TRANSIENT PERFORMANCE OF AN INCREMENTAL COMBINATION OF LMS FILTERS

Luiz F. O. Chamon and Cássio G. Lopes

Signal Processing Laboratory
Dept. of Electronic Systems Engineering, University of São Paulo – Brazil
chamon@usp.br, cassio@lps.usp.br

ABSTRACT

Incremental combinations were first introduced as a solution to the convergence stagnation issue of parallel-independent combinations, both convex and affine. Since then, this topology has been shown to enhance performance to the point of allowing a combination of LMS filters to outperform the APA with lower complexity. In order to better understand and improve this structure, the present work develops mean and mean square transient models for the incremental combination of two LMS filters, that is shown to be a generalization of the data reuse LMS (DR-LMS). By formulating the optimal supervisor and deriving its constraints, the previously proposed adaptive combiner is redesigned to improve the combination's overall performance, as shown by simulations and analysis.

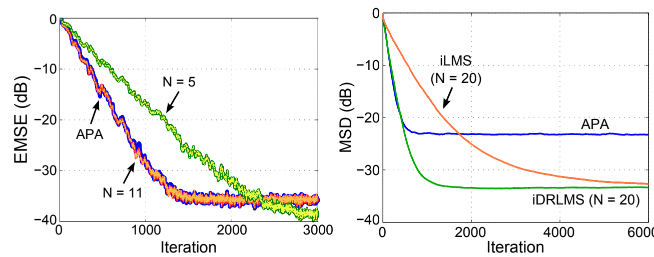
***Index Terms*—** Adaptive filtering, Combination of adaptive filters, Incremental combination, Transient analysis

1. INTRODUCTION

Combination of adaptive filters (AFs) is considered an established approach to improve performance when the accurate design of a single filter is difficult [1–4]. It consists of mixing a set of AFs—the *components*—by means of a supervisor whose task is to achieve universality, in the sense that the overall system is at least as good—normally relative to the mean-square error (MSE)—as the best filter in the set.

Usually, the component filters run independently while adaptive parameters merge their coefficients. This structure has been extensively investigated using different step sizes, orders, adaptive algorithms, and supervising rules [3–6]. However effective, this combination presents a well-known convergence stagnation due to the accurate filter's delay in reaching the MSE of the fast one. This issue was initially addressed using conditional unidirectional transfers of coefficients [6], but it was not until recently that more effective solutions have been found through changes in topology [1, 2, 7].

Among these solutions, the incremental structure put forward in [1] has been shown to have properties beyond addressing the stagnation effect, specially in relation to convergence performance (Fig. 1). Fig. 1a compares a rank K Affine



(a) White input signal (b) Correlated input signal

Fig. 1. Comparing APA, iDRLMS, and iLMS ($K = 10$)

Projection Algorithm (APA) [8] to the incremental combinations of N LMS filter with (iDR-LMS) and without (iLMS) data reuse (DR)—see Section 2. It is clear that these combinations can match the performance of APA with lower complexity [9]. Furthermore, even though the iLMS behavior is similar to that of the iDR-LMS for white signals, the use of DR is valuable for correlated signals (Fig. 1b).

So as to better understand and design this combination, the present work analyzes different aspects of its performance by (*i*) showing that this combination is, in fact, a generalization of the data reuse LMS (DR-LMS) [10, 11]; (*ii*) modeling the mean and mean-square transient performance of an incremental combination of two LMS filters (iLMS); (*iii*) formulating the optimal supervisor and showing that its unbiasedness constraint is different from that of the parallel topology; (*iv*) redesigning the adaptive supervisor from [1] according to these results; and (*v*) using simulations to show the accuracy of the models and illustrate the aforementioned contributions.

