $y_2(n)$ are the outputs of two transversal filters at time n, i.e., $y_i(n) = \mathbf{w}_i^T(n)\mathbf{u}(n),\ i = 1, 2$, with $\mathbf{w}_i^T(n)$ being the weight vectors characterizing the component filters and $\mathbf{u}(n)$ their common regressor vector. Moreover, $\lambda(n)$ is a mixing scalar parameter that lies between 0 and 1. The idea is that if $\lambda(n)$ is assigned appropriate values at each iteration, then the above combination would extract the best properties of the individual filters $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$.

We shall consider the case in which both component filters are independently adapted, using their own design rules.

*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.

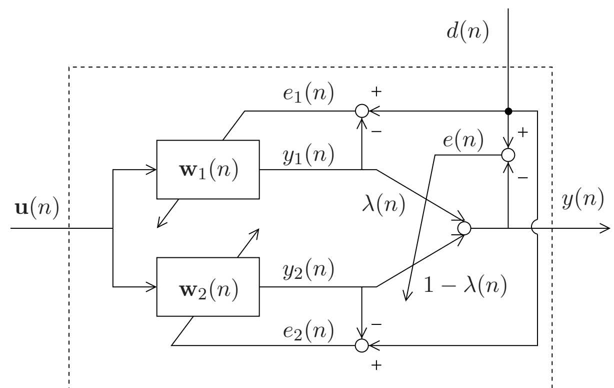

Fig. 1. Adaptive convex combination of two transversal filters. Each component is adapted using its own rules and errors, while the mixing parameter, $\lambda(n)$, is chosen to minimize the quadratic error of the overall filter.

Thus, for general transversal schemes we will assume that

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

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 both $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ have length M, so that the overall filter can also be thought of as a transversal filter with weight vector:

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

For the adaptation of the mixing parameter $\lambda(n)$ we shall use a gradient descent method to minimize the quadratic error of the overall filter, namely, $e^2(n) = [d(n) - y(n)]^2$. However, instead of directly modifying $\lambda(n)$, we adapt a variable $a(n)$ that defines $\lambda(n)$ via a sigmoidal function as

$$\lambda(n) = \text{sgm}[a(n)] = \frac{1}{1 + e^{-a(n)}}$$

The update equation for $a(n)$ is given by

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

The benefits of employing the sigmoidal activation function are twofold. First, it serves to keep $\lambda(n)$ within the desired range [0,1]. Second, as seen from (4), the adaptation rule of $a(n)$ reduces both the stochastic gradient noise and the adaptation speed near $\lambda(n) = 1$ and $\lambda(n) = 0$ when the combination is expected to perform close to one of its component filters without degradation. Still, note that the update of $a(n)$ in (4) stops whenever $\lambda(n)$ is too close to the limit values of 0 or 1. To circumvent this problem, we shall restrict the values of $a(n)$ to lie inside a symmetric 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 this way, a minimum level of adaption is always guaranteed.

In [7] we analyzed the stationary performance of the above combination scheme, concluding that $y(n)$ is *nearly* universal with respect to its components [8], i.e., it can perform as close as desired to the best component filter. In this paper we extend this result to non-stationary environments and illustrate the performance of a convex combination of two LMS filters when the optimal solution is subject to changes at different speeds.

2. NON-STATIONARY DATA MODEL AND NOTATION

In the sequel we adopt the following assumptions:

- $d(n)$ and $\mathbf{u}(n)$ are related via a linear regression model
$$d(n) = \mathbf{w}_0(n)^T \mathbf{u}(n) + e_0(n) \tag{5}$$
for some unknown weight vector $\mathbf{w}_0(n)$ of length M 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 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.

- $\mathbf{w}_0(n)$ varies according to the random walk model
$$\mathbf{w}_0(n+1) = \mathbf{w}_0(n) + \mathbf{q}(n) \tag{6}$$
where $\mathbf{q}(n)$ is an i.i.d. vector, independent of $\{\mathbf{u}(m), d(m), e_0(m)\}$, for all $m < n$, and of the initial conditions $\mathbf{w}_0(0)$, $\mathbf{w}_1(0)$, $\mathbf{w}_2(0)$ and $a(0)$.

