INCREMENTAL COMBINATION OF RLS AND LMS ADAPTIVE FILTERS IN NONSTATIONARY SCENARIOS

Wilder B. Lopes and Cassio G. Lopes

Escola Politecnica, University of Sao Paulo
Dept. of Electronic Systems
Sao Paulo, SP - Brazil
wilder@usp.br, cassio@lps.usp.br

ABSTRACT

The incremental combination of adaptive filters (AFs), recently introduced in the literature, presents intrinsic features capable of improving the overall filtering performance. In this work, the incremental combination is extended to account for AFs with different adaptive rules; when Recursive Least-Squares (RLS) and the Least-Mean-Squares (LMS) filters are employed, it is shown, by tracking analysis and extensive simulations, that the new structure is mean-square universal in terms of the combining parameter, particularly in nonstationary scenarios with highly-correlated signals. The simulations and the analytical model match well, showing that the new algorithm outperforms its parallel-independent counterpart.

Index Terms— Adaptive filtering, incremental combination, convex combination.

1. INTRODUCTION

Combinations of AFs have been explored as design solutions to enhance the overall performance of an adaptive system. In general, the component filters present their estimates to a supervisor, responsible to generate an overall estimate at least as good as the best filter in the pool (universality) in the mean-square sense [1].

Designs based on convex and affine combining rules, AFs with different step sizes, different orders, and from different families are studied in [2–8]. In stationary scenarios, convex-parallel combinations experience a stagnation effect in the adaptation [2]. In order to circumvent this, several techniques have been proposed: the transfer of coefficients [9, 10]; the cyclic feedback of coefficients [11]; and the incremental-cooperative combination [12].

Inspired by the successful parallel combination of different AFs [4], this work extends the incremental structure proposed in [12] to comprehend a hybrid chain with different adaptive rules at the component filters. In this sense, this study shows that the hybrid incremental combination is able to achieve universality in stringent scenarios (nonstationary and highly-correlated input), outperforming the parallel arrangement, while endowing robustness to the overall combination.

2. HYBRID INCREMENTAL COMBINATIONS

This section introduces a general form for the incremental combination (hereon INC), giving rise to a hybrid chain of several filters, each one with its own adaptive rule.

The first author is supported by a scholarship from CAPES, granted by the Electrical Engineering Graduate Program of the University of Sao Paulo, Brazil. The second author is supported by a CNPq research award.

2.1. Generic Case

The general form of the INC combination for K component filters with different adaptive rules is given by

$$w_{k,i} = w_{k-1,i} + \lambda_k(i)\mu_k p_k, \tag{1}$$

in which the k^{th} component filter $w_{k,i}$ is an Mx1 vector; it updates the estimate $w_{k-1,i}$ received from the previous filter in the chain, according to the local adaptive rule $p_k = -B_k \nabla^* J(w_{k-1,i})$, with B_k any positive-definite matrix, $J(w_{k-1,i})$ the underlying cost function the filter minimizes, and * denoting the conjugate transpose [13]. In addition, $\lambda_k(i)$ is the combining parameter, and μ_k is the filter step size. As adopted in [12], the values of $\lambda_k(i)$ are subject to the convex rule $\sum_{k=1}^{K} \lambda_k(i) = 1$.

The incremental chain of K filters of the form (1) results in an overall filter whose update rule is summarized by

$$w_i = w_{i-1} + \sum_{k=1}^{K} \lambda_k(i)\mu_k p_k. \tag{2}$$

2.2. Mean-Square Filters - K=2

In this work, the focus is on combinations generated by particularizing Eq. (2) for two component filters ($K = 2$) whose cost functions minimize the mean-square error.