2. COMBINATION OF ADAPTIVE FILTERS

In a system identification scenario, the n-th component of a combination of N AFs can be described as

$$w_{n,i} = w_{n,i-1} + \mu_n p_n, \tag{1}$$

where $w_{n,i}$ is the $M \times 1$ coefficient vector at iteration i, μ_n is a step size, and $p_n = -B_n \nabla^* J_n(w_{n,i-1})$ is an update direction vector, with B_n a positive definite matrix, $J_n(w_{n,i-1})$ the underlying cost function the component attempts to minimize, and * denoting the conjugate transpose operation [8, 9].

Usually, $J_n(\cdot)$ involves a function of the n-th component output estimation error $e_n(i) = d_n(i) - u_{n,i} w_{n,i-1}$, with $u_{n,i}$ selected from a sequence $\{u_i\}$ and $d_n(i)$ taken from the measurements

$$d(i) = u_i w^o + v(i), \tag{2}$$

where w^o is an $M \times 1$ vector that models the unknown system, u_i is a $1 \times M$ regressor vectors that captures samples $u(i)$ of an input signal with variance σ_u^2, and $v(i)$ is a zero-mean white Gaussian measurement noise with variance σ_v^2. An example is the usual MSE cost function $J_n(w_{n,i-1}) = \mathrm{E}\,|e_n(i)|^2$, adopted in the celebrated LMS filter [8]

$$w_{n,i} = w_{n,i-1} + \mu_n u_{n,i}^* e_n(i). \tag{3}$$

A myriad of data distribution structures can be accounted for using different definitions for $u_{n,i}$ and $d_n(i)$, including the DR method $\{u_{n,i}, d_n(i)\} = \{u_{i-n+1}, d(i-n+1)\}$ of the iDR-LMS [9]. This work will address the more common *data sharing* case $\{u_{n,i}, d_n(i)\} = \{u_i, d(i)\}$.

Finally, the output of the combination at iteration i is given by the global coefficients $w_i = f(\eta_n(i), w_{n,i})$, where $\eta_n(i)$ represents the action of the supervisor on component n. Usually, the supervisor adapts $\eta_n(i)$ attempting to minimize the global MSE: $\mathrm{E}\,|e(i)|^2$, where $e(i) = d(i) - u_i w_{i-1}$ is the global error. Using this terminology, data sharing parallel and incremental combinations are written as

Definition 1 (*Parallel combination* [2–4]).

$$\begin{aligned} w_{n,i-1} &= \delta(i-rL)\, w_{i-1} + (1-\delta(i-rL))\, w_{n,i-1} \\ w_{n,i} &= w_{n,i-1} + \mu_n u_i^* \left[d(i) - u_i\, w_{n,i-1}\right] \\ w_i &= \sum_{n=1}^{N} \eta_n(i)\, w_{n,i} \end{aligned} \tag{4}$$

where $\delta(\cdot)$ is Kronecker's delta, L is a cycle length, and $r \in \mathbb{N}$. The first equation accounts for the cyclic feedback strategy introduced in [2]. For $L \to \infty$ the parallel-independent combination is recovered [3, 4].

Definition 2 (*Incremental combination* [1, 9]).

$$\begin{aligned} w_{0,i} &= w_{i-1} \\ w_{n,i} &= w_{n-1,i} + \eta_n(i)\mu_n u_i^* [d(i) - u_i\, w_{n-1,i}] \\ w_i &= w_{N,i}. \end{aligned} \tag{5}$$

Note that for $\eta_n(i) = 1$ and $\mu_n = \mu$ in (5), the DR-LMS algorithm from [10] is recovered. The analysis hereafter is, therefore, also applicable to this AF.

