A NEW LEAST SQUARES ADAPTATION SCHEME FOR THE AFFINE COMBINATION OF TWO ADAPTIVE FILTERS

Luis A. Azpicueta-Ruiz, Aníbal R. Figueiras-Vidal, and Jerónimo Arenas-García*

Department of Signal Theory and Communications
Universidad Carlos III de Madrid
28911 Leganés-Madrid, Spain
{lazpicueta,arfv,jarenas}@tsc.uc3m.es

ABSTRACT

Adaptive combinations of adaptive filters are an efficient approach to alleviate the different tradeoffs to which adaptive filters are subject. The basic idea is to mix the outputs of two adaptive filters with complementary capabilities, so that the combination is able to retain the best properties of each component. In previous works, we proposed to use a convex combination, applying weights $\lambda(n)$ and $1-\lambda(n)$, with $\lambda(n) \in (0,1)$, to the filter components, where the mixing parameter $\lambda(n)$ was updated to minimize the overall square error using stochastic gradient descent rules. In this paper, we present a new adaptation scheme for $\lambda(n)$ based on the solution to a least-squares (LS) problem, where the mixing parameter is allowed to lie outside range $[0,1]$. Such affine combinations have recently been shown to provide additional gains. Unlike some previous proposals, the new LS combination scheme does not require any explicit knowledge about the component filters. The ability of the LS scheme to achieve optimal values of the mixing parameter is illustrated with several experiments in both stationary and tracking situations.

Index Terms— Adaptive filters, least squares (LS), combination of filters.

1. INTRODUCTION

Adaptive filtering schemes have become crucial components in many signal processing applications [1], [2]. Whatever kind of adaptive filter is used, a compromise involving speed of convergence, tracking capabilities, and steady-state error is always present. In order to alleviate these performance compromises, schemes that manage filter parameters (see, among many others, [3], [4]) have traditionally been used. A recent alternative consists in using adaptive combinations of adaptive filters with complementary capabilities [5], [6]. This approach is gaining popularity due to its simplicity and ability to deal with almost any kind of performance tradeoff in many applications [7]-[11].

In the basic combination scheme, two adaptive filters combine their outputs by means of a mixing parameter to obtain an overall output of improved performance:

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

where $y_i(n),\ i=1,2$ are the outputs of the component filters, $y(n)$ is the overall output, and n is the time index. It has been shown that, when mixing parameter $\lambda(n)$ is appropriately selected, the overall filter performs at least as the best individual component, or even better than any of them [6], [12]. Therefore, the design and study of good adaptation rules for $\lambda(n)$ is of primary importance to get the best out of the combination.

In [6] we studied the performance of one such adaptation rules, in which the mixing parameter was updated to minimize the overall square error using a gradient descent algorithm. Additionally, the mixing parameter was constrained to interval $(0,1)$ by defining it as the output of a sigmoid activation function, $\lambda(n) = [1+e^{-a(n)}]^{-1}$, so that (1) is a convex combination. Then, parameter $a(n)$ was adapted using the least-mean-square (LMS) algorithm:

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

where $e(n) = d(n) - y(n)$, and $d(n)$ is the desired output of the filter at each iteration. The value of $\lambda(n)$ can be recovered from $a(n)$ at each iteration.

Update (2) provides satisfactory performance when the step-size parameter μ_a is appropriately chosen. However, its correct adjustment depends on certain characteristics of the filtering scenario, such as the signal-to-noise (SNR) ratio or the speed of changes in tracking situations.

In order to simplify the selection of the step size for the mixing parameter, a normalized LMS (NLMS) adaptation scheme for the convex combination of two filters was presented in [13]. Both the LMS and NLMS based adaptation schemes for $\lambda(n)$ share the advantage of being completely general, in the sense that they can be used with any kind of adaptive filters and their application does not require any knowledge about the operation mechanisms of the component filters.

Recently, it has been theoretically found that the optimal values of the mixing parameter can take values outside range $[0,1]$, providing additional (though generally not very significant) performance gains [12]. For instance, for the particular case of a combination of two LMS filters with step sizes μ_1 and μ_2 ($\mu_1 > \mu_2$) for the first and second components, respectively, the optimal $\lambda(n)$ takes negative steady-state values in stationary situations.