- $E\{\mathbf{u}(n)\} = E\{\mathbf{q}(n)\} = \mathbf{0}$, $E\{d(n)\} = E\{e_0(n)\} = 0$, $E\{\mathbf{u}(n)\mathbf{u}^T(n)\} = \mathbf{R}$, and $E\{\mathbf{q}(n)\mathbf{q}^T(n)\} = \mathbf{Q}$. Unlike other studies in the adaptive filtering literature, we shall not assume that $\{\mathbf{u}(n)\}$ is a sequence of independent regressors.

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

- We define the weight error vector of a transversal filter as the difference between the optimal solution and the filter weights. Thus we define
$$\boldsymbol{\varepsilon}_i(n) = \mathbf{w}_0(n) - \mathbf{w}_i(n), \; i = 1, 2$$
for the component filters, and
$$\boldsymbol{\varepsilon}(n) = \mathbf{w}_0(n) - \mathbf{w}(n)$$
for their combination.

- A priori errors:
$$\begin{aligned} e_{a,i}(n) &= \boldsymbol{\varepsilon}_i^T(n)\mathbf{u}(n), \quad i = 1, 2 \\ e_a(n) &= \boldsymbol{\varepsilon}^T(n)\mathbf{u}(n) \end{aligned}$$

- A posteriori errors:
$$\begin{aligned} e_{p,i}(n) &= [\mathbf{w}_0(n) - \mathbf{w}_i(n+1)]^T \mathbf{u}(n), \; i = 1, 2 \\ e_p(n) &= [\mathbf{w}_0(n) - \mathbf{w}(n+1)]^T \mathbf{u}(n) \end{aligned}$$

To measure filter performance it is customary to use the excess mean-square error (EMSE), which is defined as the excess over the minimum mean-square error that can be achieved by a filter of length M in steady-state, namely σ_0^2. It can be easily seen that $e(n) = e_a(n) + e_0(n)$, so that the EMSE of the filters (isolated and combined) can be calculated as:

$$J_{ex,i}(\infty) = \lim_{n\to\infty} E\{e_{a,i}^2(n)\}, \; i = 1, 2 \quad \text{(individual)} \tag{7}$$
$$J_{ex}(\infty) = \lim_{n\to\infty} E\{e_a^2(n)\} \quad \text{(combination)} \tag{8}$$

During the analysis, it will be useful to refer to an additional variable that measures the steady-state correlation between the a priori errors of the elements of the combination, i.e.,

$$\boxed{J_{ex,12}(\infty) = \lim_{n\to\infty} E\{e_{a,1}(n)e_{a,2}(n)\}} \tag{9}$$

We shall refer to this variable as the cross-EMSE of the component filters. From its definition, and from Cauchy-Schwartz inequality, it follows that $J_{ex,12}(\infty)$ cannot be simultaneously higher than the individual EMSEs of filters $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$, i.e., $J_{ex,1}(\infty)$ and $J_{ex,2}(\infty)$.

3. UNIVERSALITY OF THE COMBINATION SCHEME

In [7, 9] we studied the steady-state performance of the adaptive combination form (1), concluding that it is nearly universal, in the sense that its stationary EMSE is as close as desired to the best component filter, $\mathbf{w}_1(n)$ or $\mathbf{w}_2(n)$, for increasing a^+. Furthermore, when the cross-EMSE of the component filters is low enough, the combination was shown to outperform both components. To be more specific, depending on the EMSEs and cross-EMSE of the component filters, and following the arguments in [7, 9], the following three cases can occur:

- Case I: $J_{ex,1}(\infty) \leq J_{ex,12}(\infty) \leq J_{ex,2}(\infty)$. In this case, we obtained that the stationary value of $a(n)$, as n goes to infinity, is essentially $a(\infty) = a^+$ a.s. A consequence of this result is that

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

with the approximation being as accurate as desired for increasing values of a^+. That is, the combination scheme performs like its best component filter in this case.

