Mean-Square Performance of a Convex Combination of Two Adaptive Filters

Jerónimo Arenas-García, *Member, IEEE*, Aníbal R. Figueiras-Vidal, *Senior Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*

Abstract—Combination approaches provide an interesting way to improve adaptive filter performance. In this paper, we study the mean-square performance of a convex combination of two transversal filters. The individual filters are independently adapted using their own error signals, while the combination is adapted by means of a stochastic gradient algorithm in order to minimize the error of the overall structure. General expressions are derived that show that the method is universal with respect to the component filters, i.e., in steady-state, it performs at least as well as the best component filter. Furthermore, when the correlation between the *a priori* errors of the components is low enough, their combination is able to outperform both of them. Using energy conservation relations, we specialize the results to a combination of least mean-square filters operating both in stationary and in nonstationary scenarios. We also show how the universality of the scheme can be exploited to design filters with improved tracking performance.

Index Terms—Adaptive filtering, convex combination, energy conservation, stochastic algorithms.

I. INTRODUCTION

V ARIABLE step-size adaptive filters allow the filters to dynamically adjust their performance in response to conditions in the input data and error signals [1]–[5]. More generally, on-line adaptation of certain filter parameters or even cost functions has been attempted to influence filter performance, such as adjusting the forgetting factor of recursive least squares (RLS) algorithms [6], [7] or minimizing adjustable cost functions [8]–[11]. Recently, there has been an interest in combination schemes, where the outputs of several filters are mixed together to get an overall output of improved quality [12]–[15],
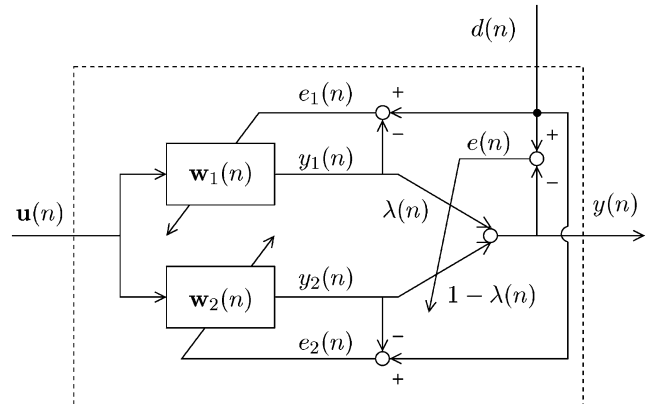
Fig. 1. Adaptive convex combination of two transversal filters. Each component is adapted using its own rules and errors, while the mixing parameter $\lambda(n)$ is chosen to minimize the quadratic error of the overall filter.

with some of these approaches sacrificing performance in lieu of analytical tractability. The issue of how to optimally combine the component filters is a challenging one and deserves closer investigation.

In [16] we presented a combination of one fast and one slow least mean-square (LMS) filter that was effective at combining fast convergence and low residual misadjustment, an approach that was later refined in [17], [18]. This scheme, which can be used with other kinds of filters, calculates the output as the convex combination of the outputs of the component filters and the mixing parameter itself is adjusted on-line by means of a nonlinear rule. The use of the nonlinearity results in improved adaptation of the combination and reduces the gradient noise. In this paper, we pursue a detailed mean-square performance analysis of this scheme for both stationary and nonstationary environments by using energy conservation arguments [19]. In particular, we will show that the combination filter structure is universal [14], [20] in the sense that it performs, in the mean-square error sense, as well as the best of its components. Furthermore, and interestingly, when certain conditions are satisfied by the component filters, their combination will be shown to outperform both of them. The analysis will also suggest a tracking filter with improved performance.

Specifically, we shall study the following adaptive convex combination scheme, which obtains the output of the overall filter as (see Fig. 1) [17], [18]:

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

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 zero and one. 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. Thus, for general transversal schemes, we will assume that

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

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 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

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

For the adaptation of the mixing parameter $\lambda(n)$ we shall use a 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 will adapt a variable $a(n)$ that defines $\lambda(n)$ via a sigmoidal function as

$$\lambda(n) = \text{sgm}[a(n)] = \frac{1}{1 + e^{-a(n)}}. \quad (4)$$

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

$$
\begin{aligned}
a(n+1) &= a(n) - \frac{\mu_a}{2}\frac{\partial e^2(n)}{\partial a(n)} \\
&= a(n) - \frac{\mu_a}{2}\frac{\partial e^2(n)}{\partial \lambda(n)}\frac{\partial \lambda(n)}{\partial a(n)} \\
&= a(n) + \mu_a e(n)[y_1(n) - y_2(n)]\lambda(n)[1 - \lambda(n)].
\end{aligned}
$$
$$(5)$$

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 (5), 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, while allowing a fast adaptation for intermediate values of $\lambda(n)$. Still, note that the update of $a(n)$ in (5) stops whenever $\lambda(n)$ is too close to the limit values of zero or one. 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 one. In this way, a minimum level of adaptation is always guaranteed.

In this paper we shall study the mean-square performance of the above combination procedure (1)–(5), as well as the behavior of the following modified version:

$$y_u(n) = \lambda_u(n)y_1(n) + [1 - \lambda_u(n)]y_2(n) \quad (6)$$

where

$$\lambda_u(n) = \begin{cases} 1, & a(n) \geq a^+ - \epsilon \\ \lambda(n), & a^+ - \epsilon > a(n) > -a^+ + \epsilon \\ 0, & a(n) \leq -a^+ + \epsilon \end{cases} \quad (7)$$

with ϵ being a small positive constant. In other words, the adaptation of $a(n)$ will continue to be given by (5) in terms of $\lambda(n)$, and we shall instead set $\lambda_u(n)$ to zero or one whenever $a(n)$ is close to the endpoints $\pm a^+$ rather than limit $\lambda_u(n)$ to the values $1 - \lambda^+$ or λ^+ as before.

The analysis that follows will show that $y_u(n)$ is universal with respect to its components, i.e., it performs, in steady-state, at least as well as the best component filter and, in certain situations, better than any of them. Similarly, $y(n)$ will be seen to be *nearly* universal, in the sense that it can perform as close as desired to the best component filter.

II. STATIONARY DATA MODEL AND NOTATION

In the sequel we adopt the following assumptions.

a) $d(n)$ and $\mathbf{u}(n)$ are related via a linear regression model

$$d(n) = \mathbf{w}_0^T \mathbf{u}(n) + e_0(n)$$

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

b) 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.

c) $E\{\mathbf{u}(n)\} = \mathbf{0}$, $E\{\mathbf{u}(n)\mathbf{u}^T(n)\} = \mathbf{R}$, $E\{d(n)\} = 0$, and $E\{e_0(n)\} = 0.$[1]

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

1) We define the weight error vector of a transversal filter as the difference between the optimal solution and the filter weights. Thus we define

$$\boldsymbol{\varepsilon}_i(n) = \mathbf{w}_0 - \mathbf{w}_i(n), \quad i = 1, 2$$

for the component filters and $\boldsymbol{\varepsilon}(n) = \mathbf{w}_0 - \mathbf{w}(n)$ for their combination.

2) *A priori* errors

$$e_{a,i}(n) = \boldsymbol{\varepsilon}_i^T(n)\mathbf{u}(n), \quad i = 1, 2 \quad \text{and} \quad e_a(n) = \boldsymbol{\varepsilon}^T(n)\mathbf{u}(n).$$

3) *A posteriori* errors

$$
\begin{aligned}
e_{p,i}(n) &= \boldsymbol{\varepsilon}_i^T(n+1)\mathbf{u}(n), \quad i = 1, 2 \\
e_p(n) &= \boldsymbol{\varepsilon}^T(n+1)\mathbf{u}(n).
\end{aligned}
$$

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

---

[1]Note that, unlike many other analyses in the adaptive filtering literature, we are not assuming that $\{\mathbf{u}(n)\}$ is a sequence of independent regressors.

of length M, namely, σ_0^2. When analyzing steady-state operation, we are mainly interested in the limiting value as n goes to ∞. Using

$$e(n) = d(n) - y(n)$$

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

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

$$J_{ex}(\infty) = \lim_{n\to\infty} E\{e_a^2(n)\} \quad \text{(combination).} \tag{9}$$

During the analysis, it will become 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.,

$$J_{ex,12}(\infty) = \lim_{n\to\infty} E\{e_{a,1}(n)e_{a,2}(n)\}. \tag{10}$$

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 the magnitude of $J_{ex,12}(\infty)$ can never 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 $J_{ex,2}(\infty)$.

III. UNIVERSALITY OF THE COMBINATION FILTER

In this section we analyze the steady-state performance of the two combination schemes (1) and (6). First, we will express $J_{ex}(\infty)$ and $J_{ex,u}(\infty)$ as a function of the EMSEs of the component filters, their cross-EMSE, and the steady-state value of the mixing parameter. Here, $J_{ex,u}(\infty)$ refers to the EMSE of the combination filter that results from using (6).

A. Steady-State EMSE of the Combination

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

$$e(n) = \lambda(n)e_1(n) + [1 - \lambda(n)]e_2(n)$$
$$= \lambda(n)e_{a,1}(n) + [1 - \lambda(n)]e_{a,2}(n) + e_0(n) \tag{11}$$

where $e_i(n) = d(n) - y_i(n)$. Likewise, the *a priori* error of the overall scheme can be expressed as a convex combination of the *a priori* errors of the component filters as

$$e_a(n) = \lambda(n)e_{a,1}(n) + [1 - \lambda(n)]e_{a,2}(n). \tag{12}$$

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

$$J_{ex}(\infty) = \lim_{n\to\infty} E\left\{\lambda^2(n)e_{a,1}^2(n) + [1 - \lambda(n)]^2 e_{a,2}^2(n) \right.$$
$$\left. + 2\lambda(n)[1 - \lambda(n)]e_{a,1}(n)e_{a,2}(n)\right\}. \tag{13}$$

[2]Note that for $\lambda(n)$ a constant, (13) simplifies to $J_{ex}(\infty) = \lambda^2 J_{ex,1}(\infty) + (1-\lambda)^2 J_{ex,2}(\infty) + 2\lambda(1-\lambda)J_{ex,12}(\infty)$.