In order to benefit from the potential advantages of negative mixing parameters, Bershad et. al [12] proposed to use

*This work was partly supported by by the Spanish Ministry of Education and Science under grant CICYT TEC-2005-00992 and by Madrid Community grant S-505/TIC/0223.

the following rule for selecting the mixing parameter.

$$\lambda(n) = 1 - \kappa \text{erf}\left(\frac{\hat{e}_1^2(n)}{\hat{e}_2^2(n)}\right), \tag{3}$$

where $\hat{e}_i^2(n), i = 1, 2$ are estimations of the instantaneous mean square errors of both component filters, which can be obtained as time averages over a rectangular window of length K, i.e.,

$$\hat{e}_i^2(n) = \frac{1}{K} \sum_{m=n-K+1}^{n} e_i^2(m), \tag{4}$$

where $e_i(n) = d(n) - y_i(n)$. This scheme has the advantage of not requiring to adjust any step size for the mixing parameter, while allowing negative values for $\lambda(n)$ when $\kappa > 1$. Selection of an appropriate value for κ, however, requires to know which is the optimal steady-state value for the mixing parameter, and this result is in general dependent on the particular filters being combined, as well as the characteristics of the filtering scenario.

Based on a statistical analysis of the LMS filter, and using several assumptions, an appropriate value of κ was found to be

$$\kappa = 1 + \frac{\mu_2}{2(\mu_1 - \mu_2)} \tag{5}$$

for the particular case of a combination of two LMS filters. However, selection of κ is not evident when an exact analysis of the mean-square performance of the component filters (and of the cross-correlation between their errors) is not plausible, or when it is subject to approximation errors. Furthermore, the fact that this adaptation scheme is not explicitly searching for the minimization of the square error can result in suboptimal operation in certain circumstances, as we will see later in the experiments section.

In this paper, we present a new adaptation scheme for $\lambda(n)$ based on the solution of a very simple least-squares (LS) problem. As in [12], we allow the mixing parameter to lie outside interval $[0, 1]$, and rely also on time averages through a window of the most recent samples to obtain good estimations of appropriate values for the parameter. Our scheme, however, does not require any specific knowledge about the particular kind of filters in the combination, but it proceeds directly to the minimization of the square error.

In the next section, we introduce the new LS adaptation rule for $\lambda(n)$, and provide some analytical insight about its expected performance. In Section 3 we will carry out several experiments, showing the efectiveness of the new scheme both in stationary and tracking conditions, and comparing its behavior to that of some previous proposals.

2. LEAST SQUARES ADAPTATION OF THE MIXING PARAMETER

2.1. Algorithm Description

Consider the following LS problem:

$$J[\lambda(n)] = \sum_{i=1}^{n} \beta(n, i) e^2(n, i) \tag{6}$$

where $\beta(n, i)$ are the weighting coefficients associated to each time instant, $e(n, i) = d(i) - y(n, i)$, and $y(n, i)$ would be the prediction of the overall filter for $d(i)$ if the outputs of the component filters were combined with parameter $\lambda(n)$, i.e,

$$\begin{aligned} y(n, i) &= \lambda(n) y_1(i) + [1 - \lambda(n)] y_2(i) \\ &= y_2(i) + \lambda(n)[y_1(i) - y_2(i)]. \end{aligned} \tag{7}$$

In order to obtain the value of the mixing parameter that minimizes (6), we can take derivatives with respect to $\lambda(n)$,

$$\frac{\partial J[\lambda(n)]}{\partial \lambda(n)} = 2 \sum_{i=1}^{n} \beta(n, i) e(n, i)[y_1(i) - y_2(i)]. \tag{8}$$

Now, making this expression equal to 0, and solving out for $\lambda(n)$, we arrive at the following equation for the optimal value of the mixing parameter at iteration n:

$$\lambda_{LS}(n) = \frac{\sum_{i=1}^{n} \beta(n, i)[d(i) - y_2(i)][y_1(i) - y_2(i)]}{\sum_{i=1}^{n} \beta(n, i)[y_1(i) - y_2(i)]^2}. \tag{9}$$

This result can be interpreted in an intuitive way in the light of the second line of (7). That expression for the overall filter output suggests that the combination of filters can be seen as a “second-level” one tap filter with weight $\lambda(n)$, whose input signal and desired output are $y_1(n) - y_2(n)$ and $d(n) - y_2(n)$, respectively. Then, (9) is coherent with the solution of an LS problem, since the denominator can then be interpreted as the autocorrelation of the input signal (with zero lag), while the numerator would be an estimation of the cross-correlation between the input and desired signals (also with zero lag).