- Case II: $J_{ex,1}(\infty) \geq J_{ex,12}(\infty) \geq J_{ex,2}(\infty)$. Applying parallel arguments to those in the previous case, we concluded that $a(\infty) = -a^+$ a.s. and

$$\boxed{J_{ex}(\infty) \approx J_{ex,2}(\infty)} \tag{11}$$

Again, the behavior of the overall filter is as good as its best element.

- Case III: $J_{ex,12}(\infty) < J_{ex,i}(\infty),\ i = 1, 2$. When the cross-EMSE is lower than the EMSEs of both individual filters, a stationary value of the mixing parameter $\lambda(n)$ is approximately characterized by

$$\bar{\lambda}(\infty) = \left[\frac{\Delta J_2}{\Delta J_1 + \Delta J_2}\right]_{1-\lambda^+}^{\lambda^+} \tag{12}$$

where we have introduced the differences

$$\Delta J_i = J_{ex,i}(\infty) - J_{ex,12}(\infty),\ i = 1, 2 \tag{13}$$

For $\bar{\lambda}(\infty) = \lambda^+$ and $\bar{\lambda}(\infty) = 1 - \lambda^+$, the two first cases show us that the performance of the combination is that of its best component. For intermediate values of $\bar{\lambda}(\infty)$, the EMSE of the overall filter can be expressed as

$$J_{ex}(\infty) = J_{ex,12}(\infty) + \frac{\Delta J_1 \Delta J_2}{\Delta J_1 + \Delta J_2} \tag{14}$$

so that, since $0 < 1 - \lambda^+ < \bar{\lambda}(\infty) < \lambda^+ < 1$, the following bounds hold:

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

i.e.,

$$\boxed{J_{ex}(\infty) < \min\{J_{ex,1}(\infty), J_{ex,2}(\infty)\}} \tag{15}$$

Since our arguments in [7, 9] relied solely on the a priori errors of the filters, and these definitions do not change when analyzing tracking operation, the conclusions about the universality of the mixture hold unaltered for a tracking scenario, as well as for the three cases mentioned above.

The analysis did not assume any particular form for the update function $\mathbf{f}_i[\cdot]$, and it consequently applies to the combination of general adaptive filters (2). To illustrate the overall filter tracking 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.

4. COMBINATION OF LMS FILTERS WITH DIFFERENT STEP-SIZES

In this section we study the non-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.

To begin with, note that when analyzing the tracking properties of LMS filters it is customary to study the influence of the step-size on the performance of the filter for a fixed covariance matrix $\mathbf{Q}$. However, our goal here is to show that CLMS is able to improve over the tracking capabilities of its components and, consequently, we will analyze the EMSE of the filters for varying $\text{Tr}(\mathbf{Q})$ and for given μ_1 and μ_2. 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 [10, Ch. 7], it can be shown that the tracking EMSEs of the LMS components are given by [10, Eq. (7.5.9)]:

$$\boxed{J_{ex,i}(\infty) = \frac{\mu_i \sigma_0^2 \text{Tr}(\mathbf{R}) + \mu_i^{-1}\text{Tr}(\mathbf{Q})}{2 - \mu_i \text{Tr}(\mathbf{R})}; \quad \mu_i < \frac{2}{\text{Tr}(\mathbf{R})}} \tag{16}$$

As it is known, the EMSE expression consists of two terms. The first term is the EMSE corresponding to a stationary environment (and it increases with μ_i), while the second term is inversely proportional to μ_i and is related to the tracking capabilities of the filter. Consequently, it can be shown [9] that (16) achieves a minimum over the interval $0 < \mu_i < 2/\text{Tr}(\mathbf{R})$ at the following optimal step-size:

$$\mu_{opt} = \sqrt{\frac{\text{Tr}(\mathbf{Q})}{\sigma_0^2 \text{Tr}(\mathbf{R})} + \frac{(\text{Tr}(\mathbf{Q}))^2}{4\sigma_0^4}} - \frac{\text{Tr}(\mathbf{Q})}{2\sigma_0^2} \tag{17}$$