Analogously, for the modified combination (6), we obtain a similar expression, but in terms of $\lambda_u(n)$

$$J_{ex,u}(\infty) = \lim_{n\to\infty} E\left\{\lambda_u^2(n)e_{a,1}^2(n) + [1 - \lambda_u(n)]^2 e_{a,2}^2(n) \right.$$
$$\left. + 2\lambda_u(n)[1 - \lambda_u(n)]e_{a,1}(n)e_{a,2}(n)\right\}. \tag{14}$$

The appearance of cross-expectation terms between the mixing parameter $\lambda(n)$ and the *a priori* errors of the component filters, together with the fact that (5) governing the adaptation of $a(n)$ is nonlinear in $\lambda(n)$, make it difficult to evaluate (13) and (14) exactly. However, there exist two cases in which the evaluation of $J_{ex}(\infty)$ and $J_{ex,u}(\infty)$ can be simplified.

- If $\lim_{n\to\infty} E\{a(n)\} = a^+$, and since $a(n) \in [-a^+, a^+]$, then necessarily $a(n) \to a^+$ as $n \to \infty$ almost surely. Consequently, $\lambda(n) \to \lambda^+$ almost surely and $\lambda_u(n) \to 1$ almost surely, allowing the simplifications

$$J_{ex}(\infty) = \lambda^{+2} J_{ex,1}(\infty) + (1 - \lambda^+)^2 J_{ex,2}(\infty)$$
$$+ 2\lambda^+(1 - \lambda^+)J_{ex,12}(\infty)$$
$$\approx J_{ex,1}(\infty) \tag{15}$$
$$J_{ex,u}(\infty) = J_{ex,1}(\infty) \tag{16}$$

where in simplifying (15) we used the fact that λ^+ is close to one. In fact, this approximation can be as accurate as desired by increasing the value of a^+.

- Similarly, when $\lim_{n\to\infty} E\{a(n)\} = -a^+$, we conclude that

$$J_{ex}(\infty) = (1 - \lambda^+)^2 J_{ex,1}(\infty) + \lambda^{+2} J_{ex,2}(\infty)$$
$$+ 2(1 - \lambda^+)\lambda^+ J_{ex,12}(\infty)$$
$$\approx J_{ex,2}(\infty) \tag{17}$$
$$J_{ex,u}(\infty) = J_{ex,2}(\infty). \tag{18}$$

Consequently, to obtain a better understanding about the performance of the combination filters, it is necessary to examine the steady-state value of $E\{a(n)\}$ and under what conditions $E\{a(n)\} \to \pm a^+$.

B. Steady-State Behavior of the Mixing Parameter

The learning rule for the mixing parameter $a(n)$ is given by (5). Taking expectations of both sides of this expression, we arrive at

$$E\{a(n+1)\} \approx [E\{a(n)\} + \mu_a E\{e(n)[y_1(n)$$
$$- y_2(n)]\lambda(n)[1 - \lambda(n)]\}]_{-a^+}^{a^+} \tag{19}$$

where the square brackets denote truncation to the indicated values, as explained after (5). 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 of 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 $\pm a^+$, the smaller the magnitude of the update factor $\lambda(n)[1 - \lambda(n)]$.

Now, introducing (11) into (19), 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

$$
\begin{aligned}
&E\{a(n+1)\} \\
&\approx \big[E\{a(n)\} + \mu_a E\{e_0(n)[e_{a,2}(n) - e_{a,1}(n)]\lambda(n)[1-\lambda(n)]\} \\
&\quad + \mu_a E\big\{ \big[e_{a,2}^2(n) - e_{a,1}(n)e_{a,2}(n)\big] \lambda(n)[1-\lambda(n)]^2 \big\} \\
&\quad + \mu_a E\big\{ \big[e_{a,1}(n)e_{a,2}(n) - e_{a,1}^2(n)\big] \lambda^2(n)[1-\lambda(n)] \big\} \big]_{-a^+}^{a^+}.
\end{aligned}
\tag{20}
$$

In the first line of (20), 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 assumption.

Assumption: In steady state, $\lambda(n)$ is independent of the *a priori* errors $e_{a,i}(n)$ of both component filters.

This condition is reasonable when using a reduced adaptation speed for $a(n)$, and therefore it is better justified when $a(n)$ approaches $\pm a^+$ in steady state given the fact that in these situations, the factor $\lambda(n)[1-\lambda(n)]$ in (5) is close to zero.

Using this assumption, and taking the limit as $n \to \infty$, (20) becomes

$$
\begin{aligned}
E\{a(n+1)\} &\approx \big[E\{a(n)\} + \mu_a E\{\lambda(n)[1-\lambda(n)]^2\}\Delta J_2 \\
&\quad - \mu_a E\{\lambda^2(n)[1-\lambda(n)]\}\Delta J_1 \big]_{-a^+}^{a^+}
\end{aligned}
\tag{21}
$$

where we introduced the differences

$$
\Delta J_i = J_{\mathrm{ex},i}(\infty) - J_{\mathrm{ex},12}(\infty), \quad i = 1,2
\tag{22}
$$

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

Equation (21) shows that the limiting value of $E\{a(n)\}$ depends on the values of $\Delta J_i, i = 1,2$. From our remark at the end of Section II, following (10), we know that these two constants cannot be simultaneously less than zero, so we need to consider the following three situations.

Case 1) $J_{\mathrm{ex},1}(\infty) \leq J_{\mathrm{ex},12}(\infty) \leq J_{\mathrm{ex},2}(\infty)$. In this case, we have $\Delta J_1 \leq 0$ and $\Delta J_2 \geq 0$. Furthermore, it can be seen that, since $\lambda(n) \in [1-\lambda^+, \lambda^+]$, 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

$$
E\{a(n+1)\} \geq [E\{a(n)\} + C]_{-a^+}^{a^+}; \quad \text{as } n \to \infty
\tag{23}
$$

for $C = \lambda^+(1-\lambda^+)^2(\Delta J_2 - \Delta J_1)$ a positive constant. Therefore, it follows that the only valid stationary point for (21) is $a(\infty) = a^+$ almost surely; and, in the light of the results of the previous subsection, we obtain in this case that

$$
\begin{aligned}
J_{\mathrm{ex}}(\infty) &\approx J_{\mathrm{ex},1}(\infty) \\
J_{\mathrm{ex},u}(\infty) &= J_{\mathrm{ex},1}(\infty).
\end{aligned}
\tag{24}
$$

That is, both combination schemes perform like their best component filter in this case.

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

$$
E\{a(n+1)\} \leq [E\{a(n)\} - C]_{-a^+}^{a^+}; \quad \text{as } n \to \infty
\tag{25}
$$

for a positive constant $C = \lambda^+(1-\lambda^+)^2(\Delta J_1 - \Delta J_2)$. Applying parallel arguments to those in the previous case, we conclude that $a(\infty) = -a^+$ almost surely, and

$$
\begin{aligned}
J_{\mathrm{ex}}(\infty) &\approx J_{\mathrm{ex},2}(\infty) \\
J_{\mathrm{ex},u}(\infty) &= J_{\mathrm{ex},2}(\infty).
\end{aligned}
\tag{26}
$$

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

Case 3) $J_{\mathrm{ex},12}(\infty) < J_{\mathrm{ex},i}(\infty), i = 1,2$. When the cross-EMSE is lower than the EMSEs of both individual filters, we have $\Delta J_i > 0, i = 1,2$, and a stationary point of (21) may be approximately characterized by the condition

$$
\begin{aligned}
&E\{\lambda(n)[1-\lambda(n)]^2\}\Delta J_2 \\
&= E\{\lambda^2(n)[1-\lambda(n)]\}\Delta J_1; \quad n \to \infty.
\end{aligned}
\tag{27}
$$

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 in this case 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, we obtain from (27)

$$
[1 - \bar\lambda(\infty)]\Delta J_2 = \bar\lambda(\infty)\Delta J_1
\tag{28}
$$

from which

$$
\bar\lambda(\infty) = \left[\frac{\Delta J_2}{\Delta J_1 + \Delta J_2} \right]_{1-\lambda^+}^{\lambda^+}.
\tag{29}
$$

It follows that

$$
\begin{aligned}
\lambda^+ &\geq \bar\lambda(\infty) > 0.5; \quad \text{if } J_{\mathrm{ex},1}(\infty) < J_{\mathrm{ex},2}(\infty) \\
0.5 &> \bar\lambda(\infty) \geq 1 - \lambda^+; \quad \text{if } J_{\mathrm{ex},1}(\infty) > J_{\mathrm{ex},2}(\infty).
\end{aligned}
\tag{30}
$$

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 (30), the overall filter will most likely not converge toward the best element in the mixture. This behavior does not imply that the combination of filters is suboptimal in this third case. In fact, it outperforms both component filters. Indeed, it can be verified that the value for $\bar\lambda(\infty)$ in (29) (neglecting the truncation) is the one that minimizes (13) and (14) under the assumption of zero variance for $\lambda(\infty)$ since then

$$
\begin{aligned}
J_{\mathrm{ex}}(\infty) &= \bar\lambda^2(\infty)J_{\mathrm{ex},1}(\infty) + [1 - \bar\lambda(\infty)]^2 J_{\mathrm{ex},2}(\infty) \\
&\quad + 2\bar\lambda(\infty)[1 - \bar\lambda(\infty)]J_{\mathrm{ex},12}(\infty)
\end{aligned}
\tag{31}
$$