Selection of the window $\beta(n, i)$ plays an important role in the proposed algorithm. The choice of an exponentially-weighted window, $\beta(n, i) = \beta^{n-i}$, with $(0 < \beta \lesssim 1)$, would make it easy to derive an RLS-like algorithm for updating $\lambda(n)$. For this one-tap filtering problem, however, computational convenience is not an issue, since the direct application of (9) is very easy and computationally efficient.

In the practice, we have found that a rectangular window

$$\beta(n, i) = \begin{cases} 1 & , n - i < K \\ 0 & , n - i > K \end{cases}$$

where K is the window length, offers improved performance with respect to the exponentially-weighted window, as we shall discuss later in the experiments section.

2.2. Analysis of the LS adaptation rule for $\lambda(n)$

In this section we give some theoretical insight about the ability of (9) to provide appropriate values for the mixing parameter. We assume a stationary data model similar to that in [6]: $d(n) = \mathbf{w}_o^T \mathbf{u}(n) + e_0(n)$, where $\mathbf{w}_o$ is an unkwnown system, $\mathbf{u}(n)$ is the input to the component filters, satisfying $E\{\mathbf{u}(n)\} = \mathbf{0}$ and $E\{\mathbf{u}(n)\mathbf{u}^T(n)\} = \mathbf{R}$, and $e_0(n)$ is zero-mean i.i.d. noise with variance σ_0^2. Under this model, we can define the *a priori* errors of the component filters as $e_{a,i}(n) = e_i(n) - e_0(n), i = 1, 2$.

Using similar arguments to those in [6] and [12], it can be shown that the optimal value for the mixing parameter is given by

$$\lambda_o(n) = \frac{\Delta J_2(n)}{\Delta J_1(n) + \Delta J_2(n)} \tag{10}$$

where we have defined $\Delta J_i(n) = E\{e_{a,i}^2(n) - e_{a,1}(n) e_{a,2}(n)\}, i = 1, 2$.

On the other hand, if we take the expectation of (9), it is easy to see that

$$E\{\lambda_{LS}(n)\} \approx \frac{\sum_{i=1}^{n} \beta(n,i) E\{[d(i) - y_2(i)][y_1(i) - y_2(i)]\}}{\sum_{i=1}^{n} \beta(n,i) E\{[y_1(i) - y_2(i)]^2\}} \tag{11}$$

where we have approximated the expectation of the quotient by the quotient of the expectations of the numerator and denominator. It can be shown that such an approximation introduces a negligible error for not too small K, since the numerator and denominator in (9) are already reasonably good estimates of the correlation and cross-correlation of the involved signals using time averages over several iterations.

Now, noting that $y_1(i) - y_2(i) = e_{\mathrm{a},2}(i) - e_{\mathrm{a},1}(i)$, and that $d(i) - y_2(i) = e_2(i) = e_{\mathrm{a},2}(i) + e_0(n)$, and after some algebraic manipulations, we arrive at

$$E\{\lambda_{LS}(n)\} \approx \frac{\sum_{i=1}^{n} \beta(n,i) \Delta J_2(i)}{\sum_{i=1}^{n} \beta(n,i) [\Delta J_1(i) + \Delta J_2(i)]} \tag{12}$$

Obviously, in steady state, this value approximates the optimal value for the mixing parameter given by (10), with the approximation being more accurate as the length of the window increases. On the other hand, in transient situations, or in other time-varying situations, $\Delta J_i(n)$ changes over time. Therefore, in these cases, it would be desirable to use shorter windows to reduce the bias in the estimation of the mixing parameter (obviously, at the cost of an increased variance).

3. EXPERIMENTS

In this section we study the performance of the new LS scheme for adapting the mixing parameter of an affine combination in a plant identification setup. Two sets of experiments have been carried out. First, we focus on the new LS rule, and study how different kinds of windows can influence its performance, both in stationary and tracking conditions. In the second block of experiments, the performance of the new rule is compared to that of previous schemes for adapting the mixing parameter.