The data-dependent matrices $H_{k,i}$, $k = 1, 2$ are defined according to the desired adaptive rule for the component filter. Thus, the INC combination for $K = 2$ is given by

$$\begin{aligned} w_1 &= w_{i-1} + \lambda(i) H_{1,i} u_i^* [d(i) - u_i w_{i-1}] \\ w_i &= w_1 + (1 - \lambda(i)) H_{2,i} u_i^* [d(i) - u_i w_1], \end{aligned} \tag{3}$$

where u_i is a 1xM vector, $d(i)$ is the desired signal, and $[d(i) - u_i w_{i-1}]$ is the estimation error. Note that selecting $H_{1,i} = \mu_1 I$ and $H_{2,i} = \mu_2 I$, with I the MxM identity matrix, the INC combination from [12] is recovered.

3. RLS–LMS COMBINATION

The INC combination of RLS and LMS filters (RLS–LMS) is obtained by selecting $H_{1,i} = P_i$ and $H_{2,i} = \mu I$, resulting in

$$\begin{aligned} w_1 &= w_{i-1} + \lambda(i) P_i u_i^* [d(i) - u_i w_{i-1}] \\ w_i &= w_1 + (1 - \lambda(i)) \mu u_i^* [d(i) - u_i w_1], \end{aligned} \tag{4}$$

in which P_i is obtained by the recursion [13],

$$P_i = \eta^{-1} \left[P_{i-1} - \frac{\eta^{-1} P_{i-1} u_i^* u_i P_{i-1}}{1 + \eta^{-1} u_i P_{i-1} u_i^*} \right], P_{-1} = \epsilon^{-1} I, \tag{5}$$

with the regularization parameter ϵ, and the forgetting factor η.

Extensive parametric simulations in terms of the combining parameter $\lambda(i) = \lambda \in [0,1]$, conducted in a system identification configuration (order M), has shown that the INC combination in (4) can clearly present universal behavior in adverse scenarios (nonstationary plants and highly correlated input signals).

In the following example, the plant is time-varying and evolves according to the random-walk model [13]

$$w_i^o = w_{i-1}^o + q_i \,. \tag{6}$$

In the literature, q_i is generated as the realization of a zero-mean independent and identically distributed (i.i.d.) process with covariance matrix $E\boldsymbol{q}_i\boldsymbol{q}_i^* \triangleq Q = \sigma_q^2 I$. The measurement noise variance is $\sigma_v^2 = 10^{-3}$, random-walk variance $\sigma_q^2 = 10^{-4}$, and the regressors are originated by a white Gaussian process filtered by a first-order auto-regressive process with transfer function $\sqrt{1-b^2}/(1-bz^{-1})$, $b = 0.98$, and $\sigma_u^2 = 1$. The LMS step size is $\mu = \mu_o/(M\sigma_u^2)$, in which $\mu_o \in [0,1]$, and M is the system order. The RLS forgetting factor is $\eta = 0.98$. The initial state of the plant coefficients ($M = 20$) is drawn from a normalized unit variance white Gaussian process. Fig. 1 depicts the curves for both parallel and INC combinations, and their component filters.

Note how both combinations achieve universality for optimized combining parameters, namely $\lambda = 0.4$ for INC and $\lambda = 0.7$ for the parallel. The INC combination performance is clearly superior (about $4dB$ lower than the parallel in steady-state). Thus, it turns out that the RLS–LMS combination is very promising if the combining parameter λ is properly chosen.

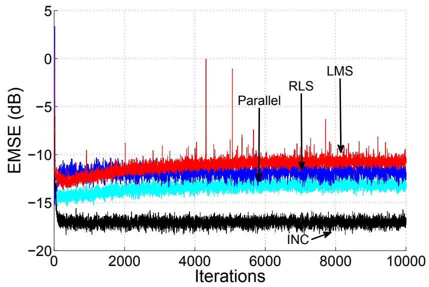


Fig. 1: Universality achieved by the combinations (INC and parallel) of RLS and LMS filters for $\mu_o = 0.3$ (500 realizations).

4. PERFORMANCE ANALYSIS - TRACKING

Motivated by the previous example, performance analysis is devised to study in which scenarios the INC combination described by (3) is able to achieve universality. The mean-square analysis is carried out in terms of the parameter $\lambda \in [0,1]$, attempting to optimize it for tracking purposes. In the sequel, Section 4.2 particularizes the analysis for the RLS–LMS case and universality in terms of mean-square error is shown. From now on, signals are modelled as stochastic processes, and random quantities are represented in boldface.