TABLE I
EMSEs OF THE UNIVERSAL AND NEARLY UNIVERSAL COMBINATIONS OF TWO FILTERS AS A FUNCTION OF THE STEADY-STATE EXPECTATION OF $a(n)$

	$\bar{a}(\infty) = a^+$	$a^+ > \bar{a}(\infty) \geq a^+ - \epsilon$	$a^+ - \epsilon > \bar{a}(\infty) > -a^+ + \epsilon$	$-a^+ + \epsilon \geq \bar{a}(\infty) > -a^+$	$\bar{a}(\infty) = -a^+$
$J_{ex}(\infty)$	$\approx J_{ex,1}(\infty)$		$< \min\{J_{ex,1}(\infty), J_{ex,2}(\infty)\}$		$\approx J_{ex,2}(\infty)$
$J_{ex,u}(\infty)$	$= J_{ex,1}(\infty)$		$< \min\{J_{ex,1}(\infty), J_{ex,2}(\infty)\}$		$= J_{ex,2}(\infty)$

$$J_{\text{ex},u}(\infty) = \bar{\lambda}^2(\infty)J_{\text{ex},1}(\infty) + [1 - \bar{\lambda}(\infty)]^2 J_{\text{ex},2}(\infty)$$
$$+ 2\bar{\lambda}(\infty)[1 - \bar{\lambda}(\infty)]J_{\text{ex},12}(\infty) \quad (32)$$

given that, for intermediate values of $\bar{\lambda}(\infty)$, and assuming zero variance for $\lambda(\infty)$, we have $\bar{\lambda}_u(\infty) = \bar{\lambda}(\infty)$. Introducing (29) without the truncation into these expressions, we get after some algebra (see Appendix I)

$$J_{\text{ex}}(\infty) = J_{\text{ex},u}(\infty) = J_{\text{ex},12}(\infty) + \frac{\Delta J_1 \Delta J_2}{\Delta J_1 + \Delta J_2} \quad (33)$$

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

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

i.e.,

$$J_{\text{ex}}(\infty) < \min\{J_{\text{ex},1}(\infty), J_{\text{ex},2}(\infty)\}$$
$$J_{\text{ex},u}(\infty) < \min\{J_{\text{ex},1}(\infty), J_{\text{ex},2}(\infty)\}. \quad (34)$$

In summary, the above three cases allow us to conclude that the combination filter defined by (1)–(5) is nearly universal, i.e., its steady-state performance is approximately at least as good as that of the best component filter (with the approximation being as accurate as desired for increasing a^+). Regarding the combination (6), the discussion shows its universality in the sense that it performs at least as well as the best element in the mixture. Furthermore, in some circumstances, both schemes are able to outperform their components. In Table I we summarize these results for the different values of $\bar{a}(\infty) = E\{a(\infty)\}$.

The conclusions so far have an intuitive interpretation. When the magnitude of the cross-EMSE lies between the EMSEs of the individual filters, the cross-correlation between their *a priori* errors is high enough so that the combination of the filters can perform at most as the best individual filter. However, when $J_{\text{ex},12}(\infty) < J_{\text{ex},i}(\infty), i = 1, 2$, further knowledge can be gained by using an appropriate weighted combination of the filter outputs, thus leading to a lower excess error for the combination.

The above analysis does 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 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.

IV. COMBINATION OF LMS FILTERS WITH DIFFERENT STEP-SIZES

In this section we study the stationary performance of an adaptive convex combination of two LMS filters (CLMS) that only differ in their step-sizes. Consequently, the adaptation rule (2) for each component is given by

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \mu_i e_i(n)\mathbf{u}(n). \quad (35)$$

Without loss of generality, we will assume that $\mu_1 = r\mu_2, r > 1$, so that the first filter adapts faster.

A. Stationary Performance of CLMS

Using the energy conservation approach of ([19, ch. 6]), and assuming that at steady-state operation $\|\mathbf{u}(n)\|^2$ is independent of the *a priori* errors of the LMS filters (in the following we shall refer to this assumption as the separation principle), it is known that the EMSEs of the LMS components, in a stationary environment, are given by [19, (6.5.11)]

$$J_{\text{ex},i}(\infty) = \frac{\mu_i \sigma_0^2 \text{Tr}(\mathbf{R})}{2 - \mu_i \text{Tr}(\mathbf{R})}; \quad \text{assuming } \mu_i < \frac{2}{\text{Tr}(\mathbf{R})}. \quad (36)$$

In passing, we note that the EMSE is an increasing function of μ_i over the indicated interval, since

$$\frac{\partial J_{\text{ex},i}(\infty)}{\partial \mu_i} = \frac{2\sigma_0^2 \text{Tr}(\mathbf{R})}{[2 - \mu_i \text{Tr}(\mathbf{R})]^2} > 0. \quad (37)$$

To derive an expression for the cross-EMSE of the two filters, we proceed from the energy conservation relation [19, (6.3.7)], which relates the weight and *a priori* and *a posteriori* errors of a general class of transversal adaptive schemes and which applies to the LMS updates, namely

$$\boldsymbol{\varepsilon}_i(n+1) + \frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2}e_{a,i}(n) = \boldsymbol{\varepsilon}_i(n) + \frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2}e_{p,i}(n). \quad (38)$$

Multiplying the transpose of (38) by (38) itself for $i = 1$ and 2, respectively, and cancelling terms, we arrive at the following generalized energy conservation relation:

$$\boldsymbol{\varepsilon}_1^T(n+1)\boldsymbol{\varepsilon}_2(n+1) + \frac{e_{a,1}(n)e_{a,2}(n)}{\|\mathbf{u}(n)\|^2}$$
$$= \boldsymbol{\varepsilon}_1^T(n)\boldsymbol{\varepsilon}_2(n) + \frac{e_{p,1}(n)e_{p,2}(n)}{\|\mathbf{u}(n)\|^2}. \quad (39)$$

Taking expectations of both sides, and using the fact that in steady state

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

we get

$$E\left\{\frac{e_{a,1}(n)e_{a,2}(n)}{\|\mathbf{u}(n)\|^2}\right\} = E\left\{\frac{e_{p,1}(n)e_{p,2}(n)}{\|\mathbf{u}(n)\|^2}\right\}; \quad n \to \infty. \tag{40}$$

The above expression holds for any filters $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ satisfying (38). When these weights arise from LMS updates, it can be shown that their *a priori* and *a posteriori* errors are related via [19, (6.3.3)]

$$e_{p,i}(n) = e_{a,i}(n) - \mu_i\|\mathbf{u}(n)\|^2 e_i(n), \quad i = 1, 2. \tag{41}$$

Introducing (41) into (40) and multiplying terms, we get

$$\mu_1\mu_2 E\{\|\mathbf{u}(n)\|^2 e_1(n)e_2(n)\}$$
$$= \mu_1 E\{e_{a,2}(n)e_1(n)\} + \mu_2 E\{e_{a,1}(n)e_2(n)\}; \quad n \to \infty. \tag{42}$$

Now, substituting $e_i(n) = e_{a,i}(n) + e_0(n)$, for $i = 1, 2$, gives

$$\mu_1\mu_2 E\left\{\|\mathbf{u}(n)\|^2 \left(e_{a,1}(n)e_{a,2}(n)\right.\right.$$
$$\left.\left. + e_0(n)[e_{a,1}(n) + e_{a,2}(n)] + e_0^2(n)\right)\right\}$$
$$= (\mu_1 + \mu_2)J_{\text{ex},12}(\infty) + E\{[\mu_1 e_{a,2}(n)$$
$$+ \mu_2 e_{a,1}(n)]e_0(n)\}; \quad n \to \infty. \tag{43}$$

Several of the above expectations vanish as a consequence of $e_0(n)$ being a zero-mean process and independent of the *a priori* errors of the filters at iteration n, so we can write

$$J_{\text{ex},12}(\infty) = \frac{\mu_{12}}{2}\left[E\{\|\mathbf{u}(n)\|^2 e_{a,1}(n)e_{a,2}(n)\} + \sigma_0^2 \text{Tr}(\mathbf{R})\right];$$
$$n \to \infty \tag{44}$$

where we have defined

$$\mu_{12} = \frac{2\mu_1\mu_2}{\mu_1 + \mu_2} \tag{45}$$

that can be easily shown to lie between μ_1 and μ_2.

Finally, applying the separation principle to (44) (i.e., that $\|\mathbf{u}(n)\|^2$ is independent of the *a priori* errors in steady state), the cross-EMSE of the filters is then given by

$$J_{\text{ex},12}(\infty) = \frac{\mu_{12}\sigma_0^2 \text{Tr}(\mathbf{R})}{2 - \mu_{12}\text{Tr}(\mathbf{R})}. \tag{46}$$

The similarity that exists between (46) and (36), together with the fact that $\mu_1 > \mu_{12} > \mu_2$, allows us to conclude that $J_{\text{ex},1}(\infty) > J_{\text{ex},12}(\infty) > J_{\text{ex},2}(\infty)$, so we are in the second of the situations considered in Section III-B. Consequently, (26) applies and both the universal and nearly universal combination schemes achieve the lower EMSE of the μ_2-LMS filter, although they can never improve on this performance.