3.1. Performance of the LS adaptation rule

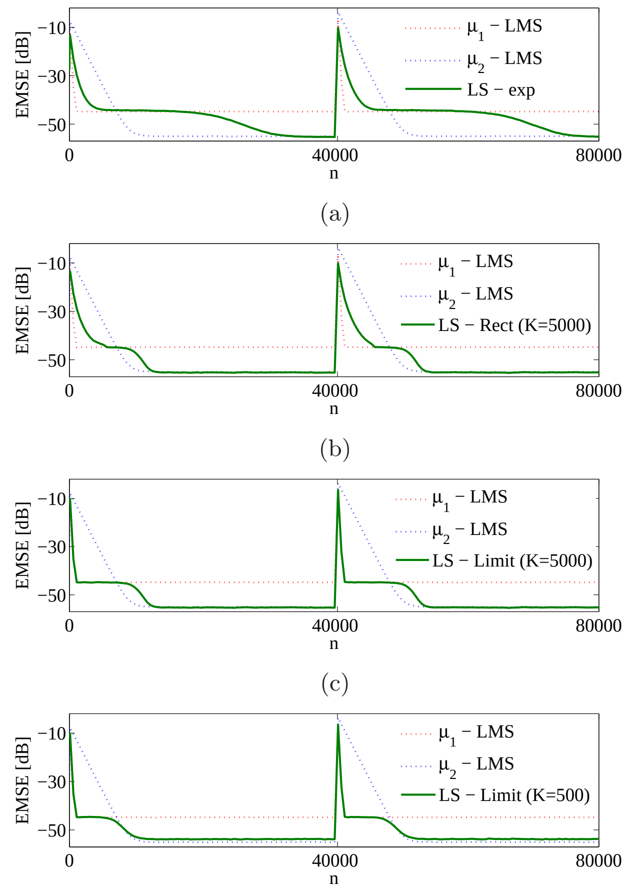
For these simulations we have considered a real plant with $M = 16$ coefficients, whose initial values are taken at random from interval $[-1, 1]$. When studying convergence and stationary behavior, plant coefficients are changed abruptly during the experiment, to study the ability of the algorithm to reconverge after a fast transition in the plant. The input signal, $u(n)$ is i.i.d. Gaussian noise with power $\sigma_u^2 = \frac{1}{16}$, so that $\mathrm{Tr}(\mathbf{R}) = 1$. The output aditive noise, $e_0(n)$, is the same kind of signal, and its variance has been selected to get an SNR of 20 dB.

We have used the excess mean-square-error (EMSE),

$$\mathrm{EMSE}(n) = E\{e^2(n) - e_0^2(n)\}$$

as a figure of merit of the behavior of the combination schemes, averaging the results over 1000 independent realizations.

In Fig. 1 we have displayed the behavior of a combination of two LMS filters when using LS adaptation for the mixing parameter. The step sizes for the LMS components are $\mu_1 =$ 0.1 and $\mu_2 = 0.01$, so that the first filter adapts faster. Two different windows were selected for the adaptation of $\lambda(n)$ as follows:



(d)

Fig. 1. EMSEs of the component LMS filters ($\mu_1 = 0.1$ and $\mu_2 = 0.2$) and of the proposed LS algorithm (9). (a) Using a exponentially-weighted window. (b) Using a uniform window ($K = 5000$). (c) Using a uniform window ($K = 5000$) and imposing $\lambda(n) \leq 1$. (d) Using a uniform window ($K = 500$) and imposing $\lambda(n) \leq 1$.

- LS-exp [Subfig. 1(a)]: exponentially-weighted window, $\beta(n,i) = \beta^{n-i}$.
- LS-rect [Subfig. 1(b)]: uniform window of length K.

The memory of both schemes, controlled by parameters β and K, was fixed long enough to guarantee that a nearly-optimal steady-state value of $\lambda(n)$ was achieved, leading to $\beta = 0.9997$ and $K = 5000$. In the light of the results, it seems evident that a more adequate performance is obtained when using the rectangular window (LS-rect). The exponentially-weighted window introduces a very long delay in the transfer between filter components. This is due to the fact that quadratic errors during convergence (i.e., just after $n = 0$ and $n = 40000$) are much larger than the errors incurred after the two component filters have converged, thus having a significant influence in the cost function (6), even when multiplied by a small factor.

As we discussed at the end of Section 2, a long memory allows for a very accurate estimation of the steady-state value of $\lambda_{\rm o}(n)$, but can actually damage the combination performance in transient situations. This is in fact the reason for observing that the overall EMSE situates well above that of the fast component following the transitions in the plant. In order to avoid this negative effect, it is enough to impose an upper limit on the mixing parameter: $\lambda_{LS}(n) \leq 1$, as we can check in Subfig. 1(c).

