A NORMALIZED ADAPTATION SCHEME FOR THE CONVEX 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 filtering schemes are subject to different tradeoffs regarding their steady-state misadjustment, speed of convergence and tracking performance. To alleviate these compromises, a new approach has recently been proposed, in which two filters with complementary capabilities adaptively mix their outputs to get an overall filter of improved performance. Following this approach, in this paper we propose a new normalized rule for adapting the mixing parameter that controls the combination. The new update rule preserves the good features of the standard scheme and is more robust to changes in the filtering scenario, for instance when the signal to noise ratio (SNR) is time varying. The benefits of the normalized scheme are illustrated analytically and with a number of experiments in both stationary and tracking situations.

Index Terms— Adaptive filters, least mean square methods, tracking filters

1. INTRODUCTION

Adaptive filtering schemes have become crucial components in many signal processing applications [1, 2]. Whatever the kind of adaptive filter used, a compromise involving speed of convergence, tracking capability and steady-state error is always present. Traditionally, schemes that manage the step-size have been proposed to alleviate this drawback (see, among others, [3, 4, 5]). However, these algorithms normally introduce several parameters, and some *a priori* knowledge about the statistics of the filtering scenario is needed for appropriately tuning them.

Recently, a new approach based on the adaptive combination of filters has been proposed [6, 7]. The basic idea is that two (or more) adaptive filters with complementary capabilities adaptively combine their outputs by means of a mixing parameter, to obtain an overall filter 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, $\lambda(n)$ is a mixing parameter in the range $(0, 1)$, and n is the time index. If $\lambda(n)$ is conveniently updated, it can be shown that the resulting filter performs like the best individual component, or even better than any of them under certain conditions. For instance, in [7] we analyzed the performance of the adaptation rule

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

where $e(n) = d(n) - y(n)$ and $e_i(n) = d(n) - y_i(n), i = 1, 2,$ are the errors of the overall filter and the components, $d(n)$ being the desired signal, and $a(n)$ is a parameter that defines $\lambda(n)$ via a sigmoidal function as[1]

$$\lambda(n) = \mathrm{sgm}[a(n)] = \left(1 + e^{-a(n)}\right)^{-1}. \tag{3}$$

Update equation (2) provides satisfactory performance when the step-size μ_a parameter is appropriately chosen. However, its correct adjustment also depends on some characteristics of the filtering scenario, such as the input signal and additive noise powers, or the step-sizes of the adaptive filters included in the combination.

In this paper we present a novel normalized scheme for the adaptation of $a(n)$. The operation of the new update scheme will be found to be independent of the signal to noise ratio (SNR), thus simplifying the selection of step-size μ_a, and providing robustness to scenarios in which the involved signals present a wide dynamic range.

In the next section we will present the new adaptation scheme, and give an analytical justification for its improved performance. Then, in Section 3, we provide several experiments comparing the performance of the normalized rule to the standard combination scheme of [7], both in stationary and tracking situations.

2. NORMALIZED ADAPTATION OF THE MIXING PARAMETER

2.1. Algorithm Description

A closer look at (2) reveals that the standard adaptation rule for the mixing parameter is equivalent to that of a Least Mean Squares (LMS) filter with varying step-size $\mu_a\lambda(n)[1 - \lambda(n)]$ and input signal $e_2(n) - e_1(n)$. Indeed, this interpretation

*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.

[1] Introduction of parameter $a(n)$ and the activation function is justified as an easy way to keep $\lambda(n) \in (0, 1)$ and to reduce gradient noise near $\lambda(n) = 1$ or $\lambda(n) = 0$. The interested reader is referred to [7] for further details.

is reinforced by the fact that the output of the combination filter can be rewritten as

$$\begin{aligned} y(n) &= y_2(n) + \lambda(n)[y_1(n) - y_2(n)] \\ &= y_2(n) + \lambda(n)[e_2(n) - e_1(n)], \end{aligned} \quad (4)$$

so that we can think of the overall combination scheme as a two-layer adaptive filter. In the first layer, the two component filters operate independently of each other and according to their own rules, while the second layer consists of a filter with input signal $e_2(n) - e_1(n)$ that minimizes the overall error.