3. MEAN PERFORMANCE

Due to space constraints, derivations will be carried on for $N = 2$, i.e., for the original combination in [1], assuming $u(i)$ arises from a real zero-mean i.i.d. process. Therefore, (5) simplifies to

$$\begin{aligned} w_{1,i} &= w_{i-1} + \eta_1(i)\mu_1 u_i^T [d(i) - u_i\, w_{i-1}] \\ w_i &= w_{1,i} + \eta_2(i)\mu_2 u_i^T [d(i) - u_i\, w_{1,i}]. \end{aligned} \tag{6}$$

Combining the equations in (6), subtracting from w^o, and employing the data model in (2) yields

$$\widetilde{w}_i = \widetilde{w}_{i-1} - \left[\bar{\mu}(i) - \mu'(i)\|u_i\|^2\right] u_i^T e(i). \tag{7}$$

where $\widetilde{w}_i = w^o - w_i$ is the global coefficients error, $\bar{\mu}(i) = \eta_1(i)\mu_1 + \eta_2(i)\mu_2$, and $\mu'(i) = \eta_1(i)\eta_2(i)\mu_1\mu_2$. Despite the algebraic similarity of (7) to variable step size (VSS) algorithms[1], it is conceptually motivated by incremental strategies [12]. In a way, $\eta_n(i)$ plays the dual role of supervisor and VSS in incremental combinations.

Before proceeding, the following assumptions are adopted so as to make the analysis tractable:

A.1 (*Data independence assumptions*) $\{u_i\}$ constitutes an i.i.d. sequence of vectors independent of $v(j)$, $\forall\, i, j$. Consequently, $\{u_i, \widetilde{w}_j\}$, $\{d(i), d(j)\}$, and $\{u_i, d(j)\}$ are independent for $i > j$.

A.2 (*Supervisor separation principle*) Due to $\eta_n(i)$ converging slowly compared to variations in u_i—and, consequently, compared to variation of the *a priori* error $e_a(i) = u_i \widetilde{w}_{i-1}$—, these variables are separable as in $\mathrm{E}[\eta_n(i) u_i] = \mathrm{E}\,\eta_n(i)\,\mathrm{E}\,u_i$ and $\mathrm{E}[\eta_n(i) e_a(i)] = \mathrm{E}\,\eta_n(i)\,\mathrm{E}\,e_a(i)$.

The assumptions in A.1 are widely adopted in the adaptive filtering literature. In some applications they accurately describe the behavior of the AF's inputs, whereas in others they are taken as approximations [8]. A.2 was inspired by the separation principle from [8] and has been employed in the analysis of both convex and affine combinations [13–15].

Now, taking the expected value of (7) and using A.1 and A.2 yields

$$\mathrm{E}\,\widetilde{w}_i = [1 - \mathrm{E}\,\bar{\mu}(i)\sigma_u^2 + \mathrm{E}\,\mu'(i)(M+2)\sigma_u^4]\,\mathrm{E}\,\widetilde{w}_{i-1}, \tag{8}$$

where $\mathrm{E}\,u_i^T u_i = \sigma_u^2 I$, $\mathrm{E}\,u_i^T u_i u_i^T u_i = (M+2)\sigma_u^4 I$, and I is the identity matrix [8]. A particular case of this result for the DR-LMS can be found in [11].

Comparing (8) to the model for a single LMS with step size μ, explicitly [8]

$$\mathrm{E}\,\widetilde{w}_i = [1 - \mu\sigma_u^2]\,\mathrm{E}\,\widetilde{w}_{i-1},$$

it is clear that, even though part of the processing in (8) takes place as with an LMS filter with step size $\mathrm{E}\,\bar{\mu}(i)$, a higher order term—$\mathrm{E}\,\mu'(i)(M+2)\sigma_u^4$—appears due to the non-linear characteristics of the adaptive algorithms involved [8].

[1]A similar interplay occurs with the NLMS algorithm [8].