4.1. Mean-Square Filters - K=2

Eq. (3) is rewritten adopting the data-dependent matrices $\boldsymbol{H}_{k,i}$, $k = 1,2$, defined according to the desired adaptive rules. Note that now, Eq. 7 is a stochastic equation, in which the boldface terms are random quantities:

$$\begin{aligned} \boldsymbol{w}_1 &= \boldsymbol{w}_{i-1} + \lambda \boldsymbol{H}_{1,i}\boldsymbol{u}_i^*[\boldsymbol{d}(i) - \boldsymbol{u}_i\boldsymbol{w}_{i-1}] \\ \boldsymbol{w}_i &= \boldsymbol{w}_1 + (1-\lambda)\boldsymbol{H}_{2,i}\boldsymbol{u}_i^*[\boldsymbol{d}(i) - \boldsymbol{u}_i\boldsymbol{w}_1]. \end{aligned} \tag{7}$$

The goal is to derive an expression for the mean-square error (MSE) in steady-state of any filter that can be written in the form (7), via energy conservation relations (ECR) [13]. The MSE is defined as MSE $= \xi \triangleq \lim_{i\to\infty} E|\boldsymbol{e}(i)|^2$, where $\boldsymbol{e}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i\boldsymbol{w}_{i-1}$. For that matter, since adaptive filters are non-linear, time-varying, and stochastic, it is necessary to adopt a set of simplifying assumptions collected into an extended nonstationary data model [13–15]:

(1) There exits a vector $\boldsymbol{w}_i^o$ such that $\boldsymbol{d}(i) = \boldsymbol{u}_i\boldsymbol{w}_i^o + \boldsymbol{v}(i)$;
(2) The weight vector varies according to $\boldsymbol{w}_i^o = \boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i$ (random-walk model);
(3) The noise sequence $\{\boldsymbol{v}(i)\}$ is i.i.d. with constant variance $\sigma_v^2 = E|\boldsymbol{v}(i)|^2$;
(4) The noise sequence $\{\boldsymbol{v}(i)\}$ is independent of $\boldsymbol{u}_i$ for all i, j ;
(5) The sequence $\{\boldsymbol{q}_i\}$ has covariance matrix $Q \triangleq E\boldsymbol{q}_i\boldsymbol{q}_i^*$ and is independent of $\{\boldsymbol{v}(i), \boldsymbol{u}_j\}$ for all i, j ;
(6) The initial conditions $\{\boldsymbol{w}_{-1}, \boldsymbol{w}_{-1}^o\}$ are independent of all $\{\boldsymbol{d}(j), \boldsymbol{u}_j, \boldsymbol{v}(j), \boldsymbol{q}_j\}$;
(7) The regressor covariance matrix is denoted by $R_u = E\boldsymbol{u}_i^*\boldsymbol{u}_i > 0$;
(8) The random variables $\{\boldsymbol{d}(i), \boldsymbol{v}(i), \boldsymbol{u}_i, \boldsymbol{q}_i\}$ are zero mean;
(9) The weight vector $\boldsymbol{w}_i^o$ has constant mean w^o .

(8)

The ECR technique is an energy balance in terms of the following error quantities

$$\begin{cases} \tilde{\boldsymbol{w}}_{i-1} \triangleq (\boldsymbol{w}_{i-1}^o - \boldsymbol{w}_{i-1}) \text{ weight-error vector} \\ \boldsymbol{e}_a(i) = \boldsymbol{u}_i(\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}) \textit{ a priori} \text{ estimation error} \\ \boldsymbol{e}_p(i) = \boldsymbol{u}_i\tilde{\boldsymbol{w}}_i \textit{ a posteriori} \text{ estimation error} \end{cases} \tag{9}$$

together with the adaptive filter's recursion. The resulting energy equation leads to a variance relation from which the MSE and the EMSE can be derived; for details see [13].