This interpretation of the combination scheme suggests that further advantages could be obtained if we used a normalized LMS rule for adapting the mixing parameter rather than standard LMS. Since $e_2(n) - e_1(n)$ plays the role of the input signal at this level, it makes sense to use the following adaptation scheme:

$$a(n+1) = a(n) + \frac{\mu_a \lambda(n)[1-\lambda(n)]}{[e_2(n) - e_1(n)]^2} e(n)[e_2(n) - e_1(n)]. \quad (5)$$

In practice, however, the performance of this scheme is quite unsatisfactory given that the instantaneous value $[e_2(n) - e_1(n)]^2$ is a very poor estimate of the power of the "second layer" input signal. Similar to the Normalized LMS (NLMS) algorithm with power normalization [1, Sec. 5.6], a better behavior is obtained from

$$a(n+1) = a(n) + \frac{\mu_a}{p(n)} \lambda(n)[1-\lambda(n)]e(n)[e_2(n) - e_1(n)], \quad (6)$$

where $p(n)$ is a rough (low-pass filtered) estimation of the power of the signal of interest

$$p(n) = \beta p(n-1) + (1-\beta)[e_2(n) - e_1(n)]^2. \quad (7)$$

Selection of the forgetting factor β is rather easy. For instance, using $\beta = 0.9$ gives a good enough approximation, and typically ensures that $p(n)$ is adapted faster than any component filter.

2.2. Analysis of the Normalized Update Rule

In this section we give some theoretical insight for the preference of (6) over the standard rule for $a(n)$ adaptation. For this, we will assume a data model similar to that in [7]: $d(n) = \mathbf{w}_0^T \mathbf{u}(\mathbf{n}) + e_0(n)$, where $\mathbf{w}_0$ is an unknown system, $\mathbf{u}(n)$ is the input to the component filters satisfying $E\{\mathbf{u}(n)\} = 0$ 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 $e_{a,i}(n) = e_i(n) - e_0(n), i = 1, 2$.

To make the analysis feasible, we will make the further assumption that $p(n)$ is a perfect estimation of $[e_2(n) - e_1(n)]^2$, i.e.,

$$\begin{aligned} p(n) &= E\{[e_2(n) - e_1(n)]^2\} = E\{[e_{a,2}(n) - e_{a,1}(n)]^2\} \\ &= \Delta J_1(n) + \Delta J_2(n) \end{aligned} \quad (8)$$

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

We proceed now similarly to [7]: Taking expectations on both sides of (6) and after some manipulations, we arrive at

$$\begin{aligned} &E\{a(n+1)\} = E\{a(n)\} \\ &\quad + \frac{\mu_a}{p(n)} E\{\lambda(n)[1-\lambda(n)]^2 [e_{a,2}^2(n) - e_{a,1}(n)e_{a,2}(n)]\} \\ &\quad - \frac{\mu_a}{p(n)} E\{\lambda^2(n)[1-\lambda(n)][e_{a,1}^2(n) - e_{a,1}(n)e_{a,2}(n)]\} \end{aligned} \quad (9)$$

Finally, we will accept that the adaptation of $a(n)$ is slow enough so that we can assume that $\lambda(n)$ is independent of the a priori errors of the filters. This assumption, which is more reasonable in steady-state (which includes both stationary and tracking situations), leads to:

$$\begin{aligned} E\{a(n+1)\} = E\{a(n)\} + \mu_a E\{\lambda(n)[1-\lambda(n)]^2\} \frac{\Delta J_2(n)}{p(n)} \\ - \mu_a E\{\lambda^2(n)[1-\lambda(n)]\} \frac{\Delta J_1(n)}{p(n)} \end{aligned} \quad (10)$$

In view of (10), we can shed some light on the advantages derived from using (6). Effectively, for most adaptive schemes $\Delta J_i(n)$ increases linearly with σ_0^2, and possibly with $\mathrm{Tr}(\mathbf{R})$ (see, e.g., [1, pp. 327, 387], [7]). However, since these quantities are divided by $p(n)$, which shows the same dependencies, we can conclude that the evolution of $E\{a(n)\}$ will be quite independent of both σ_0^2 and $\mathrm{Tr}(\mathbf{R})$. Thus, when using the normalized scheme, we can expect an easier and more robust selection of μ_a, in the sense that its optimal value will not depend on the SNR.