4. MEAN-SQUARE PERFORMANCE

The mean-square performance of AFs is usually measured by the Mean Square Deviation (MSD) and Excess MSE (EMSE)

$$\begin{aligned} \text{MSD}(i) &= \mathrm{E}\,\|\widetilde{w}_{i-1}\|^2 = \mathrm{Tr}(K_i) \\ \text{EMSE}(i) &= \mathrm{E}\,\|e_a(i)\|^2 = \mathrm{Tr}(R_u K_i) = \sigma_u^2 \text{MSD}(i), \end{aligned} \tag{9}$$

where $K_i = \mathrm{E}\,\widetilde{w}_{i-1}\widetilde{w}_{i-1}^T$. In the sequel, a recursion for K_i is derived and closed form models for (9) are deduced, first assuming only that u_i is a Gaussian vector and then simplifying the resulting expression for small step sizes. The time index of the supervisor parameters is omitted for clarity's sake.

Starting from (7),

$$\begin{aligned} K_{i+1} = \mathrm{E}\,\widetilde{w}_i\widetilde{w}_i^T &= K_i \\ &- \mathrm{E}\left[\bar{\mu} - \mu'\|u_i\|^2\right]\left[\widetilde{w}_{i-1}e^T(i)u_i + u_i^T e(i)\widetilde{w}_{i-1}^T\right] \\ &+ \mathrm{E}\left[\bar{\mu} - \mu'\|u_i\|^2\right]^2 u_i^T u_i |e(i)|^2. \end{aligned} \tag{10}$$

Using the data model (2), $e(i) = e_a(i) + v(i)$, and since all terms linearly dependent on $v(i)$ vanish due to A.1 and A.2,

$$\begin{aligned} K_{i+1} = {} & \\ & K_i - \mathrm{E}\,\bar{\mu}\left[\widetilde{w}_{i-1}\widetilde{w}_{i-1}^T u_i^T u_i + u_i^T u_i \widetilde{w}_{i-1}\widetilde{w}_{i-1}^T\right] \\ & + \mathrm{E}\,\mu'\left[\widetilde{w}_{i-1}\widetilde{w}_{i-1}^T u_i^T u_i u_i^T u_i + u_i^T u_i u_i^T u_i \widetilde{w}_{i-1}\widetilde{w}_{i-1}^T\right] \\ & + \mathrm{E}\,\bar{\mu}^2 u_i^T u_i \widetilde{w}_{i-1}\widetilde{w}_{i-1}^T u_i^T u_i \\ & - 2\,\mathrm{E}\,\bar{\mu}\mu' u_i^T u_i u_i^T u_i \widetilde{w}_{i-1}\widetilde{w}_{i-1}^T u_i^T u_i \\ & + \mathrm{E}\,|\mu'|^2 u_i^T u_i u_i^T u_i \widetilde{w}_{i-1}\widetilde{w}_{i-1}^T u_i^T u_i u_i^T u_i \\ & + \mathrm{E}\left[\bar{\mu}^2 u_i^T u_i - 2\bar{\mu}\mu' u_i^T u_i u_i^T u_i + |\mu'|^2 u_i^T u_i u_i^T u_i u_i^T u_i\right]\sigma_v^2. \end{aligned} \tag{11}$$

Even though the assumptions from Section 3 will render (11) more tractable by separating the expectations, higher order moments of u_i must yet be calculated. Assuming that u_i is a Gaussian vector, these moments can be evaluated in closed form, so that under the initial i.i.d. input assumption—i.e., $\mathrm{E}\,u_i^T u_i = \sigma_u^2 I$—one gets

$$\begin{aligned} \mathrm{E}\,u_i^T u_i W u_i^T u_i &= \sigma_u^4[\mathrm{Tr}(W)I + 2W] \\ \mathrm{E}\,u_i^T u_i u_i^T u_i W u_i^T u_i &= (M+4)\sigma_u^6[\mathrm{Tr}(W)I + 2W] \\ \mathrm{E}\,u_i^T u_i u_i^T u_i W u_i^T u_i u_i^T u_i &= (M+4)(M+6)\sigma_u^8[\mathrm{Tr}(W)I + 2W], \end{aligned}$$

where W is some deterministic symmetric matrix. These expressions can be obtained by direct evaluation of the vector products. A detailed derivation of the fourth order moment can be found in [8].