In the following, it will be useful to employ an alternative figure of merit to measure filter performance in tracking situations. We define the Normalized Square Deviation (NSD) of a filter as the ratio of its EMSE to the theoretical EMSE of an LMS filter with optimal step-size μ_{opt} (for each value of $\text{Tr}(\mathbf{Q})$). Thus, for the components of the CLMS filter we set

$$\boxed{\text{NSD}_i(\infty) = J_{ex,i}(\infty)/J_{ex,opt}(\infty),\ i = 1,2} \tag{18}$$

Similarly, the NSDs of the combined scheme, and the cross-NSD between the component filters will be defined as

$$\text{NSD}(\infty) = J_{ex}(\infty)/J_{ex,opt}(\infty) \tag{19}$$

$$\text{NSD}_{12}(\infty) = J_{ex,12}(\infty)/J_{ex,opt}(\infty) \tag{20}$$

Next, we need to obtain an expression for $J_{ex,12}(\infty)$ in the non-stationary case. Our starting point is the following relation from [10, Ch. 7], which applies to LMS updates,

$$[\mathbf{w}_0(n) - \mathbf{w}_i(n+1)] + \frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2} e_{a,i}(n) = \varepsilon_i(n) + \frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2} e_{p,i}(n),\ i = 1,2 \tag{21}$$

Multiplying the transpose of (21) by (21) itself, for $i = 1$ and $i = 2$, respectively, and after cancelling terms, we get

$$[\mathbf{w}_0(n) - \mathbf{w}_1(n+1)]^T [\mathbf{w}_0(n) - \mathbf{w}_2(n+1)] + \frac{e_{a,1}(n) e_{a,2}(n)}{\|\mathbf{u}(n)\|^2} = \varepsilon_1^T(n) \varepsilon_2(n) + \frac{e_{p,1}(n) e_{p,2}(n)}{\|\mathbf{u}(n)\|^2} \tag{22}$$

To simplify this expression, we use $\mathbf{w}_0(n) = \mathbf{w}_0(n+1) - \mathbf{q}(n)$ to write

$$\begin{aligned} E\{[\mathbf{w}_0(n) - \mathbf{w}_1(n+1)]^T [\mathbf{w}_0(n) - \mathbf{w}_2(n+1)]\} = \\ E\{\varepsilon_1^T(n+1) \varepsilon_2(n+1)\} + \text{Tr}(\mathbf{Q}) - \\ E\{[\varepsilon_1(n+1) + \varepsilon_2(n+1)]^T \mathbf{q}(n)\} \end{aligned} \tag{23}$$

Now, note that

$$\varepsilon_i(n+1) = \mathbf{w}_0(0) + \sum_{j=0}^{n} \mathbf{q}(j) - \mathbf{w}_i^T(n+1)$$

But since $\mathbf{q}(n)$ is independent of $\mathbf{w}_0(0)$ by assumption, and $\mathbf{q}(n)$ is also independent of $\mathbf{w}_i^T(n+1)$ as a consequence of $\mathbf{q}(n)$ being i.i.d. and independent of $\{\mathbf{u}(m), d(m)\}, m \leq n$, we can simplify (23) to

$$\begin{aligned} E\{[\mathbf{w}_0(n) - \mathbf{w}_1(n+1)]^T [\mathbf{w}_0(n) - \mathbf{w}_2(n+1)]\} = \\ E\{\varepsilon_1^T(n+1) \varepsilon_2(n+1)\} - \text{Tr}(\mathbf{Q}) \end{aligned} \tag{24}$$

Using this result, and the fact that in steady state

$$E\{\varepsilon_1^T(n+1) \varepsilon_2(n+1)\} = E\{\varepsilon_1^T(n) \varepsilon_2(n)\}$$

we get from (22) that

$$E\left\{\frac{e_{a,1}(n) e_{a,2}(n)}{\|\mathbf{u}(n)\|^2}\right\} = E\left\{\frac{e_{p,1}(n) e_{p,2}(n)}{\|\mathbf{u}(n)\|^2}\right\} + \text{Tr}(\mathbf{Q}) \tag{25}$$

as n goes to infinity.