3. EXPERIMENTS

In this section we study the performance of both the standard [Eq. (2)] and the normalized [Eq. (6)] adaptation rules in a plant identification setup. Two kinds of experiments have been carried out: In the first set of experiments we pay attention to convergence and stationary behavior, while the second set of experiments considers a tracking situation. To include results with different types of component filters, NLMS and LMS schemes have been used for the first and second group of experiments, respectively.

3.1. Convergence and Stationary Performance

For these simulations the real plant consists of $M = 7$ coefficients whose initial values are randomly taken from interval $[-1, 1]$: $\mathbf{w}_0^T = [0.9003, -0.5377, 0.2137, -0.028, 0.7826, 0.5242, -0.0871]$. At $n = 15000$ the plant coefficients are changed abruptly to $\mathbf{w}_0^T = [0.2542, -0.4696, -0.3753, 0.0454, -0.1827, 0.7859, 0.1475]$ in order to study convergence properties when a rapid transition in the plant occurs. The input signal, $u(n)$, is the output of a first order AR model with transfer function $0.6/(1 - 0.8z^{-1})$, fed with i.i.d. Gaussian noise with unitary variance. The output additive noise, $e_0(n)$, is also white Gaussian noise whose variance is changed along the experiment: initally, σ_0^2 is set to get an SNR of 50 dB, and it suddenly increases at $n = 30000$ to get SNR$= 10$ dB.

Two NLMS filters with $\mu_1 = 0.5$ and $\mu_2 = 0.05$ have been used as the filter components. Step-sizes for the mixing parameter were independently adjusted to $\mu_a = 8000$ and $\mu_a = 1$ for the standard and normalized schemes, respectively, to get a satisfactory convergence. For the normalized scheme, we also used $\beta = 0.9$.

All displayed results have been averaged over 1000 independent runs using the Excess Mean-Square-Error (EMSE) as the figure of merit: $\mathrm{EMSE}(n) = E\{[e(n) - e_0(n)]^2\}$.

As it can be seen in Figs. 1(a) and 2(a), both combination approaches perform adequately for high SNR, showing the fast convergence of the fast NLMS filter, and the lower steady-state misadjustment of the μ_2-NLMS. However, when

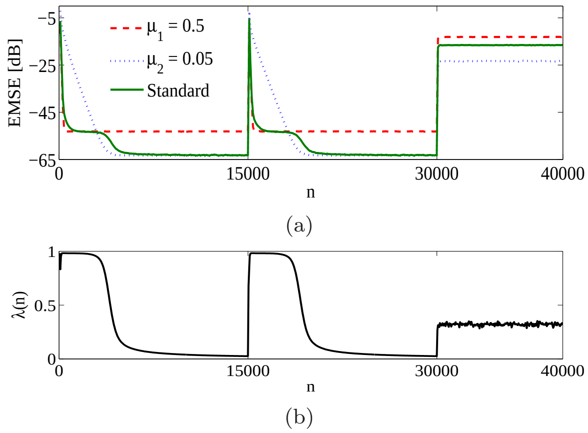

Fig. 1. (a) EMSEs of the component NLMS filters and of their adaptive combination using (2). (b) Evolution of the mixing paramenter $\lambda(n)$.

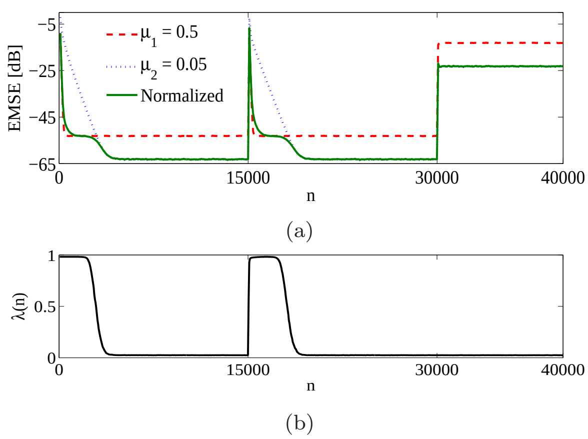

Fig. 2. (a) EMSEs of the component NLMS filters and of their adaptive combination using the normalized scheme. (b) Evolution of the mixing paramenter $\lambda(n)$.

the SNR decreases to 10 dB, using (2) results in a larger error as a consequence of the increased misadjustment of the components (i.e., larger ΔJ_1 and ΔJ_2). The normalized scheme, on the other hand, stays robust to this situation, and its steady state performance remains as low as that of the μ_2-NLMS. The superior performance of the normalized scheme can also be concluded from Figs. 1(b) and 2(b), examining the steady-state value of $\lambda(n)$ when SNR= 10 dB.