Fig. 2 plots (46) for different selections of μ_1 and $\mu_2 = \mu_1/r$. The optimal solution in this example was formed with independent random values between -1 and 1, and is given by $\mathbf{w}_0^T = [0.9003, -0.5377, 0.2137, -0.028, 0.7826, 0.5242, -0.0871](M = 7)$. 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 $u(n)$ was
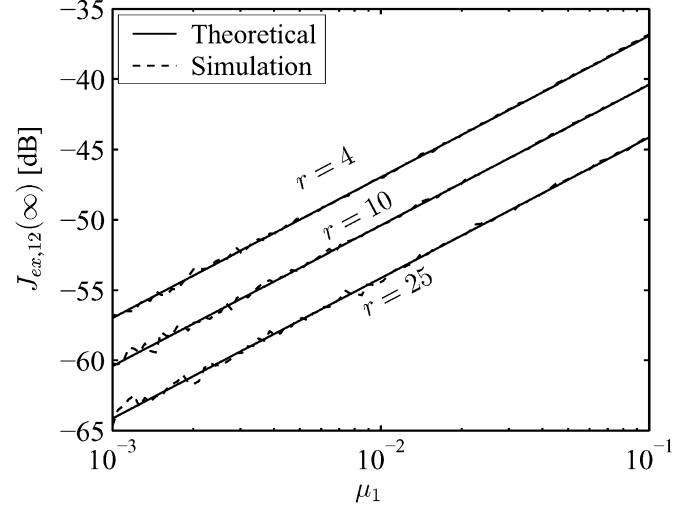

Fig. 2. Steady-state theoretical and estimated cross-EMSE of two LMS filters for different values of μ_1 and $\mu_2 = \mu_1/r$.
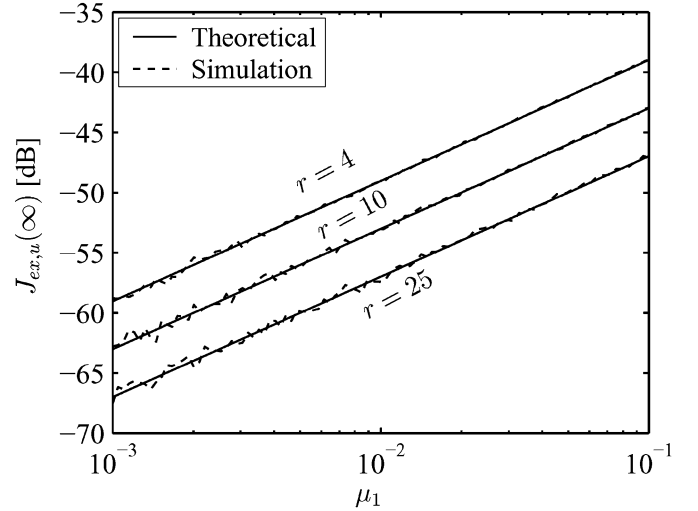

Fig. 3. Stationary theoretical and estimated EMSE of the universal combination $y_u(n)$.

obtained from a first-order autoregressive model with transfer function $\sqrt{1-a^2}/(1-az^{-1}), a = 0.8$, fed with an i.i.d. Gaussian random process whose variance was selected to get $\text{Tr}(\mathbf{R}) = 1$. Finally, additive i.i.d. noise $e_0(n)$ with variance $\sigma_0^2 = 0.01$, and independent of $u(n)$, was added to form the desired signal. Each estimated cross-EMSE was calculated by averaging $e_{a,1}(n)e_{a,2}(n)$ for 20 000 iterations after the convergence of both LMS filters, and over 50 independent runs. The figure indicates that (46) is appropriate to model the cross-correlation between the *a priori* errors of two independent LMS filters for a wide range of step-sizes. Fig. 3 represents the theoretical and estimated stationary EMSE of the universal combination $y_u(n)$ with $\mu_a = 100$. Again, there exists a close match between the results of our analysis and the real values of $J_{\text{ex},u}(\infty)$ for different selections of μ_1 and μ_2.
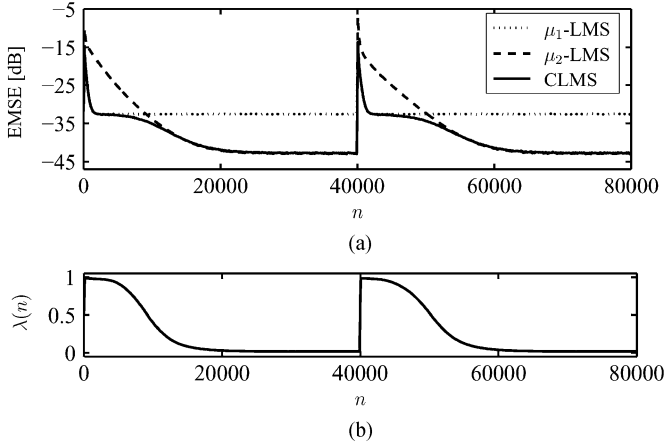
Fig. 4. (a) EMSEs of two LMS filters with step-sizes $\mu_1 = 0.1$ and $\mu_2 = 0.01$, as well as that of their adaptive combination (CLMS) with $\mu_a = 100$ and $a^+ = 4$. (b) Evolution of the mixing parameter.

Finally, a typical performance of the combination scheme (1) is depicted in Fig. 4(a) for a case with $\mu_1 = 0.1$, $\mu_2 = 0.01$, $\mu_a = 100$, and $a^+ = 4$. At $n = 40000$, $\mathbf{w}_0$ is changed into $\mathbf{w}_0^T = [0.2542, -0.4696, -0.3753, 0.0454, -0.1827, 0.7859, 0.1475]$ to study the ability of the combined scheme to react to changes in the filtering scenario. The curves in this figure have been averaged over 10 000 independent runs. As predicted by our analysis, the combined filter attains the lower stationary EMSE of the μ_2-LMS filter. Furthermore, we see that it also presents the faster convergence of the fast LMS component. Fig. 4(b) represents the time evolution of the mixing coefficient. We can see that it rapidly changes toward $\lambda(n) = \lambda^+$ during the initial convergence and after $n = 40\,000$, while its steady-state value is $\lambda(n) = 1 - \lambda^+$ in stationary intervals.

B. Tracking Performance of CLMS

We now examine the tracking performance of the CLMS filter in a nonstationary environment. For this purpose, we first modify the data model of Section II and assume that the weight vector $\mathbf{w}_0$ varies according to the random-walk model

$$\mathbf{w}_0(n+1) = \mathbf{w}_0(n) + \mathbf{q}(n)$$

where $\mathbf{q}(n)$ is an i.i.d. random zero-mean vector with covariance matrix $\mathbf{Q} = E\{\mathbf{q}(n)\mathbf{q}^T(n)\}$, and independent of $\{\mathbf{u}(m), d(m), e_0(m)\}$, for all $m \leq n$, and of the initial conditions $\mathbf{w}_0(0), \mathbf{w}_1(0), \mathbf{w}_2(0),$ and $a(0)$. From its definition, $\mathrm{Tr}(\mathbf{Q})$ can be seen as a measure of the speed of changes in $\mathbf{w}_0$.

When using this modified model we can keep the definitions given in Section II for the weight and *a priori* errors of the filters

$$\varepsilon_i(n) = \mathbf{w}_0(n) - \mathbf{w}_i(n), \quad e_{a,i}(n) = \varepsilon_i^T(n)\mathbf{u}(n), \quad i = 1, 2$$

and redefine the *a posteriori* errors as

$$e_{p,i}(n) = [\mathbf{w}_0(n) - \mathbf{w}_i(n+1)]^T \mathbf{u}(n), \quad i = 1, 2.$$

Recall that the argument for the universality of the combination scheme relied solely on the *a priori* errors of the filters. Consequently, and since the definitions for these errors have not changed, the conclusions about the universality of the mixtures

obtained in Section III still apply to the nonstationary scenario, as well as the three cases studied in Section III-B. We now proceed to evaluate more explicitly the EMSE and cross-EMSE expressions in the nonstationary case.

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 $\mathrm{Tr}(\mathbf{Q})$ and for given μ_1 and μ_2.

Again, energy conservation arguments can be used to show that, in the tracking scenario described above, the EMSE of an LMS filter with step-size μ_i is given by [19, (7.5.9)]

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

$$\text{assuming } \mu_i < \frac{2}{\mathrm{Tr}(\mathbf{R})}. \qquad (47)$$

As 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. We show in Appendix II that (47) achieves a minimum over the interval $0 < \mu_i < 2/\mathrm{Tr}(\mathbf{R})$ at the following optimal step-size:

$$\mu_{\mathrm{opt}} = \sqrt{\frac{\mathrm{Tr}(\mathbf{Q})}{\sigma_0^2 \mathrm{Tr}(\mathbf{R})} + \frac{[\mathrm{Tr}(\mathbf{Q})]^2}{4\sigma_0^4}} - \frac{\mathrm{Tr}(\mathbf{Q})}{2\sigma_0^2}. \qquad (48)$$

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}. Thus, for the components of the CLMS filter, we set

$$\mathrm{NSD}_i(\infty) = J_{\mathrm{ex},i}(\infty)/J_{\mathrm{ex,opt}}(\infty), \quad i = 1, 2. \qquad (49)$$

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

$$\mathrm{NSD}(\infty) = J_{\mathrm{ex}}(\infty)/J_{\mathrm{ex,opt}}(\infty) \qquad (50)$$

$$\mathrm{NSD}_u(\infty) = J_{\mathrm{ex},u}(\infty)/J_{\mathrm{ex,opt}}(\infty) \qquad (51)$$

$$\mathrm{NSD}_{12}(\infty) = J_{\mathrm{ex},12}(\infty)/J_{\mathrm{ex,opt}}(\infty). \qquad (52)$$