To proceed further, we need to specialize this result for $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$. From [10, Eq. (6.3.3)],

$$e_{p,i}(n) = e_{a,i}(n) - \mu_i \|\mathbf{u}(n)\|^2 e_i(n),\ \ i = 1,2 \tag{26}$$

and $e_i(n) = e_{a,i}(n) + e_0(n)$, we arrive after some manipulations at

$$\begin{aligned} &(\mu_1 + \mu_2) J_{ex,12}(\infty) = \\ &\text{Tr}(\mathbf{Q}) + \mu_1 \mu_2 \left[E\{\|\mathbf{u}(n)\|^2 e_{a,1}(n) e_{a,2}(n)\} + \sigma_0^2 \text{Tr}(\mathbf{R})\right] \end{aligned} \tag{27}$$

Finally, application of the separation principle (viz., that $\|\mathbf{u}(n)\|^2$ is independent of $e_{a,i}(n)$ in steady-state) allows us to rearrange terms and to get

$$\boxed{J_{ex,12}(\infty) = \frac{\mu_{12} \sigma_0^2 \text{Tr}(\mathbf{R}) + 2\text{Tr}(\mathbf{Q})/(\mu_1 + \mu_2)}{2 - \mu_{12} \text{Tr}(\mathbf{R})}} \tag{28}$$

As discussed in Section 3, the performance of the CLMS filter depends on the signs of $\Delta J_i,\ i = 1,2$. So, to study CLMS performance we need to analyze the relations among $J_{ex,1}(\infty)$, $J_{ex,2}(\infty)$ and $J_{ex,12}(\infty)$ for any value of $\text{Tr}(\mathbf{Q})$ in the non-stationary case. In order to do this, let us consider first a small step-size approximation for (16), (17) and (28), namely

$$J_{ex,i}(\infty) \approx \frac{\mu_i \sigma_0^2 \text{Tr}(\mathbf{R}) + \mu_i^{-1} \text{Tr}(\mathbf{Q})}{2},\ i = 1,2 \tag{29}$$

$$\mu_{opt} \approx \sqrt{\frac{\text{Tr}(\mathbf{Q})}{\sigma_0^2 \text{Tr}(\mathbf{R})}} \tag{30}$$

$$J_{ex,12}(\infty) \approx \frac{\mu_{12} \sigma_0^2 \text{Tr}(\mathbf{R}) + 2\text{Tr}(\mathbf{Q})/(\mu_1 + \mu_2)}{2} \tag{31}$$

Then, subtracting (31) from (29), and defining $r = \mu_1/\mu_2$, it is straightforward to verify that

$$\Delta J_1 = \frac{r-1}{r+1} \cdot \frac{\mu_1 \sigma_0^2 \text{Tr}(\mathbf{R}) - \mu_1^{-1} \text{Tr}(\mathbf{Q})}{2} \quad \begin{matrix} < 0, & \text{for } \text{Tr}(\mathbf{Q}) > q_1 \\ > 0, & \text{for } \text{Tr}(\mathbf{Q}) < q_1 \end{matrix} \tag{32}$$

$$\Delta J_2 = \frac{1-r}{r+1} \cdot \frac{\mu_2 \sigma_0^2 \text{Tr}(\mathbf{R}) - \mu_2^{-1} \text{Tr}(\mathbf{Q})}{2} \quad \begin{matrix} > 0, & \text{for } \text{Tr}(\mathbf{Q}) > q_2 \\ < 0, & \text{for } \text{Tr}(\mathbf{Q}) < q_2 \end{matrix} \tag{33}$$

where we have defined[1]

$$q_i = \mu_i^2 \sigma_0^2 \text{Tr}(\mathbf{R}) \tag{34}$$

Thus, we find that, depending on the value of $\text{Tr}(\mathbf{Q})$, all three cases described in Section 3 can occur – see Table I:

[1] Note that under the small step-size approximation, q_1 (q_2) is the value of $\text{Tr}(\mathbf{Q})$ for which μ_1 (μ_2) is the optimal step-size (see Eq. (30)).

