CONVEX COMBINATION OF ADAPTIVE FILTERS WITH DIFFERENT TRACKING CAPABILITIES

Magno T. M. Silva and Vítor H. Nascimento

Electronic Systems Engineering Department, Escola Politécnica, Universidade de São Paulo
Av. Prof. Luciano Gualberto, trav. 3, nº 158, CEP 05508-900 – São Paulo, SP, Brazil
E-mails: {magno, vitor}@lps.usp.br

ABSTRACT

Recently, an adaptive convex combination of two LMS (Least Mean-Square) filters was proposed and its tracking performance analyzed. Motivated by the performance of such scheme and by the differences between the tracking capabilities of the RLS (Recursive Least-Squares) and LMS algorithms, we propose a convex combination of one LMS and one RLS filter. The resulting combination should profit of the best tracking behavior of each component filter. A steady-state analysis via energy conservation relation is also presented for stationary and non-stationary environments.

Index Terms— Adaptive filters, convex combination, energy conservation, tracking analysis, LMS algorithm, RLS algorithm.

1. INTRODUCTION

It is well-known in the literature that adaptive filters that exhibit good convergence properties in stationary environments, do not necessarily present good tracking performance in nonstationary environments [1]. A classical example is the tracking behavior of the LMS (Least-Mean Square) and RLS (Recursive Least-Squares) algorithms [2]. Although the convergence performance of RLS is superior to that of LMS in stationary environments, there are situations where one outperforms the other and vice-versa in terms of tracking capability. In [2], Eweda showed that the performance of RLS and LMS is similar when the matrix $\mathbf{Q}$ (autocorrelation of the perturbation of the optimal solution, see Eq. (8) below) is a multiple of the identity matrix. For other choices of $\mathbf{Q}$, one algorithm may perform better than the other, as for example: (i) if $\mathbf{Q}$ is a multiple of $\mathbf{R}$ (autocorrelation of the input signal), LMS is superior and (ii) if $\mathbf{Q}$ is a multiple of $\mathbf{R}^{-1}$, RLS is superior.

In order to improve adaptive filter performance, a convex combination of one fast and one slow LMS filter was introduced in [3]. The mean-square performance of this combination was analyzed via energy conservation relations in [4]. Furthermore, it was shown that this structure is universal in the mean-square error sense, presenting a worst case performance as good as the best of its components, outperforming both of them when the correlation between the *a priori* errors of the component filters is low enough [4].

In order to obtain an algorithm that has good tracking performance for a larger range of environments, we propose an adaptive convex combination of one LMS and one RLS filter. The proposed combination should acquire the good initial convergence properties of RLS and be able to perform as well as the best of its components in regards to tracking behavior, for all kinds of nonstationary environments. Moreover, extending the steady-state analysis of [4], we obtain an analytical expression for the Excess Mean-Square Error (EMSE) for stationary and nonstationary environments.

The paper is organized as follows. In Section 2, the proposed combination is described. In Section 3, the steady-state analysis is presented. Simulation results and the conclusions are shown in sections 4 and 5, respectively.

2. PROBLEM FORMULATION

The convex combination of two adaptive filters proposed in [3] is depicted in Figure 1. In this scheme, the output of the overall filter is given by

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

where $y_i(n)$, $i = 1, 2$ are the outputs of the transversal filters, i.e., $y_i(n) = \mathbf{u}^T(n)\mathbf{w}_i(n-1)$, $\mathbf{u}(n) \in \mathbb{R}^M$ is the common regressor vector, and $\mathbf{w}_i(n-1) \in \mathbb{R}^M$ are the weight vectors of each length-M component filter. The mixing parameter $\eta(n)$ is modified via an auxiliary variable $a(n-1)$ and a sigmoidal function [3, 4], that is,

$$\eta(n) = \mathrm{sgm}[a(n-1)] = \frac{1}{1 + e^{-a(n-1)}}, \qquad (2)$$

with $a(n)$ being updated as

$$a(n) = a(n-1) + \mu_a e(n)[y_1(n) - y_2(n)]\eta(n)[1-\eta(n)], \quad (3)$$

This work was supported in part by FAPESP under grants 02/12216-3 and 04/15114-2, and CNPq under grant 303.361/2004-2.