Next, we need to obtain an expression for $J_{\mathrm{ex},12}(\infty)$ in the nonstationary case. Since the derivation is similar to what we carried out in the stationary case, we shall be brief and point out only the main results. Our starting point in this case is the relation [19, ch. 7]

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

$$= \varepsilon_i(n) + \frac{\mathbf{u}(n)}{\|\mathbf{u}(n)\|^2} e_{p,i}(n), \quad i = 1, 2 \quad (53)$$

TABLE II
CLMS NORMALIZED SQUARE DEVIATION AS A FUNCTION OF $\mathrm{Tr}(\mathbf{Q})$

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

which holds for a general class of transversal filters. Multiplying the transpose of (53) by (53) itself, for $i = 1$ and $i = 2$, respectively, and after cancelling terms, we get

$$[\mathbf{w}_0(n) - \mathbf{w}_1(n+1)]^T[\mathbf{w}_0(n) - \mathbf{w}_2(n+1)]$$
$$+ \frac{e_{a,1}(n)e_{a,2}(n)}{\|\mathbf{u}(n)\|^2} = \boldsymbol{\varepsilon}_1^T(n)\boldsymbol{\varepsilon}_2(n) + \frac{e_{p,1}(n)e_{p,2}(n)}{\|\mathbf{u}(n)\|^2}. \quad (54)$$

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

$$E\{[\mathbf{w}_0(n) - \mathbf{w}_1(n+1)]^T[\mathbf{w}_0(n) - \mathbf{w}_2(n+1)]\}$$
$$= E\{\boldsymbol{\varepsilon}_1^T(n+1)\boldsymbol{\varepsilon}_2(n+1)\} + \mathrm{Tr}(\mathbf{Q})$$
$$- E\{[\boldsymbol{\varepsilon}_1(n+1) + \boldsymbol{\varepsilon}_2(n+1)]^T\mathbf{q}(n)\}. \quad (55)$$

Now, note that

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

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 (55) to

$$E\{[\mathbf{w}_0(n) - \mathbf{w}_1(n+1)]^T[\mathbf{w}_0(n) - \mathbf{w}_2(n+1)]\}$$
$$= E\{\boldsymbol{\varepsilon}_1^T(n+1)\boldsymbol{\varepsilon}_2(n+1)\} - \mathrm{Tr}(\mathbf{Q}). \quad (56)$$

Using this result, and the fact that in steady state $E\{\boldsymbol{\varepsilon}_1^T(n+1)\boldsymbol{\varepsilon}_2(n+1)\} = E\{\boldsymbol{\varepsilon}_1^T(n)\boldsymbol{\varepsilon}_2(n)\}$, we get from (54) that

$$E\left\{\frac{e_{a,1}(n)e_{a,2}(n)}{\|\mathbf{u}(n)\|^2}\right\} = E\left\{\frac{e_{p,1}(n)e_{p,2}(n)}{\|\mathbf{u}(n)\|^2}\right\} + \mathrm{Tr}(\mathbf{Q});$$
$$n \to \infty \quad (57)$$

which is the extension of (40) to the nonstationary case.

To proceed further, we need to specialize this result for $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$. Using (41) and $e_i(n) = e_{a,i}(n) + e_0(n)$, we arrive after some manipulations at

$$(\mu_1 + \mu_2)J_{\mathrm{ex},12}(\infty) = \mathrm{Tr}(\mathbf{Q})$$
$$+ \mu_1\mu_2\left[E\{\|\mathbf{u}(n)\|^2 e_{a,1}(n)e_{a,2}(n)\} + \sigma_0^2\mathrm{Tr}(\mathbf{R})\right]. \quad (58)$$

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

$$J_{\mathrm{ex},12}(\infty) = \frac{\mu_{12}\sigma_0^2\mathrm{Tr}(\mathbf{R}) + 2\mathrm{Tr}(\mathbf{Q})/(\mu_1 + \mu_2)}{2 - \mu_{12}\mathrm{Tr}(\mathbf{R})}. \quad (59)$$

Again, as discussed in Section III-B, 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 $J_{\mathrm{ex},1}(\infty), J_{\mathrm{ex},2}(\infty)$, and $J_{\mathrm{ex},12}(\infty)$ for any value of $\mathrm{Tr}(\mathbf{Q})$ in the nonstationary case. In order to do this, let us consider first a small step-size approximation for (47), (48), and (59), i.e.,

$$J_{\mathrm{ex},i}(\infty) \approx \frac{\mu_i\sigma_0^2\mathrm{Tr}(\mathbf{R}) + \mu_i^{-1}\mathrm{Tr}(\mathbf{Q})}{2}, \quad i = 1, 2 \quad (60)$$

$$\mu_{\mathrm{opt}} \approx \sqrt{\frac{\mathrm{Tr}(\mathbf{Q})}{\sigma_0^2\mathrm{Tr}(\mathbf{R})}} \quad (61)$$

$$J_{\mathrm{ex},12}(\infty) \approx \frac{\mu_{12}\sigma_0^2\mathrm{Tr}(\mathbf{R}) + 2\mathrm{Tr}(\mathbf{Q})/(\mu_1 + \mu_2)}{2}. \quad (62)$$

Then, subtracting (62) from (60), and recalling that $r = \mu_1/\mu_2$, it is straightforward to verify that

$$\Delta J_1 = \frac{r-1}{r+1} \cdot \frac{\mu_1\sigma_0^2\mathrm{Tr}(\mathbf{R}) - \mu_1^{-1}\mathrm{Tr}(\mathbf{Q})}{2}$$
$$\times \begin{cases} < 0, & \text{for } \mathrm{Tr}(\mathbf{Q}) > q_1 \\ > 0, & \text{for } \mathrm{Tr}(\mathbf{Q}) < q_1 \end{cases} \quad (63)$$

$$\Delta J_2 = \frac{1-r}{r+1} \cdot \frac{\mu_2\sigma_0^2\mathrm{Tr}(\mathbf{R}) - \mu_2^{-1}\mathrm{Tr}(\mathbf{Q})}{2}$$
$$\times \begin{cases} > 0, & \text{for } \mathrm{Tr}(\mathbf{Q}) > q_2 \\ < 0, & \text{for } \mathrm{Tr}(\mathbf{Q}) < q_2 \end{cases} \quad (64)$$

where we have defined[3]

$$q_i = \mu_i^2\sigma_0^2\mathrm{Tr}(\mathbf{R}). \quad (65)$$

Thus, we find that, depending on the value of $\mathrm{Tr}(\mathbf{Q})$, all three cases described in Section III-B can occur—see also Table II.

- If $\mathrm{Tr}(\mathbf{Q}) > q_1$, we have that $J_{\mathrm{ex},1}(\infty) \leq J_{\mathrm{ex},12}(\infty) \leq J_{\mathrm{ex},2}(\infty)$ (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 $J_{\mathrm{ex}}(\infty) \approx J_{\mathrm{ex},2}(\infty)$ and $J_{\mathrm{ex},u}(\infty) = J_{\mathrm{ex},2}(\infty)$.
- Finally, when $q_2 < \mathrm{Tr}(\mathbf{Q}) < q_1$, we have that both ΔJ_1 and ΔJ_2 are greater than zero, i.e., (34) applies, and the combination outperforms both components.

Fig. 5(a) and (b) illustrates the above theoretical conclusions. We have depicted the theoretical steady-state 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 $\sigma_0^2 = 0.01$.

[3]Note that under the small step-size approximation, $q_1(q_2)$ is the value of $\mathrm{Tr}(\mathbf{Q})$ for which $\mu_1(\mu_2)$ is the optimal step-size [see (61)].
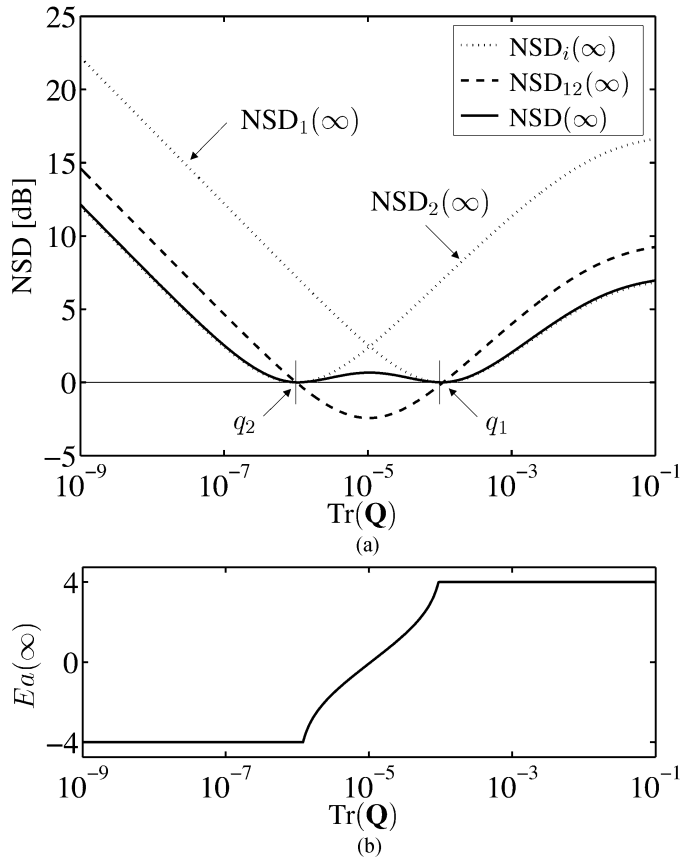
Fig. 5.   (a) Theoretical normalized mean-square deviation of two LMS filters with steps $\mu_1 = 0.1$ and $\mu_2 = 0.01$ as a function of the trace of the covariance matrix of $\mathbf{q}(n)$. Their cross-NSD is also depicted using a dashed line, as is the NSD of their adaptive combination. (b) Theoretical steady-state value for the mixing parameter.
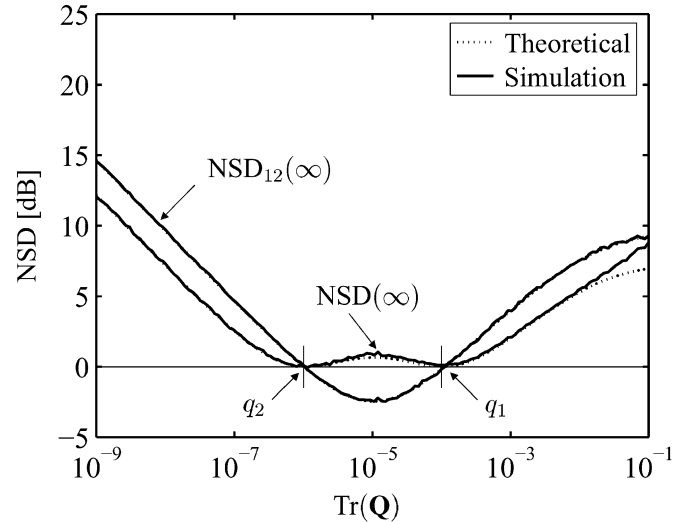