	$\mathrm{NSD}(\infty)$
$\mathrm{Tr}(\mathbf{Q}) > q_1$	$\approx \mathrm{NSD}_1(\infty)$
$q_1 > \mathrm{Tr}(\mathbf{Q}) > q_2$	$< \min\{\mathrm{NSD}_1(\infty), \mathrm{NSD}_2(\infty)\}$
$\mathrm{Tr}(\mathbf{Q}) < q_2$	$\approx \mathrm{NSD}_2(\infty)$

Table 1. CLMS normalized square deviation as a function of $\mathrm{Tr}(\mathbf{Q})$.

- If $\mathrm{Tr}(\mathbf{Q}) > q_1$, we have $J_{ex,1}(\infty) \leq J_{ex,12}(\infty) \leq J_{ex,2}(\infty)$ (or, equivalently, $\mathrm{NSD}_1(\infty) \leq \mathrm{NSD}_{12}(\infty) \leq \mathrm{NSD}_2(\infty)$), so we are in Case I, and the combination performs like the μ_1-LMS.

- Just the opposite occurs for $\mathrm{Tr}(\mathbf{Q}) < q_2$, with $J_{ex}(\infty) \approx J_{ex,2}(\infty)$.

- Finally, when $q_2 < \mathrm{Tr}(\mathbf{Q}) < q_1$, we have that both ΔJ_1 and ΔJ_2 are greater than 0, i.e., (15) applies, and the combination outperforms both components.

Figures 2(a) and (b) illustrate the above theoretical conclusions. We have depicted the theoretical tracking values for the NSDs of two LMS filters with $\mu_1 = 0.1$ and $\mu_2 = 0.01$, for their cross-NSD, and for the NSD achieved by their adaptive convex combination with $a^+ = 4$, for different values of $\mathrm{Tr}(\mathbf{Q})$. Additional settings were $\mathrm{Tr}(\mathbf{R}) = 1$ and $\sigma_0^2 = 0.01$. We can see that the CLMS scheme offers improved tracking performance, not only because it inherits the best tracking properties of each LMS, but also because it performs better than either for certain rates of variations, as can be seen in Figure 2(a) for $q_2 < \mathrm{Tr}(\mathbf{Q}) < q_1$.

We have carried out simulations for an example where the initial optimal solution ($M = 7$) was formed with independent random values between -1 and 1. The regressor $\mathbf{u}(n)$ is obtained from a process $u(n)$ as

$$\mathbf{u}^T(n) = [u(n), u(n-1), \ldots, u(n-6)]$$

where $u(n)$ is colored Gaussian noise with input power. Additive i.i.d. noise $e_0(n)$ with variance $\sigma_0^2 = 0.01$ was added to form the desired signal. Finally, the entries of $\mathbf{q}(n)$ were taken as independent Gaussian values with equal variances. As for the settings for the CLMS filter, we have used step-sizes $\mu_1 = 0.1$ and $\mu_2 = 0.01$ for the components, and $\mu_a = 100$ and $a^+ = 4$ to adapt the combination. Figure 3 shows a close match between the theoretical and estimated values for the residual NSD of the CLMS filter, and the cross-NSD of the component filters for most values of $\mathrm{Tr}(\mathbf{Q})$. All results have been averaged over 20000 samples after filter convergence, and over 50 independent runs.