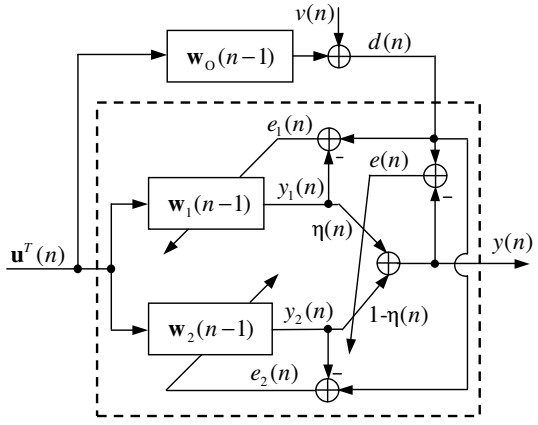

Fig. 1. Adaptive convex combination of two transversal filters.

where $e(n) = d(n) - y(n)$ and $d(n)$ is the desired signal. The sigmoidal function keeps $\eta(n)$ within the desired range $[0,\ 1]$. However, the update of $a(n)$ in (3) stops whenever $\eta(n)$ is close to 0 or 1. To avoid this, the values of $a(n)$ are restricted to lie inside a symmetric interval $[-a^+,\ a^+]$. Thus, a minimum level of adaptation is always guaranteed [4].

We assume that the coefficients $\mathbf{w}_1$ and $\mathbf{w}_2$ are adapted respectively by the RLS and LMS algorithms, that is,

$$\mathbf{w}_1(n) = \mathbf{w}_1(n-1) + \widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n)e_1(n) \quad (4)$$

$$\mathbf{w}_2(n) = \mathbf{w}_2(n-1) + \mu e_2(n)\mathbf{u}(n), \quad (5)$$

where μ is the LMS step-size and $e_i(n) = d(n) - y_i(n)$, $i = 1, 2$. A linear regression model is assumed, that is, $d(n) = \mathbf{u}^T(n)\mathbf{w}_{\rm o}(n-1)+v(n)$, with $\mathbf{w}_{\rm o}(n-1)$ being the time-variant optimal solution and $v(n)$ an i.i.d. (independent and identically distributed) and zero mean random process with variance $\sigma_v^2 = \mathrm{E}\{v^2(n)\}$, which plays the role of a disturbance uncorrelated with $\mathbf{u}(n)$ [1, Sec. 6.2.1]. The matrix $\widehat{\mathbf{R}}^{-1}(n)$ is obtained via the matrix inversion lemma [1, Eq. (2.6.4)] applied to

$$\widehat{\mathbf{R}}(n) = \lambda\widehat{\mathbf{R}}(n-1) + \mathbf{u}(n)\mathbf{u}^T(n), \quad (6)$$

where $0 \ll \lambda < 1$ is the forgetting factor, $\widehat{\mathbf{R}}(0) = \epsilon\mathbf{I}$, $\mathbf{I}$ is the identity matrix, and ϵ is a positive constant [1]. The autocorrelation matrix $\mathbf{R} \triangleq \mathrm{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}$ ($\mathrm{E}\{\cdot\}$ is the expectation operator) is related to $\widehat{\mathbf{R}}(n)$ via

$$\mathbf{R} = (1-\lambda)\mathrm{E}\{\widehat{\mathbf{R}}(n)\}. \quad (7)$$

3. TRACKING ANALYSIS

We assume that in a nonstationary environment, the variation in the optimal solution $\mathbf{w}_{\rm o}$ follows the random-walk model [1, p. 359], that is,

$$\mathbf{w}_{\rm o}(n) = \mathbf{w}_{\rm o}(n-1) + \mathbf{q}(n). \quad (8)$$

In this model, $\mathbf{q}(n)$ is an i.i.d. vector with positive-definite autocorrelation matrix $\mathbf{Q} = \mathrm{E}\{\mathbf{q}(n)\mathbf{q}^T(n)\}$ and is independent of the initial conditions $\{\mathbf{w}_{\rm o}(-1), \mathbf{w}(-1)\}$ and of $\{\mathbf{u}(l)\}$ for all $l < n$ [1, Sec. 7.4].