Under A.1 and A.2, (11) becomes

$$\begin{aligned} K_{i+1} &= \alpha\,K_i + \beta\,\mathrm{Tr}(K_i)I + \gamma\,\sigma_u^2\sigma_v^2 I \\ \alpha &= 1 - 2\,\mathrm{E}\,\bar{\mu}\sigma_u^2 + 2(M+2)\,\mathrm{E}\,\mu'\sigma_u^4 + 2\beta \\ \beta &= \mathrm{E}\,\bar{\mu}^2\sigma_u^4 - 2(M+4)\,\mathrm{E}\,\bar{\mu}\mu'\sigma_u^6 \\ &\quad + (M+4)(M+6)\,\mathrm{E}\,|\mu'|^2\sigma_u^8 \\ \gamma &= \mathrm{E}\,\bar{\mu}^2 - 2\,\mathrm{E}\,\bar{\mu}\mu'(M+2)\sigma_u^2 \\ &\quad + \mathrm{E}\,|\mu'|^2(M+2)(M+4)\sigma_u^4, \end{aligned}$$

whose trace leads to the MSD recursion

$$\begin{aligned} \text{MSD}(i+1) &= A \cdot \text{MSD}(i) + b \cdot \mathrm{Tr}(R_u)\sigma_v^2 \\ A &= 1 - 2\,\mathrm{E}\,\bar{\mu}\sigma_u^2 \\ &\quad + 2(M+2)\,\mathrm{E}\,\mu'\sigma_u^4 + (M+2)\beta \\ b &= \mathrm{E}\,\bar{\mu}^2 - 2\,\mathrm{E}\,\bar{\mu}\mu'(M+2)\sigma_u^2 \\ &\quad + \mathrm{E}\,|\mu'|^2(M+2)(M+4)\sigma_u^4, \end{aligned} \tag{12}$$

with $\text{MSD}(0) = \|w^o\|^2$ assuming $w_{-1} = 0$.

This expression can be further simplified assuming small step sizes, so that higher order powers of μ_n can be neglected. Formally, the assumption adopted is that $[\mathrm{Tr}(R_u)\mu_1]^\ell \to 0$, $\forall\,\ell > 2$, where $\mu_1 > \mu_2$ without loss of generality. As a result, A and b in (12) can be replaced by

$$\begin{aligned} A' &= 1 - 2\,\mathrm{E}\,\bar{\mu}\sigma_u^2 + (M+2)(\mathrm{E}\,\bar{\mu}^2 + 2\,\mathrm{E}\,\mu')\sigma_u^4 \\ b' &= \mathrm{E}\,\bar{\mu}^2. \end{aligned} \tag{13}$$

The accuracy of (13)—as shown in Section 6—suggests that the small step sizes assumption yields negligible errors, making this recursion valid over a wide range of step sizes.

5. SUPERVISOR ANALYSIS

The design of an effective adaptive supervisor for the incremental combination is still an open issue, although a simple solution based on error filtering was introduced in [1] for stationary scenarios. This section will employ the models obtained so far to derive general properties of the incremental supervisor and improve the existing design.

5.1. Optimal supervisor

The pair $\{\eta_1(i), \eta_2(i)\}$ that results in the highest local convergence rate is determined by solving

$$\nabla_{\eta_1,\eta_2}\,\text{MSD}(i+1) = 0, \tag{14}$$

since $\text{MSD}(i+1)$ is convex for a given $\text{MSD}(i)$ under the adopted assumptions.