Stationary performance has also been studied for other SNRs, and the limiting value of the EMSE ($n \to \infty$) has been displayed in Fig. 3 as a function of the SNR. The displayed results have been obtained by averaging the EMSE over 10000 iterations once the algorithm convergence has taken place, and over 200 independent realizations. It can be seen that for all values of the SNR the combination filter with normalized updates behaves as the slower NLMS component (which achieves a lower error), while the standard scheme significantly deviates from the best component for SNR< 30 dB.

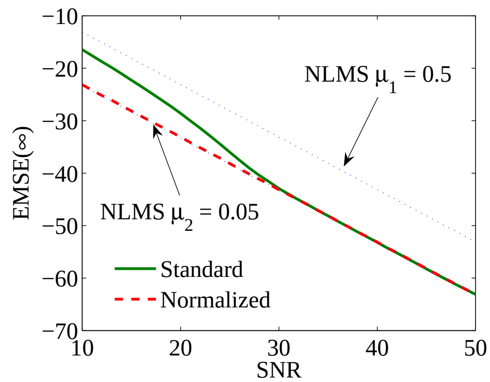

Fig. 3. Stationary steady-state EMSE of two NLMS filters and of their combination using standard and normalized adaptation for $a(n)$.

3.2. Tracking Performance

In order to evaluate tracking performance, in this section we will let the plant vary at each iteration following a random-walk model

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

where $\mathbf{q}(n)$ are i.i.d. zero-mean Gaussian random vectors with covariance matrix $\mathbf{Q} = E\{\mathbf{q}(n)\mathbf{q}^T(n)\} = \sigma_q^2\mathbf{I}$. From its definition, $\mathrm{Tr}(\mathbf{Q})$ can be seen as a measure of the speed of changes in the plant.

In this case, to obtain results comparable with [7], the input signal is obtained from the same AR model used in the previous subsection, but adjusting the variance of the exciting signal to get $\mathrm{Tr}(\mathbf{R}) = 1$. The variance of the i.i.d. additive noise $e_0(n)$ was again varied to obtain different SNRs.

For these experiments, the combination components are two LMS filters with $\mu_1 = 0.1$ and $\mu_2 = 0.01$. The step-size for the standard combination rule has been set to $\mu_a = 10000$, to assure good performance when operating with a high SNR of 50 dB. For the normalized scheme we have set $\beta = 0.9$ and $\mu_a = 0.05$ (note that this value is different from the one used in the previous subsection, as a consequence of using a different kind of adaptive filter components).

In these experiments, the Normalized Square Deviation (NSD) of a filter, defined as the quotient between its EMSE and the EMSE 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 [7, Eq. (48)]:

$$\mu_{\mathrm{opt}} = \sqrt{\frac{\mathrm{Tr}(\mathbf{Q})}{{\sigma_0}^2\mathrm{Tr}(\mathbf{R})} + \frac{[\mathrm{Tr}(\mathbf{Q})]^2}{{\sigma_0}^4}} - \frac{\mathrm{Tr}(\mathbf{Q})}{2{\sigma_0}^2}. \tag{11}$$

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

$$\mathrm{NSD}(\infty) = \mathrm{EMSE}(\infty)/\mathrm{EMSE}_{\mathrm{opt}}(\infty), \tag{12}$$

where $\mathrm{EMSE}_{\mathrm{opt}}(\infty)$ is the steady-state EMSE of the LMS filter with step-size given by (11).

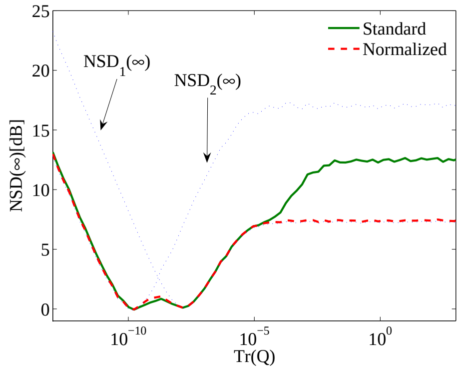

Fig. 4. Steady-state NSD of two LMS filters [$\mathrm{NSD}_1(\infty)$ and $\mathrm{NSD}_2(\infty)$] and of their adaptive combination using standard and normalized update rules for $a(n)$. SNR $=$ 50 dB.