One measure of the filter performance is given by the EMSE, defined as

$$\zeta \triangleq \lim_{n\to\infty} \mathrm{E}\{e_a^2(n)\}, \qquad e_a(n) = \mathbf{u}^T(n)\widetilde{\mathbf{w}}(n-1),$$

and $\widetilde{\mathbf{w}}(n-1) = \mathbf{w}_{\rm o}(n-1) - \mathbf{w}(n-1)$. The *a priori* error $e_a(n)$ of the overall scheme can be written as a function of the *a priori* errors of the component filters, that is,

$$e_a(n) = \eta(n)e_{a,1}(n) + [1-\eta(n)]e_{a,2}(n), \quad (9)$$

where $e_{a,i}(n) = \mathbf{u}^T\widetilde{\mathbf{w}}_i(n-1)$ and $\widetilde{\mathbf{w}}_i(n-1) = \mathbf{w}_{\rm o}(n-1) - \mathbf{w}_i(n-1),\ i = 1, 2$.

Using the energy conservation approach of [1, Ch. 7], the EMSE of RLS (component 1 of the combination), in a nonstationary environment, is given by [1, Eq. (7.10.18)]

$$\zeta_1 = \frac{\sigma_v^2(1-\lambda)M + \dfrac{1}{(1-\lambda)}\mathrm{Tr}(\mathbf{QR})}{2-(1-\lambda)M}, \quad (10)$$

in which $\mathrm{Tr}(\mathbf{A})$ stands for the trace of the matrix $\mathbf{A}$. For LMS (component 2), we have [1, Eq. (7.5.9)]

$$\zeta_2 = \frac{\mu\sigma_v^2\mathrm{Tr}(\mathbf{R}) + \mu^{-1}\mathrm{Tr}(\mathbf{Q})}{2-\mu\mathrm{Tr}(\mathbf{R})}. \quad (11)$$

Using the same arguments of [4, Sec. III], it is possible to show that the considered scheme is universal in the mean square error sense. Thus, when RLS outperforms LMS in the steady-state, the behavior of the overall filter will be close to that of RLS and $\zeta \approx \zeta_1$. On the other hand, when LMS is superior, $\zeta \approx \zeta_2$. Moreover, there are situations where the combination will outperform both of them. In this case, the EMSE of the overall filter will be close to [4, Eq. (33)]

$$\zeta \approx \zeta_{12} + \frac{\Delta\zeta_1\Delta\zeta_2}{\Delta\zeta_1 + \Delta\zeta_2}, \quad (12)$$

where ζ_{12} is the cross-EMSE , defined as

$$\zeta_{12} \triangleq \lim_{n\to\infty} \mathrm{E}\{e_{a,1}(n)e_{a,2}(n)\}, \quad (13)$$

and $\Delta\zeta_i = \zeta_i - \zeta_{12},\ i = 1, 2$.

The EMSE of the overall filter is the minimum of the values calculated by the expressions (10), (11), and (12).

Analytical expressions for ζ_{12} have not been computed before for combinations of LMS and RLS. In order to evaluate ζ_{12}, we first subtract both sides of (4) and (5) from $\mathbf{w}_{\rm o}(n)$. Using (8), we arrive at

$$\widetilde{\mathbf{w}}_1(n) - \mathbf{q}(n) = \widetilde{\mathbf{w}}_1(n-1) - \widehat{\mathbf{R}}^{-1}(n)\mathbf{u}(n)e_1(n), \quad (14)$$

$$\widetilde{\mathbf{w}}_2(n) - \mathbf{q}(n) = \widetilde{\mathbf{w}}_2(n-1) - \mu e_2(n)\mathbf{u}(n). \quad (15)$$

In the sequel, we multiply the transpose of (15) by (14), using $\widehat{\mathbf{R}}(n)$ as a weighting matrix. After simple algebraic manipulations, we obtain a long expression and take the expectations of both of its sides. To simplify the resulting expression, the following assumptions are employed:

A1. In steady state,

$$\mathrm{E}\{\widetilde{\mathbf{w}}_2^T(n)\widehat{\mathbf{R}}(n)\widetilde{\mathbf{w}}_1(n)\} = \mathrm{E}\{\widetilde{\mathbf{w}}_2^T(n\text{–}1)\widehat{\mathbf{R}}(n)\widetilde{\mathbf{w}}_1(n\text{–}1)\}.$$

This assumes that the filters are operating in stable conditions, and have reached steady-state.