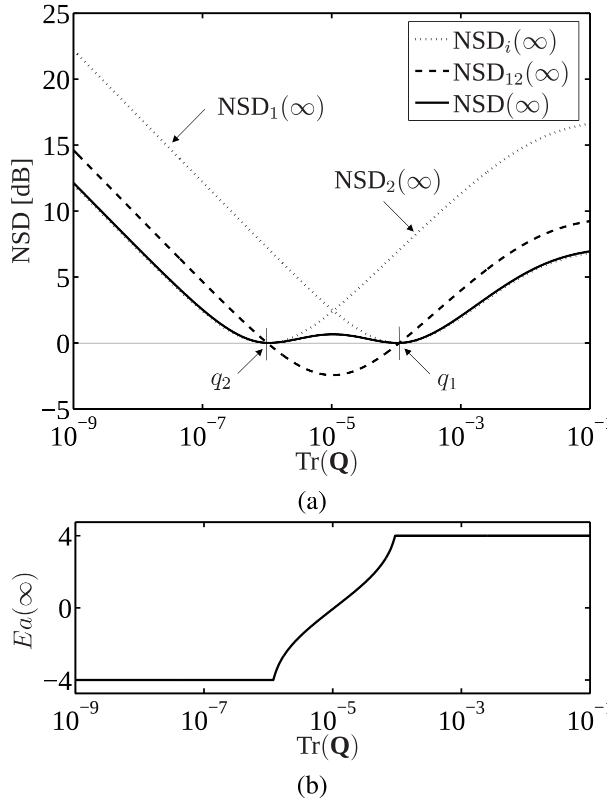

Fig. 2. (a) Theoretical normalized mean-square deviation of two LMS filters with steps $\mu_1 = 0.1$ and $\mu_2 = 0.01$ as a function of the trace of the covariance matrix of $\mathbf{q}(n)$. Their cross-NSD is also depicted using a solid line, as well as it is the NSD of their adaptive combination. (b) Theoretical steady-state value for the mixing parameter.

5. CONCLUSION

Combination approaches can help improve adaptive filter performance. In this paper we have analyzed the tracking behavior of one such approach, showing that it performs as close as desired to the best of its components, and, possibly, better than any of them. These results have been illustrated by studying the tracking performance of a combination of two LMS filters with different step-sizes in a non-stationary environment.

6. 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, Feb. 1992.

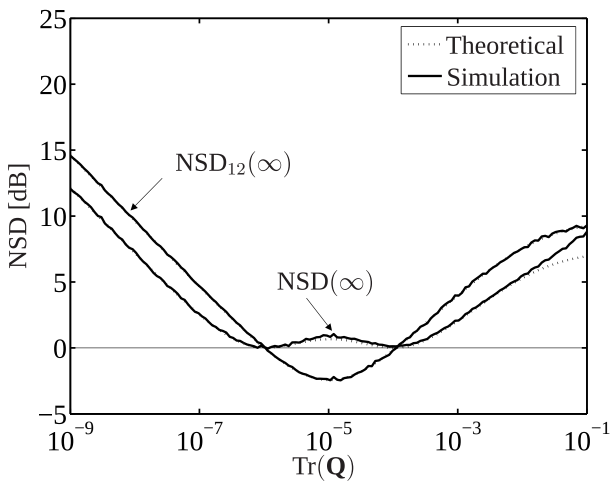

Fig. 3. Steady-state theoretical and estimated cross-NSD of two LMS filters with $\mu_1 = 0.1$ and $\mu_2 = 0.01$, and NSD for the resulting CLMS combination.

[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, pp. 380–384, Mar. 2000.

[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. 13th IEEE Intl. Workshop on Neural Networks for Signal Proc.*, Toulouse, France, pp. 239–248, Sep. 2003.

[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, pp. 137–142, Sep. 2003.

[7] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Steady-state performance of convex combinations of adaptive filters," in *Proc. ICASSP*, vol. IV, Philadelphia, PA, pp. 33-36, Mar. 2005.

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

[9] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," to appear in *IEEE Trans. on Signal Proc.*, 2006.

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