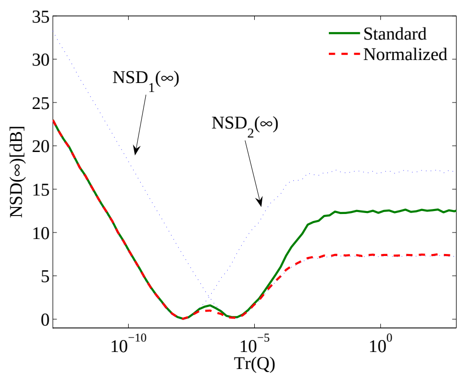

Fig. 5. Steady-state NSD of two LMS filters [$\mathrm{NSD}_1(\infty)$ and $\mathrm{NSD}_2(\infty)$] and of their adaptive combination using standard and normalized update rules for $a(n)$. SNR $=$ 30 dB.

The steady-state NSD of the two LMS components, as well as the NSD incurred by their combination using the standard and normalized update equations is shown as a function of $\mathrm{Tr}(\mathbf{Q})$ in Figs. 4, 5 and 6 for SNR= 50, 30 and 10 dB, respectively. All results have been averaged over 20000 iterations once the algorithms reached steady-state, and over 50 independent runs.

As it can be seen in the figures, the combination scheme with standard adaptation for the mixing parameter results in suboptimal performance when the plant changes very fast [i.e., for large $\mathrm{Tr}(\mathbf{Q})$]. This is due to the fact that both component filters are incurring in a very significant error, what results in non-negligible gradient noise when applying (2). Not only that, but we can also see how the performance of this scheme degrades, whatever the value of $\mathrm{Tr}(\mathbf{Q})$, as the SNR is decreased. On the other hand, when the normalized adaptation is applied, the combination shows a very stable operation, and behaves as well as the best component filter not only for any SNR, but also for all values of $\mathrm{Tr}(\mathbf{Q})$. This can be explained in similar terms to our discussion at the end of Subsection 3.2: In tracking situations $\Delta J_i(\infty)$ can also be shown to be proportional to $\mathrm{Tr}(\mathbf{Q})$, and normalization by $p(n)$ can compensate this effect.

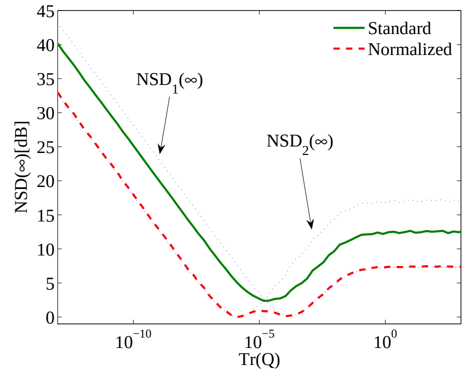

Fig. 6. Steady-state NSD of two LMS filters [$\mathrm{NSD}_1(\infty)$ and $\mathrm{NSD}_2(\infty)$] and of their adaptive combination using standard and normalized update rules for $a(n)$. SNR $=$ 10 dB.

4. CONCLUSIONS

Adaptive combinations of adaptive filters constitute a flexible and versatile approach to improve the performance of adaptive filters. In this paper, we have presented a normalized rule for updating the parameter that controls the combination. When compared to the standard (unnormalized) update scheme, the new approach results in a more stable behavior of the combination, and simplifies the selection of step-size μ_a. The superior behavior of the normalized rule has been justified theoretically, and illustrated by means of several experiments in both stationary and tracking situations, where very satisfactory results were achieved for very different SNR and tracking conditions.

5. REFERENCES

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

[2] S. Haykin, *Adaptive Filter Theory (4th ed.)*, Upper Saddle River, NY: Prentice Hall, 2002.

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

[4] 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.

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

[6] 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* pp. 1195–1198, 2002.

[7] 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. on Signal Process.*, vol. 54 ,pp. 1078–1090, Mar. 2006.