For the previously derived MSD models, (14) becomes

$$\begin{aligned} \frac{\partial}{\partial\eta_1}\text{MSD}(i+1) &= 2\sigma_u^2\mu_1[A_2 \cdot \text{MSD}(i) + b_2 \cdot M\sigma_v^2] = 0 \\ \frac{\partial}{\partial\eta_2}\text{MSD}(i+1) &= 2\sigma_u^2\mu_2[A_1 \cdot \text{MSD}(i) + b_1 \cdot M\sigma_v^2] = 0, \end{aligned} \tag{15}$$

with

$$A_n = (M+2)\sigma_u^2 \Big\{\mu'(M+4)\sigma_u^2 \left[\eta_n\mu_n(M+6)\sigma_u^2 - 1\right] + \eta_n\mu_n \left[1 - \bar{\mu}(M+4)\sigma_u^2\right] + \bar{\mu}\Big\} - 1$$

$$b_n = (M+2)\sigma_u^2 \Big\{\eta_n\mu_n \left[\mu'(M+4)\sigma_u^2 - \bar{\mu}\right] - \mu'\Big\} + \bar{\mu}$$

for the Gaussian regressor model (12) and $A_n = (\bar{\mu} + \eta_n\mu_n)(M+2)\sigma_u^2 - 1$ and $b_n = \bar{\mu}$ for the small step sizes model (13). Note that the equations in (15) can only be simultaneously satisfied for $\eta_1\mu_1 = \eta_2\mu_2$. Since the closed form expression for $\eta_n(i)$ in the Gaussian regressor case is intricate and the small step sizes approximation leads to good results—see Section 6—, the former will be omitted.

Hence, solving (15) in the small step sizes case leads to

$$\eta_n^o(i) = \frac{1}{\mu_n}\frac{\text{MSD}(i)}{3(M+2)\sigma_u^2\,\text{MSD}(i) + 2M\sigma_v^2},\ n = 1, 2. \tag{16}$$

Note that (16) only serves analytical purposes and cannot be evaluated exactly in practice since it requires w^o to be known.

5.2. Supervisor constraint

In combination of AFs, the supervisor must be constrained to guarantee the unbiasedness of the overall estimator. In parallel combinations, such as convex and affine combinations, this constraint consists of $\sum \eta_n(i) = 1$ [3, 4]. In the sequel, this restriction is shown not to apply to incremental combinations, so that the supervisor proposed in [1], which adheres to this constraint, can be readily improved.

In order to guarantee convergence in the mean, i.e., $\lim_{i\to\infty} \mathrm{E}\,\widetilde{w}_i = 0$, the recursion (8) must meet the condition $|1 - \mathrm{E}\,\bar{\mu}(i)\sigma_u^2 + \mathrm{E}\,\mu'(i)(M+2)\sigma_u^4| < 1$ [8].

From Section 5.1, the optimal supervisor has $\eta_1(i)\mu_1 = \eta_2(i)\mu_2$, so that the supervisor constraint simplifies to

$$0 < \mathrm{E}\,\eta_n(i)\mu_n < \frac{2}{(M+2)\sigma_u^2}, \quad n = 1, 2. \tag{17}$$

Notice that, from (17), it is possible to have $\sum \eta_n(i) > 1$.

5.3. Adaptive supervisor

In lieu of the results from the previous sections, the supervisor design proposed in [1] can be readily improved by lifting the convex constraint and using $\eta_1(i) = \eta(i)/\mu_1$ and $\eta_2(i) = \eta(i)/\mu_2$. The supervising parameter $\eta(i)$ can then be adapted in a deterministic manner using, e.g.,

$$\eta(i) = \frac{1}{1 + e^{s(i-\ell)}}, \tag{18}$$

or using a filtered error approach

$$\begin{aligned} p(i) &= \alpha\, p(i-1) + (1-\alpha)\,|e(i)|^2 \\ \eta(i) &= f[p(i)], \end{aligned} \tag{19}$$

where $0 < \alpha < 1$ and $f(x)$ is a linear activation function, instead of the sigmoidal function employed in [1]. $\eta(i)$ must be constrained as per (17) to guarantee unbiasedness. This, however, does not guarantee mean-square stability. Hence, a more appropriate way to limit the supervising parameter is by means of the optimal supervisor (16) value when far from w^o:

$$\eta(i) \leq \lim_{\text{MSD}(i)\to\infty} \mu_n\eta_n^o(i) = \frac{1}{3(M+2)\sigma_u^2} \tag{20}$$