A2. From relation (7) and the assumed independence between $\mathbf{q}(n)$ and the regressor $\mathbf{u}(n)$, it follows that, in steady-state,

$$\mathrm{E}\{\mathbf{q}^T(n)\widehat{\mathbf{R}}(n)\mathbf{q}(n)\} = \frac{\mathrm{Tr}(\mathbf{QR})}{1-\lambda}.$$

A3. $\|\mathbf{u}(n)\|^2$ is independent of the *a priori* errors in steady-state.

This assumption, usually called *separation principle*, is very used in steady-state analysis of adaptive algorithms [1, Sec. 6.5.2]. Since $e_i(n) = e_{a,i}(n) + v(n),\ \ i = 1, 2$ and $v(n)$ is a zero mean process, which is independent of the *a priori* errors and of the regressor vector, an immediate consequence of the separation principle is that $\|\mathbf{u}(n)\|^2$ and $e_i(n),\ \ i = 1, 2$ are independent, when $n \to \infty$.

A4. In steady-state,

$$\mu\mathrm{E}\{e_2(n)\mathbf{u}^T(n)\widehat{\mathbf{R}}(n)\widetilde{\mathbf{w}}_1(n-1)\} \approx \frac{\mu\mathrm{Tr}(\mathbf{R})}{(1-\lambda)M}\zeta_{12}.$$

To obtain this approximation, we assume that the regressor $\mathbf{u}(n)$ follows the model proposed in [5] — since the most important characteristic of $\mathbf{u}(n)$ for adaptive filters is its autocorrelation matrix, we use a simple model that retains the correct autocorrelation, but is easy to analyze:

$$\mathbf{u}(n) = s(n)\bar{\mathbf{u}}(n),$$

where $\bar{\mathbf{u}}(n)$ may point to the direction of each of the eigenvectors of $\mathbf{R}$ with equal probability, i.e.,

$$\begin{cases} \Pr\{s(n) = \pm 1\} = 0.5, \text{ (± 1 have equal probab.)} \\ \Pr\{\bar{\mathbf{u}}(n) = \sqrt{M\lambda_i}\mathbf{b}_i\} = \frac{1}{M},\ \ i = 1 \ldots M. \end{cases}$$

Vector $\mathbf{b}_i$ is a unit-norm eigenvector of $\mathbf{R}$ with respect to eigenvalue λ_i. Since $\mathbf{R}$ is symmetric and positive-definite, it follows that $\mathbf{b}_i^T\mathbf{b}_j = \delta_{i,j}$. Variables $s(n)$ and $\bar{\mathbf{u}}(n)$ are independent from each other, and are also assumed i.i.d.. In this case, we get, using (7) and approximating $\widehat{\mathbf{R}}$ by its mean,

$$\mathbf{u}^T(n)\widehat{\mathbf{R}}(n) \approx \frac{\mathbf{u}^T(n)\mathbf{R}}{1-\lambda},$$

and the model for $\mathbf{u}(n)$ implies that $\mathbf{u}^T(n)\mathbf{R} = \lambda_i\mathbf{u}^T(n)$ with probability $1/M$. Therefore, the expectation in Assumption A4 will take the average of the eigenvalues of $\mathbf{R}$, and we obtain the desired result.