Fig. 6.   Steady-state theoretical and estimated cross-NSD of two LMS filters with $\mu_1 = 0.1$ and $\mu_2 = 0.01$, and NSD for the resulting CLMS combination.
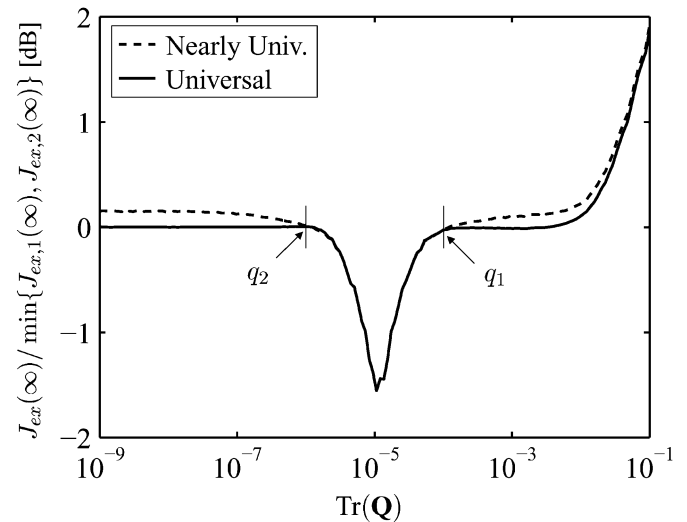

Fig. 7.   Steady-state EMSE of the component filters in relation to the excess mean-square error of the best LMS component.

We have only depicted the theoretical values for $\mathrm{NSD}(\infty)$; this curve and that for $\mathrm{NSD}_u(\infty)$ are indistinguishable for the level of detail in the figure. 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 Fig. 5(a) for $q_2 < \mathrm{Tr}(\mathbf{Q}) < q_1$.

To further illustrate our analysis, we have carried out simulations for the same example considered in the previous subsection, except that now random changes modify the optimal solution at each iteration. The entries of $\mathbf{q}(n)$ are taken as independent Gaussian values with equal variances. As for the settings for the CLMS filter, we have again 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. Fig. 6 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 20 000 samples after filter convergence, and over 50 independent runs.

To further examine the universal capabilities of the combined filters (1) and (6), it is necessary to study their behavior in relation to the performance of their two component LMS filters. In Fig. 7 we depict the ratio between $J_{\mathrm{ex}}(\infty)$ ($J_{\mathrm{ex},u}(\infty)$) for the strictly universal scheme) and the lowest EMSE of the

components for each value of $\mathrm{Tr}(\mathbf{Q})$, so that a level of 0 dB is associated to universality. As expected, both $\mathbf{w}(n)$ and $\mathbf{w}_u(n)$ obtain similar results, but the nearly universal combination suffers some degradation for $\mathrm{Tr}(\mathbf{Q}) < q_2$ and $\mathrm{Tr}(\mathbf{Q}) > q_1$ given that, unlike $\lambda_u(n)$, $\lambda(n)$ is not allowed to take exactly the values zero and one. For $q_2 < \mathrm{Tr}(\mathbf{Q}) < q_1$, both combination schemes show almost identical behaviors, achieving a lower EMSE than any of the components. Although universality is satisfied for at least seven orders of magnitude, this property is lost for $\mathrm{Tr}(\mathbf{Q}) \gtrsim 10^{-2}$. The reason is that, for these values of $\mathrm{Tr}(\mathbf{Q})$, the error incurred by the combination is very high and, since the adaptation of $a(n)$ is directly proportional to $e(n)$, this fact introduces a nonnegligible error in the subsequent analysis.

V. COMBINATION OF LMS FILTERS WITH IMPROVED TRACKING PERFORMANCE

The tracking analysis of the CLMS filter in the previous section suggests that the combination method can be exploited to

get from the mixture an improved tracking performance for any value of $\text{Tr}(\mathbf{Q})$. To see this, recall from Section IV-B (see also Fig. 5) that the limiting value of $a(\infty)$ indicates if the combined filter is operating in Case I, II, or III as discussed in Section III-B.

- When $\bar{a}(\infty) = a^+(\text{Tr}(\mathbf{Q}) > q_1)$, the filter operates in Case I, and so we have $J_{\text{ex}}(\infty) \approx J_{\text{ex},1}(\infty)$.
- If $\bar{a}(\infty) = -a^+(q_2 > \text{Tr}(\mathbf{Q}))$, the filter operates in Case II.
- For intermediate values of $\bar{a}(\infty)$, we have $q_2 < \text{Tr}(\mathbf{Q}) < q_1$, and Case III appears, allowing the combined filter to obtain a lower EMSE than any of its components.

Furthermore, note that in Case II we have $\mu_i > \mu_{\text{opt}}, i = 1, 2$, while in Case I we have $\mu_i < \mu_{\text{opt}}, i = 1, 2$. Consequently, it is possible to improve the tracking performance of the combination of filters by using the value of the mixing coefficient to manage the step-sizes of the component LMS filters.

1) If $\lambda(n) > \beta$, with β being a threshold close to one (i.e., when the combination is most likely in Case I), multiply the step-sizes of both filters by r, thus getting $\mu'_1 = r\mu_1$ and $\mu'_2 = \mu_1$, where the prime superscript is being used to denote the new values of the step-sizes.
2) If $\lambda(n) < 1 - \beta$, decrease the values of both step-sizes accordingly to $\mu'_1 = \mu_2$ and $\mu'_2 = \mu_2/r$.

Obviously, application of these procedures only makes sense if $\beta < \lambda^+$, so it is no longer necessary to recur to the truncation of the mixing parameter $a(n)$.

To show why application of these two simple rules makes it possible to get $J'_{\text{ex}}(\infty) \leq J_{\text{ex}}(\infty)$, let us assume first that $\text{Tr}(\mathbf{Q})$ is such that the filter is operating in Case I. Consequently, $a(n) \to a^+$, and the first rule will be applied. Then we have that the new step-sizes necessarily verify any of the two following conditions.

- $\mu'_i < \mu_{\text{opt}}, i = 1, 2$: in this case we still have that $\text{Tr}(\mathbf{Q}) > q'_1$ (Case I) so that

$$J'_{\text{ex},1}(\infty) < J'_{\text{ex},2}(\infty) = J_{\text{ex},1}(\infty) \qquad (66)$$

allowing us to conclude that

$$J'_{\text{ex}}(\infty) \approx J'_{\text{ex},1}(\infty) < J_{\text{ex},1}(\infty) \approx J_{\text{ex}}(\infty). \qquad (67)$$

- $\mu'_1 > \mu_{\text{opt}} > \mu'_2$: now, we can see that $q'_2 < \text{Tr}(\mathbf{Q}) < q'_1$ so that the new values of the step-sizes make the filter operate in Case III and

$$\begin{aligned} J'_{\text{ex}}(\infty) &< \min\{J'_{\text{ex},1}(\infty), J'_{\text{ex},2}(\infty)\} \\ &= \min\{J'_{\text{ex},1}(\infty), J_{\text{ex},1}(\infty)\} \\ &\leq J_{\text{ex},1}(\infty). \end{aligned} \qquad (68)$$

So, again we have $J'_{\text{ex}}(\infty) \leq J_{\text{ex}}(\infty)$. Parallel arguments can be used to show that dividing the step-sizes by r when $\lambda(n) < 1 - \beta$ serves also to decrease the EMSE of the combination.