Merging the two equations in (7) results in a simple recursion,

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \boldsymbol{H}_i\boldsymbol{u}_i^*\boldsymbol{e}(i), \tag{10}$$

in terms of the data-dependent matrix

$$\boldsymbol{H}_i \;= [\lambda\boldsymbol{H}_{1,i} + (1-\lambda)\boldsymbol{H}_{2,i}(1-\lambda\|\boldsymbol{u}_i\|^2_{\boldsymbol{H}_{1,i}})], \tag{11}$$

where $\|\boldsymbol{u}_i\|^2_{\boldsymbol{H}_i} \triangleq \boldsymbol{u}_i\boldsymbol{H}_i\boldsymbol{u}_i^*$ is the weighted norm of $\boldsymbol{u}_i$. In the general case, $\|x\|^2_\Sigma = x\Sigma x^*$.

Subtracting (10) from $\boldsymbol{w}_i^o$ gives

$$(\boldsymbol{w}_i^o - \boldsymbol{w}_i) = (\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}) - \boldsymbol{H}_i\boldsymbol{u}_i^*\boldsymbol{e}(i). \tag{12}$$

Multiplying (12) from the left by $\boldsymbol{u}_i$ results in

$$\boldsymbol{e}_p(i) = \boldsymbol{e}_a(i) - \|\boldsymbol{u}_i\|^2_{\boldsymbol{H}_i}\boldsymbol{e}(i). \tag{13}$$

Substituting (13) in (12) gives

$$(\boldsymbol{w}_i^o - \boldsymbol{w}_i) + \frac{\boldsymbol{H}_i\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{H}_i}}\boldsymbol{e}_a(i) = (\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}) + \frac{\boldsymbol{H}_i\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{H}_i}}\boldsymbol{e}_p(i)\,. \tag{14}$$

Using $\boldsymbol{H}_i^{-1}$ as weighting matrix and equating the squared weighted norms of (4.1) results in

$$\begin{aligned} &\|\boldsymbol{w}_i^o - \boldsymbol{w}_i\|^2_{\boldsymbol{H}_i^{-1}} + \overline{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \\ &\|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2_{\boldsymbol{H}_i^{-1}} + \overline{\mu}(i)|\boldsymbol{e}_p(i)|^2, \end{aligned} \tag{15}$$

where $\overline{\mu}(i) \triangleq (\|\boldsymbol{u}_i\|^2_{\boldsymbol{H}_i})^\dagger = \dfrac{1}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{H}_i}}$, if $\boldsymbol{u}_i \neq 0$ or equals zero otherwise, with † representing the pseudoinverse operator [16].

Taking the expectations of (15) gives

$$\begin{aligned} &E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{H}_i^{-1}} + E\overline{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \\ &E\|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2_{\boldsymbol{H}_i^{-1}} + E\overline{\mu}(i)|\boldsymbol{e}_p(i)|^2 \end{aligned} \tag{16}$$

Using the random-walk model into the first term of the right-hand side of (16) yields

$$E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{H}_i^{-1}} + E\overline{\mu}(i)|\boldsymbol{e}_a(i)|^2 = E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{H}_i^{-1}} + E\|\boldsymbol{q}_i\|^2_{\boldsymbol{H}_i^{-1}} + E\overline{\mu}(i)|\boldsymbol{e}_p(i)|^2 \tag{17}$$

In steady-state ($i \to \infty$), $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{H}_i^{-1}} = E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{H}_i^{-1}}$ holds. Moreover, whenever it is reasonable to assume that $\tilde{\boldsymbol{w}}_{i-1}$ is independent of $\boldsymbol{u}_i$ (for instance, small step sizes), one has

$$\begin{cases} E(\boldsymbol{H}_i^{-1}) \approx [E(\boldsymbol{H}_i)]^{-1} = H^{-1} \\ H \triangleq E\boldsymbol{H}_i. \end{cases} \tag{18}$$