We have finally explored how the length of the window (parameter K) affects the EMSE of the combination. In Subfig. 1(d) we have represented EMSE evolution for $K = 500$. As we can see, the transfer between filter components occurs now with almost no delay, but the steady-state EMSE suffers a slight increment.

In order to assess the performance of the proposed LS adaptation rule for $\lambda(n)$, it is necessary to study its tracking abilities. Tracking situations have been analyzed assuming a random-walk model for the plant coefficients

$$\mathbf{w}_{\rm o}(n+1) = \mathbf{w}_{\rm o}(n) + \mathbf{q}(n),$$

where $\mathbf{q}(n)$ are i.i.d. zero-mean Gaussian random vectors with covariance matrix $\mathbf{Q} = E\{\mathbf{q}(n)\mathbf{q}^T(n)\} = \sigma_q^2\mathbf{I}$. We will study filter performance for varying $\mathrm{Tr}(\mathbf{Q})$, which can be seen as a measure of the degree of non-stationarity in the plant.

In these experiments, the normalized square deviation (NSD) of a filter, defined as the ratio of its EMSE to that of the LMS filter with optimal step size, is used as the figure of merit. Note that in the tracking situation we are considering, there exists an LMS filter with optimal performance, whose step size depends on the speed of changes and is given by [6, Eq. (48)]:

$$\mu_{\rm opt} = \sqrt{\frac{\mathrm{Tr}(\mathbf{Q})}{{\sigma_0}^2\mathrm{Tr}(\mathbf{R})} + \frac{[\mathrm{Tr}(\mathbf{Q})]^2}{{\sigma_0}^4}} - \frac{\mathrm{Tr}(\mathbf{Q})}{2{\sigma_0}^2}. \quad (13)$$

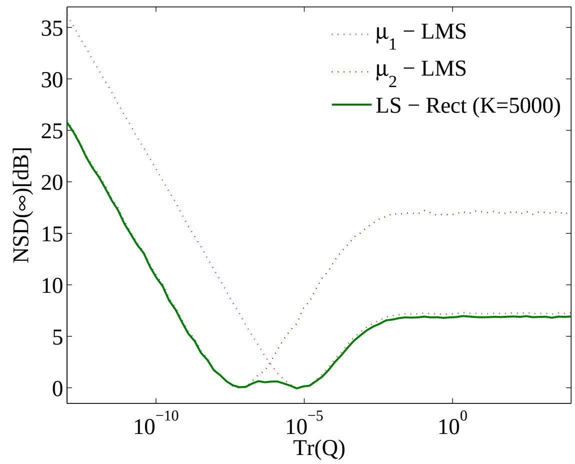
Therefore, the steady-state ($n \to \infty$) NSD of any filter (either a component or the combination) is defined as

$$\mathrm{NSD}(\infty) = \frac{\mathrm{EMSE}(\infty)}{\mathrm{EMSE}_{\rm opt}(\infty)}, \quad (14)$$

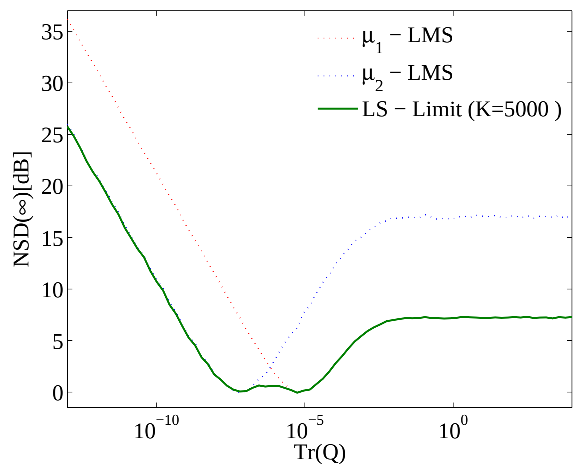
where $\mathrm{EMSE}_{\rm opt}(\infty)$ is the steady-state EMSE of the LMS filter with step size given by (13).

Fig. 2 represents, as a function of $\mathrm{Tr}(\mathbf{Q})$, the steady-state NSD achieved by the two LMS components, as well as that incurred by their combination using the proposed LS adaptation rule for the mixing parameter. We have considered both the cases where no limits are imposed on $\lambda_{LS}(n)$ [Subfig. 2(a)], and the case in which an upper limit of 1 has been imposed [Subfig. 2(b)]. All results have been averaged over 25000 iterations once the algorithms reached steady-state, and over 20 independent runs.