6. SIMULATIONS

The simulations in this section were conducted using $\sigma_u^2 = 1$, $\sigma_v^2 = 10^{-3}$, $w^o = \text{col}\{1\}/\sqrt{M}$, and $M = 10$. For the small step sizes simulations, $\mu_1 = 0.005$ and $\mu_2 = 0.003$. Everywhere else, $\mu_1 = 0.05$ and $\mu_2 = 0.005$. All curves were averaged over 200 independent realizations.

Fig. 2 and Fig. 3 show the transient models (dashed curves) overlaid on the numerical experiments. Note that both models are equally accurate, which confirms the claim that errors due to neglecting higher order terms in (13) are small, so that it is valid over a wide range of step sizes.

Fig. 2a uses the deterministic supervisor in (18) with a convex constraint ($\eta_1(i) = \eta(i)$, $\eta_2(i) = 1 - \eta(i)$, $s = 0.05$, and $\ell = 1100$) and without ($\eta_1(i) = \eta(i)$, $\eta_2(i) = 1$, $s = 0.05$, and $\ell = 700$) to illustrate the effect of lifting such restriction. Note that, while the convexly constrained combination is at most as fast as the fastest component (μ_1), the unconstrained supervisor is able to outperform it. This effect is enhanced by the small step sizes, so that it is not evident in Fig. 2b. In this case, the convexly constrained and unconstrained supervisor curves are indistinguishable.

Fig. 3 shows the behavior of the combination using the optimal supervisor (16). In this case, the components' step sizes have no impact on the performance of the combination since $\eta_n^o(i)$ is "normalized" by their values—see (16). Note that the incremental combination is able to improve both transient and steady-state performance of the combination. This is a result of the supervisor–VSS duality of η_n.

Last, Fig. 4 compares the performance of different combinations with adaptive supervisors in a scenario with abrupt nonstationarity—at iteration $i = 2000$, $w^o \to -w^o$: *normalized convex combination* ($\mu_a = 1.5$, $\beta = 0.9$, $\epsilon = 10^{-4}$) [16], *convex combination with transfer of coefficients* ($\mu_a = 500$, $\alpha = 0.95$, $\beta = 0.98$) [6], *normalized convex combination with coefficients feedback* ($\mu_a = 1.1$, $\beta = 0.9$, $\epsilon = 10^{-4}$, $L = 50$) [2], *convexly constrained incremental combination* ($\alpha = 0.98$ and $f(x) = 2(1 + e^{-x})^{-1} - 1$) [1], and *incremental combination with the redesigned supervisor* ($\alpha = 0.97$ and $f(x) = 1.1x$ in (19)). The incremental combination clearly outperforms the other structures and the proposed supervisor considerably improves its performance with relation to the previous convexly constrained approach.

7. CONCLUSION

This work studied the transient performance of an incremental combination of LMS filters. Mean and mean-square models were developed and used to show that this combination is not equivalent to a VSS algorithm and that it is, in fact, a generalization of the DR-LMS. Supervisor constraints and the optimal supervisor of the combination were then formulated. Based on these results, an adaptive combiner was designed and its performance was shown by simulations and analyses.