Applying (18) in (17) results in the *variance relation*:

$$E\|\boldsymbol{u}_i\|^2_H|\boldsymbol{e}(i)|^2 + E\|\boldsymbol{q}_i\|^2_{H^{-1}} = 2R_e\{E\boldsymbol{e}_a^*(i)\boldsymbol{e}(i)\} \tag{19}$$

The separation principle states that in steady-state $\|\boldsymbol{u}_i\|^2_H$ is independent of $\boldsymbol{e}_a(i)$ [13]. From the data model one has

$$\boldsymbol{e}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i\boldsymbol{w}_{i-1} = \boldsymbol{e}_a(i) + \boldsymbol{v}(i) \tag{20}$$

which substituting into (19) leads to

$$2E|\boldsymbol{e}_a(i)|^2 = \sigma_v^2 E\|\boldsymbol{u}_i\|^2_H + E\|\boldsymbol{u}_i\|^2_H E|\boldsymbol{e}_a(i)|^2 + E\|\boldsymbol{q}_i\|^2_{H^{-1}}, \tag{21}$$

where $E|\boldsymbol{e}_a(i)|^2$ is the very definition of the Excess-Mean-Square Error: EMSE $= \zeta \triangleq E|\boldsymbol{e}_a(i)|^2$.

Plugging $E\|\boldsymbol{u}_i\|^2_H = Tr(R_u H)$ and $E\|\boldsymbol{q}_i\|^2_{H^{-1}} = Tr(QH^{-1})$ into (21) results in the steady-state EMSE, in nonstationary scenarios, for the INC combination described by (7):

$$\boxed{\zeta = \frac{\sigma_v^2 Tr(R_u H) + Tr(QH^{-1})}{2 - Tr(R_u H)}}. \tag{22}$$

Eq. (22) holds for filters of the form (3), and whenever assumption (18) is reasonable.

The MSE is obtained as follows (see Eq. (20)),

$$\xi = \zeta + \sigma_v^2. \tag{23}$$

Eq. (22) holds under the data model (8), and requires the calculation of the data moments R_u and H (which is filter dependent).

4.2. RLS–LMS case

For the RLS–LMS[1] combination (refer to (7)), $\boldsymbol{H}_{1,i} = \boldsymbol{P}_i$ (the stochastic version of Eq. 5) and $\boldsymbol{H}_{2,i} = \mu I$, with I the MxM identity matrix, and $\mu = \mu_o/(M\sigma_u^2)$, $\mu_o \in [0,1]$. Thus,

$$\boldsymbol{H}_i = [\lambda\boldsymbol{P}_i + \mu(1-\lambda)(1 - \lambda\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i})I]. \tag{24}$$

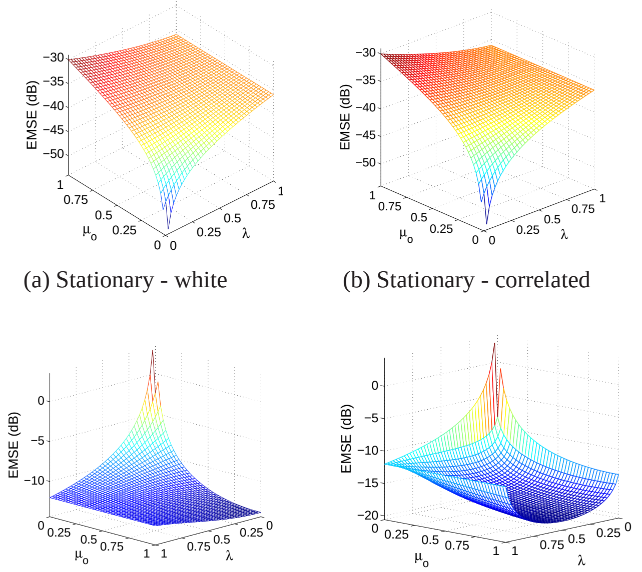
To calculate the EMSE and MSE of the structure, one needs to determine $E\boldsymbol{H}_i$ from (24). This is accomplished by particularizing (18) for the RLS–LMS case (see [13] p.288),