In Table III we list the pseudocode for the proposed combination of variable step LMS algorithms (CVS-LMS) based on the

TABLE III
SUMMARY OF THE CVS-LMS ALGORITHM

1.-	Inputs:
	$\mu_1(0), \mu_2(0), \mu_a, \mu_{\max}, \beta, r$
2.-	Initialization:
	$\mathbf{w}_1(0) = \mathbf{w}_2(0) = \mathbf{0}; a(0) = 0; \lambda(0) = 0.5$
3.-	Loop $n = 0, 1, 2, \ldots$
	$y_i(n) = \mathbf{w}_i^T(n)\mathbf{u}(n), \; i = 1, 2$
	$e_i(n) = d(n) - y_i(n), \; i = 1, 2$
	$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \mu_i(n)e_i(n)\mathbf{u}(n), \; i = 1, 2$
	$y(n) = \lambda(n)y_1(n) + [1 - \lambda(n)]y_2(n)$
	$e(n) = d(n) - y(n)$
	$a(n+1) = a(n) + \mu_a(n)e(n)[y_1(n) - y_2(n)]\lambda(n)[1 - \lambda(n)]$
	$\lambda(n+1) = \text{sgm}[a(n+1)]$
3.1.-	If $\lambda(n+1) < 1 - \beta$
	$\mu_i(n+1) = \mu_i(n)/r, \; i = 1, 2$
	$\mathbf{w}_1(n+1) = \mathbf{w}_2(n+1)$
	$a(n+1) = 0; \lambda(n+1) = 0.5$
	Endif
3.2.-	If $(\lambda(n+1) > \beta$ and $r\mu_1(n) < \mu_{\max})$
	$\mu_i(n+1) = r\mu_i(n), \; i = 1, 2$
	$\mathbf{w}_2(n+1) = \mathbf{w}_1(n+1)$
	$a(n+1) = 0; \lambda(n+1) = 0.5$
	Endif
3.3.-	If $(\lambda(n+1) > \beta$ and $r\mu_1(n) > \mu_{\max})$
	$a(n+1) = a^+; \lambda(n+1) = \beta$
	Endif

above ideas. We have added an additional parameter $\mu_{\max}$ that serves to guarantee the stability of the filter when $\mathbf{w}_0$ changes very fast, leading to many successive increments of μ_1 and μ_2.

We have carried out experiments using the proposed CVS-LMS algorithm in the tracking scenario of Section IV-B. We set the filter parameters to $\mu_1(0) = 0.05, r = 4, \mu_a = 100$, and $\mu_{\max} = 0.2$, and the NSD for each value of $\text{Tr}(\mathbf{Q})$ has been calculated by averaging the error over $20\,000$ samples (after CVS-LMS convergence is complete) and over 50 independent runs. We can see in Fig. 8 how the new filter improves the performance of a conventional LMS with fixed step-size $\mu = 0.0125$, achieving an NSD of about 0 dB for several orders of magnitude in $\text{Tr}(\mathbf{Q})$. The NSD increment that can be observed for the highest values of $\text{Tr}(\mathbf{Q})$ is due to the upper limit imposed on the step-sizes ($\mu_{\max}$).

A. Comparison With Previous Variable Step-Size LMS Algorithms

We have carried out a number of experiments comparing the combination approach with several variable step-size LMS algorithms. We will focus here on the robust variable step-size (RVSS) LMS method of [3] and on the adaptive learning rate (ALR) LMS algorithm of [4], which are good representatives of these versions.

When using an adaptive step-size, the LMS update takes the form

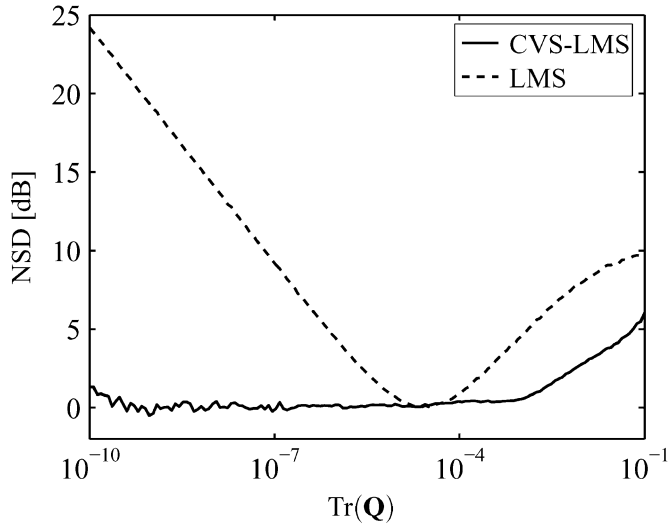$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)e(n)\mathbf{u}(n) \qquad (69)$$

Fig. 8. Tracking performance of a conventional LMS filter with fixed step-size $\mu = 0.0125$ and of a CVS-LMS scheme with parameters $\mu_1(0) = 0.05$, $r = 4$, $\mu_a = 100$, and $\mu_{\max} = 0.2$.

TABLE IV
SUMMARY OF THE RVSS-LMS ALGORITHM

1.- Inputs:
$\mu(0), \mu_{\min}, \mu_{\max}, \alpha_R, \beta_R, \gamma_R$

2.- Initialization:
$\mathbf{w}(0) = \mathbf{0}$
$p = 0$

3.- Loop $n = 0, 1, 2, \dots$
$e(n) = d(n) - \mathbf{w}^T(n)\mathbf{u}(n)$
$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)e(n)\mathbf{u}(n)$
$p = \beta_R\, p + (1 - \beta_R)e(n)e(n-1)$
$\mu(n+1) = \left[\alpha_R\mu(n) + \gamma_R\, p^2\right]_{\mu_{\min}}^{\mu_{\max}}$

where $\mu(n)$ is adjusted according to a certain rule. In [2] it was proposed to use the instantaneous value of the quadratic error incurred by the filter to control $\mu(n)$. RVSS-LMS [3] elaborates on the same idea and manages the step-size by means of a low-pass filtered estimation of the autocorrelation $E\{e(n)e(n-1)\}$. Using a different criterion, ALR-LMS [4] estimates the norm of the mean flow used to adapt the LMS filter (i.e., $E\{-e(n)\mathbf{u}(n)\}$). This value is a measure of the proximity of $\mathbf{w}(n)$ to the solution of the filtering problem, and, therefore, it can be used to manage the step size. Summaries of the RVSS-LMS and ALR-LMS algorithms are provided in Tables IV and V, respectively.

Fig. 9(a) shows the tracking performance of RVSS-LMS as a function of $\mathrm{Tr}(\mathbf{Q})$ for different values of the free parameters. Following the guidelines in [3], we used $\alpha_R = 0.97$ in all cases, exploring different values for β_R and γ_R. Regarding the minimum and maximum values for $\mu(n)$, we selected $\mu_{\min} = 0$ and $\mu_{\max} = 0.2$ to obtain a fair comparison with CVS-LMS. Again, all results have been obtained by averaging the error over 20 000 samples (after filter convergence) and 50 runs.

Although better than conventional LMS, RVSS-LMS is still subject to a compromise regarding its performance for different

TABLE V
SUMMARY OF THE ALR-LMS ALGORITHM

1.- Inputs:
$\mu(0), \mu_{\min}, \mu_{\max}, \alpha_A, \beta_A, \delta_A$

2.- Initialization:
$\mathbf{w}(0) = \mathbf{0}$
$\mathbf{f} = 0$

3.- Loop $n = 0, 1, 2, \dots$
$e(n) = d(n) - \mathbf{w}^T(n)\mathbf{u}(n)$
$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)e(n)\mathbf{u}(n)$
$\mathbf{f} = (1 - \delta_A)\mathbf{f} - \delta_A e(n)\mathbf{u}(n)$
$\mu(n+1) = \left[\mu(n) + \alpha_A\mu(n)\left[\beta_A\|\mathbf{f}\| - \mu(n)\right]\right]_{\mu_{\min}}^{\mu_{\max}}$
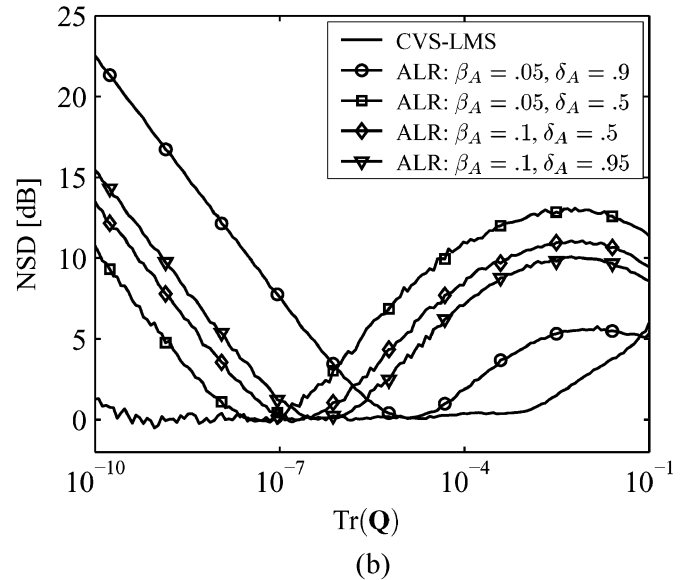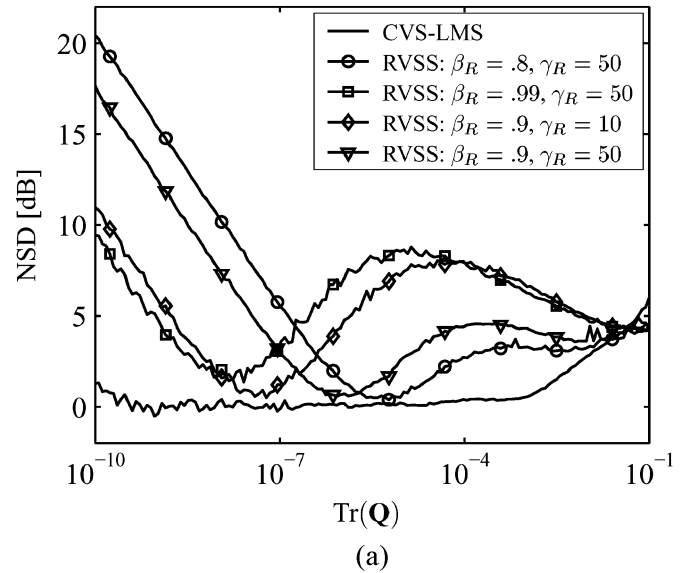

(a)

(b)

Fig. 9. Comparison between the tracking NSDs of the CVS-LMS filter and those of the RVSS algorithm of [3] and the ALR method of [4], for different values of the free parameters.

values of $\mathrm{Tr}(\mathbf{Q})$. Consequently, a straightforward manner to improve its tracking capability would be to use an adaptive combi-

nation of two filters with complementary settings (for instance, with a high and a low β_R). In this regard, it is useful to remark that, unlike RVSS-LMS, CVS-LMS is not subject to this performance compromise and its NSD remains close to zero for the displayed range of $\mathrm{Tr}(\mathbf{Q})$.