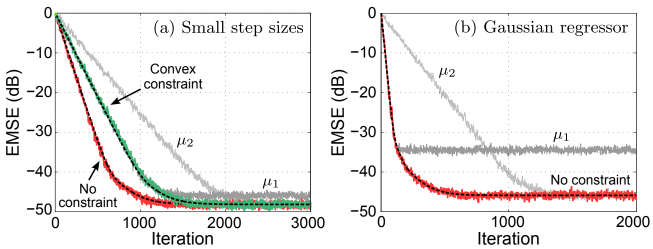


Fig. 2. Mean square model and supervisor constraint

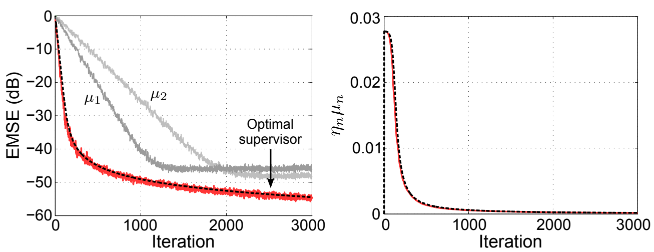


Fig. 3. Optimal supervisor (small step sizes)

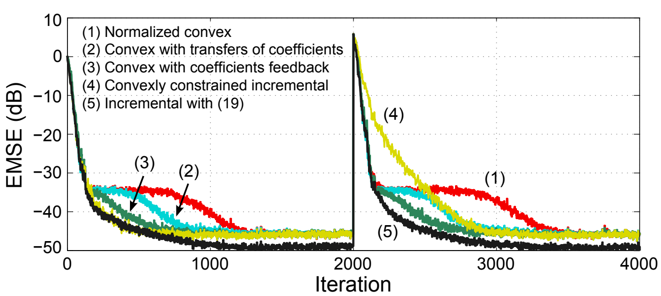


Fig. 4. Adaptive supervisor performance

8. REFERENCES

[1] W.B. Lopes and C.G. Lopes, "Incremental-cooperative strategies in combination of adaptive filters," in *ICASSP*, 2011, pp. 4132–4135.

[2] L.F.O. Chamon, W.B. Lopes, and C.G. Lopes, "Combination of adaptive filters with coefficients feedback," in *ICASSP*, 2012, pp. 3785–3788.

[3] J. Arenas-García, A.R. Figueiras-Vidal, and A.H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54[3], pp. 1078–1090, 2006.

[4] 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[5], pp. 1853—-1864, 2008.

[5] 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," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46[4], pp. 1865–1879, 2010.

[6] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A.R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Process.*, vol. 86[9], pp. 2430–2438, 2006.

[7] V.H. Nascimento and R.C. de Lamare, "A low-complexity strategy for speeding up the convergence of convex combinations of adaptive filters," in *ICASSP*, 2012, pp. 3553–3556.

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

[9] L.F.O. Chamon, H.F. Ferro, and C.G. Lopes, "A data reusage algorithm based on incremental combination of LMS filters," in *Asilomar Conference on Signals, Syst. and Comput.*, 2012, pp. 406–410.

[10] S. Shaffer and C.S. Williams, "Comparison of LMS, alpha LMS, and data reusing LMS algorithms," in *Asilomar Conference on Signals, Syst. and Comput.*, 1983, pp. 260–264.

[11] S. Roy and J.J. Shynk, "Analysis of the data-reusing LMS algorithm," in *Midwest Symposium on Circuits and Systems*, 1989, pp. 1127–1130.

[12] D.P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM J. Optim.*, vol. 7[4], pp. 913–926, 1997.

[13] M.T.M. Silva, V.H. Nascimento, and J. Arenas-García, "A transient analysis for the convex combination of two adaptive filters with transfer of coefficients," in *ICASSP*, 2010, pp. 3842—-3845.

[14] V.H. Nascimento, M.T.M. Silva, R. Candido, and J. Arenas-García, "A transient analysis for the convex combination of adaptive filters," in *SSP*, 2009, pp. 53–56.

[15] S.S. Kozat, A.T. Erdogan, A.C. Singer, and A.H. Sayed, "Transient analysis of adaptive affine combinations," *IEEE Trans. Signal Process.*, vol. 59[12], pp. 6227–6232, 2011.

[16] 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 *ICASSP*, 2008, pp. 3301–3304.