As it can be seen, the combination scheme offers an appropriate behavior for a wide range of degrees of non-stationarity, achieving, at least, the smallest of the NSDs of the component filters. When comparing Subfigs. 2(a) and 2(b), it is clear that imposing an upper limit of 1 on the mixing parameter has a very minor influence on filter performance. Therefore, in the next subsection, we will focus on this particular implementation of the LS combination scheme, and compare its performance to that of some other schemes for learning the mixing parameter.



(a)



(b)

Fig. 2. Steady-state NSD of two LMS filters, and of their adaptive combination using LS adaptation for $\lambda(n)$ with a uniform window (K=5000). (a) No constraints are imposed on $\lambda_{LS}(n)$. (b) $\lambda_{LS}(n) \leq 1$.

3.2. Comparison to other adaptation rules

For comparison purposes, in this subsection we have considered two other methods for the adaptation of the mixing parameter: 1) a normalized gradient descent algorithm from [13], and 2) update rule (3), taken from [12].

Fig. 3 represents the convergence and steady-state performance of all three combination schemes for the same scenario used in the previous subsection (notice the different scaling of the time axis), and compare then to the optimal (but unrealizable) EMSE of any affine combination of the μ_1 and μ_2-LMS filters, that would be obtained if using $\lambda_{\rm o}(n)$.

All three combination schemes are able to follow the fast convergence of the μ_1-LMS and achieve the lower steady-state

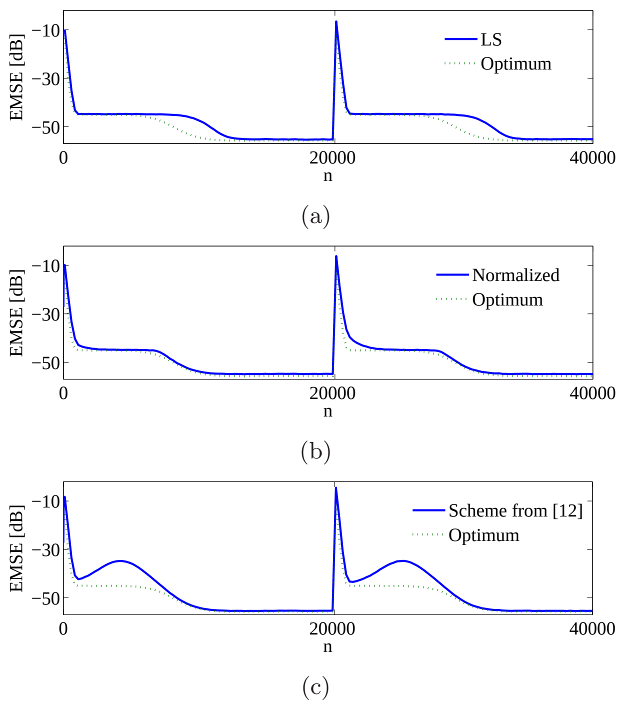


Fig. 3. EMSEs of different schemes for adaptively combining adaptive filters. The optimal (but unrealizable) EMSE of any affine combination of filters μ_1 and μ_2-LMS filters is also included for comparison purposes. (a) LS adaptation of the mixing parameter (this paper). (b) Normalized adaptation rule from [13]. (c) Scheme from [12].

error of the second component. Different deviations are observed, however, with respect to the optimal affine combiner. Regarding LS [Subfig. 3(a)], we can observe certain delay in the transition between filters. As for the normalized adaptation rule [Subfig. 3(b)], in this case it seems to present a better behavior than any of the two other schemes, but it has the inconvenience of not allowing negative values for the mixing parameter, what can result in suboptimal operation in other cases. Finally, a bump is observed for the scheme from [12] [Subfig. 3(c)], which is probably a consequence of the fact that this scheme is not explictly minimizing the overall square error. In some other filtering scenarios this bump may not appear, and this scheme could provide a more satisfactory performance.

In Fig. 4 we have represented the steady-state NSD of the three combination schemes, again as a function of $\mathrm{Tr}(\mathbf{Q})$. The performance of the LS and the normalized adaptation rules is very similar, with LS achieving a very minor advantage for slowly changing scenarios, as a consequence of using negative values for the mixing parameter. The scheme from [12] can also exploit the advantages of negative values for $\lambda(n)$. On the other hand, it shows suboptimal performance for a wide range of $\mathrm{Tr}(\mathbf{Q})$.