The simulations in next section confirm that the assumptions are reasonable. Considering now the random-walk model for $\mathbf{q}(n)$ and Assumption A1, we get

$$\begin{aligned} -\mathrm{E}\{\mathbf{q}^T(n)\widehat{\mathbf{R}}(n)\mathbf{q}(n)\} \approx\ & -\mathrm{E}\{e_{a,2}(n)e_1(n)\} - \\ & \mu\mathrm{E}\{e_2(n)\mathbf{u}^T(n)\widehat{\mathbf{R}}(n)\widetilde{\mathbf{w}}_1(n-1)\} + \\ & \mu\mathrm{E}\{e_1(n)e_2(n)\mathbf{u}^T(n)\mathbf{u}(n)\}. \end{aligned} \tag{16}$$

Using A2-A4, after some algebra in (16), we arrive at

$$\zeta_{12} = \frac{\mu\sigma_v^2\mathrm{Tr}(\mathbf{R}) + \dfrac{\mathrm{Tr}(\mathbf{QR})}{1-\lambda}}{1 + \mu\mathrm{Tr}(\mathbf{R})\left[\dfrac{1}{(1-\lambda)M} - 1\right]}. \tag{17}$$

In stationary environments, the expressions can be simplified, making $\mathbf{Q} = \mathbf{0}$.

4. SIMULATION RESULTS

To verify the behavior of the proposed scheme, we consider a system identification application. The initial optimal solution is formed with $M = 5$ independent random values between 0 and 1, and is given by

$$\mathbf{w}_{\mathrm{o}}^T(0) = [0.5349\ \ 0.9527\ \ -0.9620\ \ -0.0158\ \ -0.1254].$$

The regressor $\mathbf{u}(n)$ is obtained from a process $u(n)$ as $\mathbf{u}^T(n) = [u(n)\ \ u(n-1)\ \ \cdots\ \ u(n-4)]$, where $u(n)$ is generated with a first-order autoregressive model, whose transfer function is $\sqrt{1-\alpha^2}/(1-\alpha z^{-1})$. This model is fed with an i.i.d. Gaussian random process, whose variance is such that $\mathrm{Tr}(\mathbf{R}) = 1$. Moreover, additive i.i.d. noise $v(n)$ with variance $\sigma_v^2 = 0.01$ is added to form the desired signal.

Figure 2 shows the EMSE and $\mathrm{E}\{\eta(n)\}$ estimated from the ensemble-average of 400 independent runs for RLS, LMS, and their convex combination, with $\lambda = 0.95$, $\mu = 0.01$, $\mu_a = 100$, and $a^+ = 4$. At iteration $n = 35000$, matrix $\mathbf{Q}$ is changed from $\mathbf{Q} = \beta^2\mathbf{R}$ to $\mathbf{Q} = \beta^2\mathbf{R}^{-1}$, with $\beta = 0.001$. As RLS presents faster convergence than LMS, the combined scheme performs close to RLS during the first 6000 iterations. After the initial convergence, LMS presents better tracking performance than that of RLS. As predicted by the analysis, the proposed scheme performs close to LMS and $\mathrm{E}\{\eta(n)\} \approx 0$. When the matrix $\mathbf{Q}$ becomes a multiple of $\mathbf{R}^{-1}$, this behavior changes: RLS becomes superior to LMS and the combination performs slightly better than RLS, as shown in the figure and predicted by the analysis. In this case, $\mathrm{E}\{\eta(n)\} \approx 0.75$, that is, LMS plays an useful role in the combination, providing a slight improvement to the overall performance in relation to that of RLS.