Similar conclusions can be extracted when comparing CVS-LMS and ALR-LMS [Fig. 9(b)]. In this case, the settings for ALR-LMS were $\alpha_A = 0.5, \mu_{\min} = 0$, and $\mu_{\max} = 0.2$, and we explored different values for β_A and δ_A.

VI. CONCLUSION

Combination approaches are an effective way to improve the performance of adaptive filters. In this paper we have analyzed the behavior of one such approach, showing its universality in the sense that it performs at least as well as the best of the components. Furthermore, when the component filters satisfy certain conditions, we have seen that their combination outperforms both of them.

Using energy conservation arguments, we have specialized the results to a convex combination of LMS filters with different step-sizes, operating both in stationary and in nonstationary environments. The understanding about the functioning of the combination has suggested a filter with improved tracking performance for different speeds of changes.

Logical lines for further research include the analysis of convex combinations of multiple filters [18], or combinations with a different mixing coefficient for each tap [17], as well as the application of the proposed method to other scenarios, such as adaptive equalization or noise cancellation.

APPENDIX I

To derive (33) we proceed from (31). Rearranging terms we get

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

Now, we use the definitions (22) to rewrite the above expression as

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

But since $\bar{\lambda}(\infty) = \Delta J_2/(\Delta J_1 + \Delta J_2)$ and $1 - \bar{\lambda}(\infty) = \Delta J_1/(\Delta J_1 + \Delta J_2)$, we have

$$J_{\mathrm{ex}}(\infty) = \bar{\lambda}(\infty)\left[J_{\mathrm{ex},12}(\infty) + \frac{\Delta J_1 \Delta J_2}{\Delta J_1 + \Delta J_2}\right]$$
$$+ [1 - \bar{\lambda}(\infty)]\left[J_{\mathrm{ex},12}(\infty) + \frac{\Delta J_1 \Delta J_2}{\Delta J_1 + \Delta J_2}\right]$$

from which (33) is obtained.

APPENDIX II

To show that (48) minimizes the tracking EMSE of an LMS filter for $0 < \mu_i < 2/\mathrm{Tr}(\mathbf{R})$, we note first that $J_{\mathrm{ex},i}(\infty)$ given by (47) and its first derivative with respect to μ_i are continuous functions over that interval, and that, since $\sigma_0^2, \mathrm{Tr}(\mathbf{Q}), \mathrm{Tr}(\mathbf{R})$ are all positive constants, $0 < J_{\mathrm{ex},i}(\infty) < +\infty$. Furthermore, we have that

$$\lim_{\mu_i \to 0^+} J_{\mathrm{ex},i}(\infty) = \lim_{\mu_i \to [2/\mathrm{Tr}(\mathbf{R})]^-} J_{\mathrm{ex},i}(\infty) = +\infty.$$

Consequently, $J_{\mathrm{ex},i}(\infty)$ must have at least one minimum inside the interval $(0, 2/\mathrm{Tr}(\mathbf{R}))$. Setting the derivative of $J_{\mathrm{ex},i}(\infty)$ with respect to μ_i equal to zero, we get that only two candidate minima exist

$$\tilde{\mu}_{\mathrm{opt}} = \pm\sqrt{\frac{\mathrm{Tr}(\mathbf{Q})}{\sigma_0^2 \mathrm{Tr}(\mathbf{R})} + \frac{[\mathrm{Tr}(\mathbf{Q})]^2}{4\sigma_0^4}} - \frac{\mathrm{Tr}(\mathbf{Q})}{2\sigma_0^2}.$$

From these two values, it can be easily verified that only (48) satisfies $0 < \tilde{\mu}_{\mathrm{opt}} < 2/\mathrm{Tr}(\mathbf{R})$, and the proof concludes.

REFERENCES

[1] R. W. Harris, D. M. Chabries, and F. A. Bishop, "Variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, pp. 309–316, 1986.

[2] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, pp. 1663–1642, 1992.

[3] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: Analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, pp. 631–639, 1997.

[4] K.-R. Müller, A. Ziehe, N. Murata, and S.-I. Amari, "On-line learning in switching and drifting environments with application to blind source separation," in *On-Line Learning in Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[5] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, pp. 132–135, 2004.

[6] B. E. Ydstie, L. S. Kershenbaum, and R. W. H. Sargent, "Theory and application of an extended horizon self-tuning controller," in *AIChE J.*, vol. 31, 1985, pp. 1771–1780.

[7] W. Zhuang, "RLS algorithm with variable forgetting factor for decision feedback equalizer over time-variant fading channels," *Wireless Personal Commun.*, vol. 8, pp. 15–29, 1998.

[8] J. A. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least mean mixed-norm adaptive filtering," *Electron. Lett.*, vol. 30, pp. 1574–1575, 1994.

[9] D. I. Pazaitis and A. G. Constantinides, "LMS + F algorithm," *Electron. Lett.*, vol. 31, pp. 1423–1424, 1995.

[10] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," *IEEE Signal Process. Lett.*, vol. 4, pp. 46–48, 1997.

[11] E. V. Papoulis and T. Stathaki, "A normalized robust mixed-norm adaptive algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 11, pp. 56–59, 2004.

[12] P. Anderson, "Adaptive forgetting in recursive identification through multiple models," *Int. J. Contr.*, vol. 42, pp. 1175–1193, 1985.

[13] M. Niedźwiecki, "Multiple-model approach to finite memory adaptive filtering," *IEEE Trans. Signal Process.*, vol. 40, pp. 470–473, 1992.

[14] A. C. Singer and M. Feder, "Universal linear prediction by model order weighting," *IEEE Trans. Signal Process.*, vol. 47, pp. 2685–2700, 1999.

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

[16] M. Martínez-Ramón, J. Arenas-García, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "An adaptive combination of adaptive filters for plant identification," in *Proc. 14th Int. Conf. Digital Signal Processing*, Santorini, Greece, 2002, pp. 1195–1198.

[17] 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 Int. Workshop Neural Networks Signal Processing*, Toulouse, France, 2003, pp. 239–248.
[18] 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 Int. Symp. Intel. Signal Processing*, Budapest, Hungary, 2003, pp. 137–142.
[19] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
[20] N. Merhav and M. Feder, "Universal prediction," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2124–2147, 1998.

Jerónimo Arenas-García (S'00–M'04) was born in Seville, Spain, in 1977. He received the telecommunication engineer degree (honors) from Universidad Politecnica de Madrid, Madrid, Spain, in 2000 and the Ph.D. degree from Universidad Carlos III de Madrid.

He is now a Visiting Postdoc at the Technical University of Denmark. His present research interests are focused in the fields of adaptive signal processing, machine learning, and their applications.

Aníbal R. Figueiras-Vidal (S'74–M'76–SM'84) received the telecommunication engineer degree (honors) from Universidad Politécnica de Madrid, Madrid, Spain, in 1973 and the doctor degree (honors) from Universidad Politécnica de Barcelona, Barcelona, Spain, in 1976.

He is a Professor of signal theory and communications at Universidad Carlos III, Madrid. His research interests are digital signal processing, digital communications, neural networks, and learning theory. He has (co)authored more than 300 journal and conference papers in these areas.

Dr. Figueiras-Vidal received an "Honoris Causa" doctorate degree from Universidad de Vigo, Vigo, Spain, in 1999. He is currently General Secretary of the Royal Academy of Engineering of Spain.

Ali H. Sayed (S'90–M'92–SM'99–F'01) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1992.

He is a Professor and Chairman of electrical engineering at the University of California, Los Angeles (UCLA). He is also the Principal Investigator of the UCLA Adaptive Systems Laboratory. He is the author of *Fundamentals of Adaptive Filtering* (New York: Wiley, 2003) and is coauthor of *Indefinite Quadratic Estimation and Control* (Philadelphia, PA: SIAM, 1999) and of *Linear Estimation* (Englewood Cliffs, NJ: Prentice-Hall, 2000). He is also coeditor of *Fast Reliable Algorithms for Matrices with Structure* (Philadelphia, PA: SIAM, 1999). He has served on the editorial boards of the *SIAM Journal on Matrix Analysis and Its Applications,* the *International Journal of Adaptive Control and Signal Processing,* and the European *Signal Processing* journal. He has contributed several articles to engineering and mathematical encyclopedias and handbooks and has served on the program committees of several international meetings. He has also consulted with industry in the areas of adaptive filtering, adaptive equalization, and echo cancellation. His research interests span several areas, including adaptive and statistical signal processing, filtering and estimation theories, signal processing for communications, interplays between signal processing and control methodologies, system theory, and fast algorithms for large-scale problems. He has published widely in these areas.

Dr. Sayed received the 1996 IEEE Donald G. Fink Award and a 2002 Best Paper Award from the IEEE Signal Processing Society in the area of Signal Processing Theory and Methods. He was coauthor of two Best Student Paper awards at international meetings. He received the 2003 Kuwait Prize in Basic Sciences. He is a member of the IEEE Signal Processing Society technical committees on Signal Processing Theory and Methods and on Signal Processing for Communications. He has served as Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is now serving as Editor-in-Chief of the *EURASIP Journal on Applied Signal Processing*.