$$\begin{cases} \lim\limits_{i\to\infty} E(\boldsymbol{P}_i^{-1}) = \dfrac{R_u}{1-\eta} \triangleq P^{-1} \\ E\boldsymbol{P}_i \approx [E(\boldsymbol{P}_i^{-1})]^{-1} = (1-\eta)R_u^{-1} = P \\ \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}i} \text{ is independent of } \boldsymbol{e}_a(i) \\ E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}i} \approx E\|\boldsymbol{u}_i\|^2_P = Tr(R_u P) = (1-\eta)M, \end{cases} \tag{25}$$

resulting in

$$H = \{\lambda(1-\eta)R_u^{-1} + \mu(1-\lambda)[1 - \lambda(1-\eta)M]I\}. \tag{26}$$

Substituting (26) into (22), together with (23) returns the EMSE (MSE) of the RLS–LMS incremental structure. It is a function of several parameters. Here η and M are fixed, and for a given R_u and Q the ζ behavior is explored in terms of μ_o and λ. Note that for this case, an expression for an optimal combining parameter λ^o can be derived by making $\lambda^o = \{\lambda|(\partial\zeta/\partial\lambda) = 0\}$, which is not shown here due to space constraints. In any event, it is a guideline



(a) Stationary - white
(b) Stationary - correlated
(c) Nonstationary - white
(d) Nonstationary - correlated

Fig. 2: RLS–LMS - EMSE theoretical surfaces in terms of μ_o and λ in both stationary and nonstationary scenarios, with white and correlated input data.

for adaptive designs ($\lambda(i) \to \lambda^o$). Such adaptive implementations will be pursued in forthcoming publications.

Fig. 2 depicts an example of the EMSE surface for RLS–LMS in four cases, varying μ_o and λ, in both stationary and nonstationary scenarios. Regressors are either white Gaussian or highly correlated, obtained from a first-order auto-regressive process with transfer function $\sqrt{1-b^2}/(1-bz^{-1})$, $b = 0.98$. The signals variances are $\sigma_u^2 = 1$, $\sigma_v^2 = 10^{-3}$, and $\sigma_q^2 = 10^{-4}$ for the random walk[2]. The RLS forgetting factor is $\eta = 0.98$, regularization factor $\epsilon = 10^{-5}$ and the system order is $M = 20$. Figs. 2 (a), (b) and (c) do not show improvement from the combination. On the other hand, it is noticeable from the convex shape of the correlated data surface (d) that there are optimum pairs (μ_o, λ) that attain the minimum EMSE. This shows that the structure is universal and considerably outperforms the component filters in nonstationary scenario with highly correlated input.

5. SIMULATIONS

In the next two examples, simulations are presented comparing the parallel with the INC structure. The EMSE curves are generated for $\mu_o = \{0.2, 0.9\}$ and for $\lambda \in [0,1]$. The nonstationary plant is initialized with w_{-1}^o drawn from a normalized unit variance white Gaussian process, and follows the random-walk model with $\sigma_q^2 = 10^{-4}$. Regressors are correlated and generated according to Section 3, namely $\sigma_u^2 = 1$, $\sigma_v^2 = 10^{-3}$, $\eta = 0.98$, $\epsilon = 10^{-5}$.

The first example is run with $\mu_o = 0.2$. Fig. 3 (a) is an abacus comparing in steady-state the parallel and the INC structure (theory and simulations). The simulated curves were generated as the average of the last 300 estimates after convergence. INC combination clearly outperforms the parallel as well as its component filters. This figure can be regarded as a slice of Fig. 2 (d). Fig. 3 (b) presents the learning curves for $\lambda = 0.4$, with an improvement of nearly $4dB$

[1] Analysis for the LMS–RLS case is obtained by swapping $H_{1,i}$ and $H_{2,i}$.