4. CONCLUSIONS

Adaptive combinations are a flexible and versatile approach to improve the performance of adaptive filters. In order to get the best out of them, however, the combination needs to be appropriately adjusted at each iteration. In this paper, we have presented a novel rule for adapting the parameter that controls the combination, which is based on the solution to a very simple LS problem. The behavior of our proposal has been illustrated by means of several experiments in both stationary and tracking situations, and shown to be competitive with respect to existing schemes.

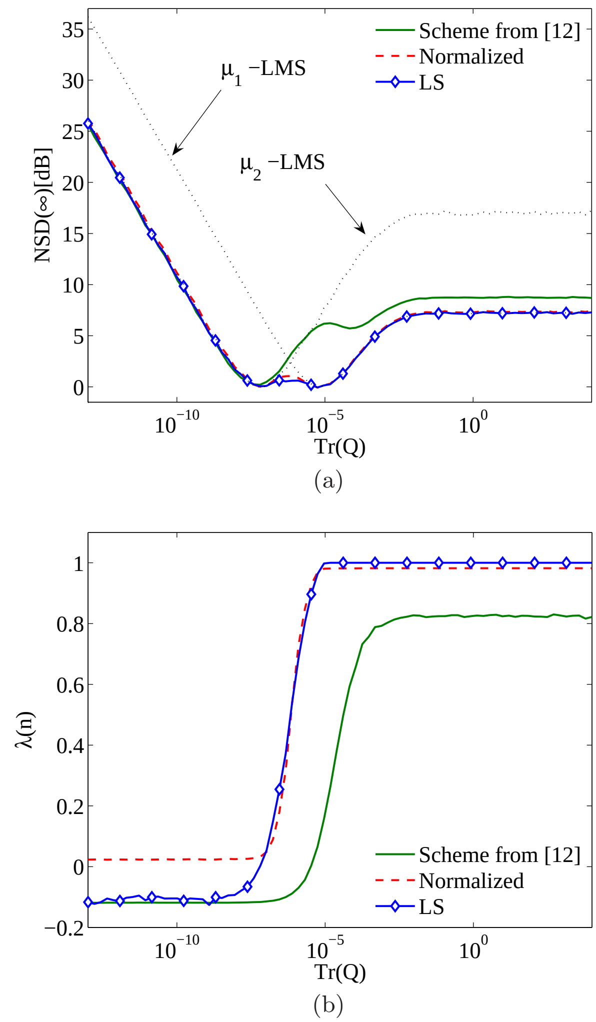


Fig. 4. Tracking performance of different schemes for adaptively combining adaptive filters. (a) NSD steady-state performance. (b) Steady-state value of the mixing parameter for each method.

5. REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*, New York: Wiley, 2003.

[2] S. Haykin, *Adaptive Filter Theory*, Upper Saddle River, NY: Prentice Hall, 2002.

[3] 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, Feb. 2004.

[4] 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, Aug. 1998.

[5] 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 Intl. Conf. Digital Signal Processing*, Santorini, Greece, 2002, pp. 1195–1198.

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

[7] Y. G. Zhang and J. A. Chambers, "Convex Combination of Adaptive Filters for a variable tap-length LMS algorithm," *IEEE Signal Proc. Letters*, vol. 13, pp. 628-631, 2006.

[8] A. H. Sayed and C. Lopes, "Adaptive Processing over Distributed Networks," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90A, pp. 1504-1510, 2007.

[9] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "Acoustic echo cancellation in frequency domain using combinations of filters," in *19th Intl. Congress on Acoustics (ICA)*, Madrid, Spain, 2007.

[10] D. P. Mandic, et al. "Online tracking of the degree of nonlinearity within complex signals," in *Proc. ICASSP'08*, Las Vegas, NV, 2008, pp. 2061-2064.

[11] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capabilities of adaptive filters via adaptive combination," *IEEE Trans. Signal Process.*, to appear, 2008.

[12] N. J. Bershad, J. C. M. Bermudez, and J.-Y. Tourneret, "An affine combination of two LMS adaptive filters – transient mean-square analysis," *IEEE Trans. Signal Process.*, vol. 56, pp. 1853–1864, 2008.

[13] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. ICASSP'08*, Las Vegas, NV, 2008, pp. 3301-3304.