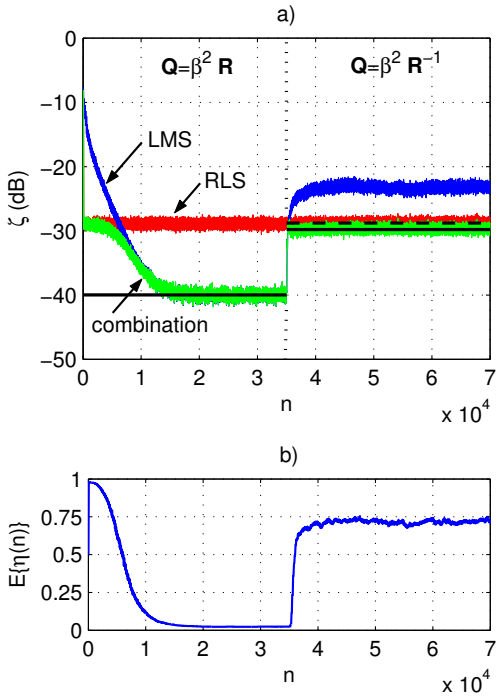

Fig. 2. a) EMSE for RLS, LMS and their convex combination and b) ensemble-average of $\eta(n)$; $\lambda = 0.95$, $\epsilon = 20$, $\mu = 0.01$, $\mu_a = 100$, $a^+ = 4$, $\alpha = 0.8$, $\beta = 0.001$; mean of 400 independent runs. In a), the solid lines represent the predicted values of ζ and the dashed line is the predicted value of ζ_1.

To verify the validity of the tracking analysis, we depict, in Figure 3, the EMSE for different values of β^2, considering the theoretical and experimental results for the convex combination and RLS, and the theoretical results for LMS. We assume $\mathbf{Q} = \beta^2\mathbf{R}^{-1}$, $\lambda = 0.92$, $\mu = 0.04$, $\mu_a = 100$, $a^+ = 4$ and the ensemble-average of 50 independent runs for the experimental curves. The simulation results are in good agreement with the analysis. This agreement also occurs for other kinds of nonstationary environments, with $\mathbf{Q} = \beta^2\mathbf{I}$ and $\mathbf{Q} = \beta^2\mathbf{R}$. We can also observe that, from $\beta^2 = 5 \times 10^{-7}$ to $\beta^2 = 10^{-5}$, the convex combination outperforms both component filters.

We should notice that the cross-EMSE, predicted by (17), is not always in a good agreement with experimental results. However, through exhaustive simulations, we observe that the EMSE of the complete scheme is not very sensitive to variations on the approximation to ζ_{12} (however, this term cannot be simply disregarded). Thus, using (17), in all our simulations, we observed good agreement between theoretical and experimental EMSE. Note also that the small disagreement between our model and the simulations observed in Figure 3 for large β^2 is due to an imprecision in the model for RLS: this can be seen by comparing the theoretical and simulation curves for RLS alone.

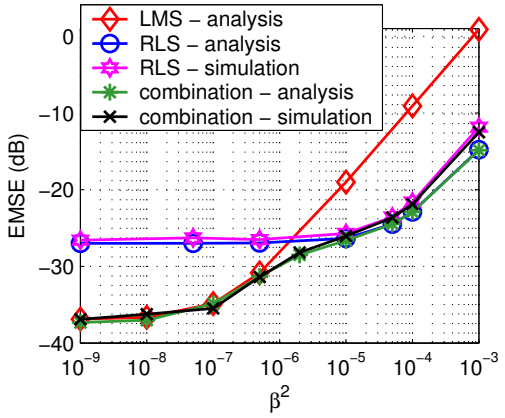

Fig. 3. EMSE for different values of β^2 considering the theoretical and experimental results for the convex combination and the theoretical results for LMS and RLS, with $\lambda = 0.92$, $\epsilon = 12.5$, $\mu = 0.04$, $\mu_a = 100$, $a^+ = 4$, $\alpha = 0.8$; mean of 50 independent runs.

5. CONCLUSIONS

In order to take advantage of the fact that RLS and LMS may outperform each other in different nonstationary scenarios, we propose a convex combination of both algorithms that performs at least as well as the best component filter. Using an energy conservation relation, the tracking analysis of the convex combination of two LMS filters was extended for the proposed scheme. Close agreement between analytical and simulation results for the EMSE of the overall scheme was observed.

6. REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*, John Wiley & Sons, NJ, 2003.

[2] E. Eweda, "Comparison of RLS, LMS and sign algorithms for tracking randomly time-varying channels," *IEEE Transactions on Signal Processing*, vol. 42, pp. 2937–2944, Nov. 1994.

[3] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Processing*, vol. 86, pp. 2430–2438, Sept. 2006.

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

[5] V. H. Nascimento, "A simple model for the effect of normalization on the convergence rate or adaptive filters," in *Proc. of ICASSP'04*. IEEE, 2004, vol. II, pp. 453–456.