[2] Note how high the nonstationary degree is compared to the typical literature range $[10^{-6}, 10^{-8}]$

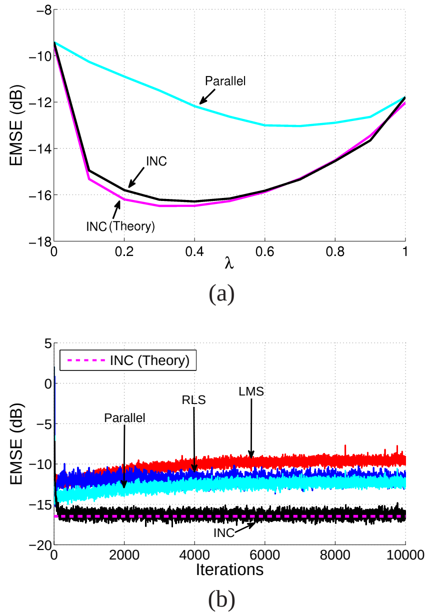


Fig. 3: (a) EMSE versus λ for both combinations with $\mu_o = 0.2$ and $\sigma_q^2 = 10^{-4}$. (b) EMSE of the individual filters and the combinations for $\mu_o = 0.2$ and $\lambda = 0.4$ (500 realizations).

over the parallel combination. Note how the theoretical and the simulated curves match.

In the same vein, Fig. 4 presents the same scenario, only with the step size increased to $\mu_o = 0.9$, while keeping $\lambda = 0.6$. The abacus has not been depicted as it is nearly meaningless: due to the highly correlated input signal, the LMS component filter experiences severe spikes (outliers), and so does the parallel as it can not fully combat this phenomenon. Steady-state performance for the INC case achieves $7dB$ lower than the best component filter (RLS). Although the RLS can converge, the LMS is unstable, driving the parallel combination with fixed λ (CVX_1) into divergence. The parallel combination with adaptive λ (CVX_2: $\lambda \rightarrow \lambda(i)$) [2, 4] was included to illustrate that, even with an update rule, in this stringent scenario the parallel combination is not able to cope with LMS instability: the best it may do is to track the RLS performance, by setting $\lambda = 1$. Once again, the simulated curve corroborates the theory.

Fig. 4 also illustrates the stabilization effect provided by the incremental nature of the structure. Even with a poorly designed step size for the LMS (in the example, $\mu_o = 0.9$), the INC combination is able to circumvent this issue, making the final EMSE curve converge. In this situation, the only way the INC can diverge is setting λ very close or equal to 0, turning the RLS component into a relay filter, and consequently the combination becomes the unstable LMS filter.

Fig. 5 presents an abacus showing the universal behavior of the INC combination for different degrees of non-stationarity (σ_q^2) with $\mu_o = 0.2$. Note that an improvement of about $5dB$ of the combination over the component filters remains over a wide range of σ_q^2.

6. CONCLUSION

By extending the INC combination of AFs [12] to account for different adaptive rules, this work resorts to theoretical analysis and

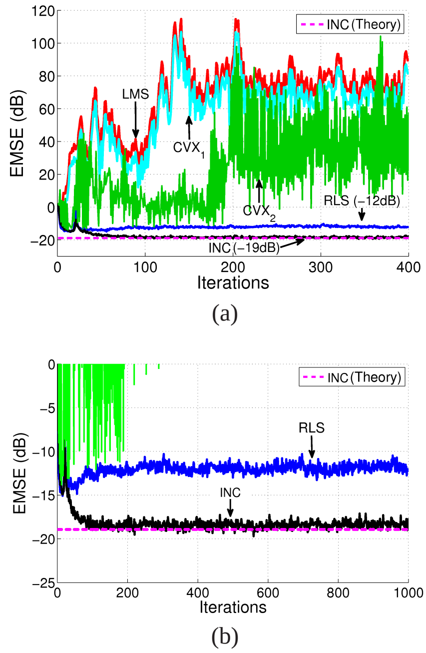


Fig. 4: (a) EMSE of the individual filters and the combinations for $\mu_o = 0.9$, $\lambda = 0.6$ and $\sigma_q^2 = 10^{-4}$ (500 realizations). (b) A zoomed-in view of the convergent curves.

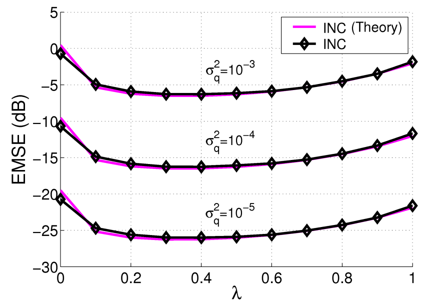


Fig. 5: Abacus for RLS–LMS combination with $\mu_o = 0.2$ and different values of σ_q^2.

parametric simulations (in terms of λ) to show that the RLS–LMS incremental combination is able to achieve universality in stringent scenarios (nonstationary and highly correlated input). Moreover, the INC scheme provides an stabilization effect, improving the combination robustness and allowing it to outperform the parallel [2,4] in the same conditions. In this way, the INC structure is a good candidate when the combination runs in challenging scenarios. The analysis framework presented may be easily employed for different component filters (other than RLS and LMS) whose adaptive rules can be properly described by the filter-dependent matrices $H_{k,i}$.

Future work considers the procedure to obtain an optimal λ via $(\partial\zeta/\partial\lambda) = 0$, what may provide a guideline for adapting $\lambda(i)$. Also, the study of the INC combination removing the constraint $\sum_{k=1}^{K} \lambda_k(i) = 1$ will be pursued.

7. REFERENCES

[1] N. Merhav and M. Feder, "Universal prediction," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2124 –2147, oct 1998.

[2] J. Arenas-Garcia, A.R. Figueiras-Vidal, and A.H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1078–1090, 2006.

[3] N.J. Bershad, J.C.M. Bermudez, and J.Y. Tourneret, "An affine combination of two LMS adaptive filters – Transient mean-square analysis," *Signal Processing, IEEE Transactions on*, vol. 56, no. 5, pp. 1853–1864, 2008.

[4] M.T.M. Silva and V.H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 3137–3149, 2008.

[5] C.G. Lopes, E.H. Satorius, and A.H. Sayed, "Adaptive carrier tracking for direct-to-earth Mars communications," in *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on*. IEEE, 2007, pp. 1042–1046.

[6] C.G. Lopes, E.H. Satorius, P. Estabrook, and A.H. Sayed, "Efficient Adaptive Carrier Tracking for Mars to Earth Communications During Entry, Descent and Landing," in *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*. IEEE, 2007, pp. 517–521.

[7] C.G. Lopes, E.H. Satorius, P. Estabrook, and A.H. Sayed, "Adaptive Carrier Tracking for Mars to Earth Communications During Entry, Descent, and Landing," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 4, pp. 1865–1879, 2010.

[8] Yonggang Zhang and J.A. Chambers, "Convex combination of adaptive filters for a variable tap-length lms algorithm," *Signal Processing Letters, IEEE*, vol. 13, no. 10, pp. 628 –631, oct. 2006.

[9] M.T.M. Silva, V.H. Nascimento, and J. Arenas-Garcia, "A transient analysis for the convex combination of two adaptive filters with transfer of coefficients," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3842–3845.

[10] V.H. Nascimento and R.C. de Lamare, "A low-complexity strategy for speeding up the convergence of convex combinations of adaptive filters," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 3553 –3556.

[11] L.F.O. Chamon, W.B. Lopes, and C.G. Lopes, "Combination of adaptive filters with coefficients feedback," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 3785 –3788.

[12] W.B. Lopes and C.G. Lopes, "Incremental-cooperative strategies in combination of adaptive filters," in *Int. Conf. in Acoust. Speech and Signal Process.* IEEE, 2011, pp. 4132 –4135.

[13] A.H. Sayed, *Adaptive filters*, Wiley-IEEE Press, 2008.

[14] Paulo Sergio Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Kluwer Academic Publishers, Norwell, MA, USA, 2 edition, 2002.

[15] S.S. Haykin, *Adaptive filter theory*, Prentice-hall Englewood Cliffs, NJ, 1986.

[16] Alan J. Laub, *Matrix analysis - for scientists and engineers.*, SIAM, 2005.