INCREMENTAL-COOPERATIVE STRATEGIES IN COMBINATION OF ADAPTIVE FILTERS

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

A new topology for combination of adaptive filters is proposed. Based on incremental strategies, the standard convexly combined parallel-independent filters are rearranged into a series-cooperative configuration without changing the computational complexity. Two new algorithms are derived from the new topology. Simulations in a stationary system identification scenario show the superior performance of the new algorithms.

Index Terms— convex combination, adaptive filtering, incremental strategies.

1. INTRODUCTION

Combination of adaptive filters (AF) has been explored in the literature in order to improve filtering performance when an accurate design of a single filter is difficult. In such an approach, a set of AFs is aggregated via a supervisor which attempts to achieve universal behavior, in which the overall system performs at least as well as the best filter in the set, usually in the mean-square error sense. Combinations of adaptive filters with different step-sizes, different orders and different adaptive algorithms are available in [1]–[5]. In such schemes, an adaptive parameter aggregates the component filters via a convex combination so that the resulting overall structure achieves fast convergence and accuracy in steady-state, as well as better tracking properties, if the combining parameter is properly adapted.

In this work, a new combination structure and algorithms are introduced, inspired by incremental cooperative strategies and the celebrated convex combination schemes [1, 6]. The component filters are rearranged in a series-cooperative topology and two new algorithms are introduced with advantages with respect to convex parallel–independent counterpart. The new techniques are motivated in the system identification formulation with LMS component filters handling stationary signals, although the approach may be readily extended to other learning rules.

2. PARALLEL-INDEPENDENT STRUCTURE

The common ground for the convex structures currently available in the literature is that the component adaptive filters (AFs) are independent and operate, in a sense, in parallel. As depicted in Fig. 1, the outputs of a fast filter (LMS_1) and an accurate filter (LMS_2) are convexly aggregated via a combining parameter $\lambda(i)$

$$w_{i-1} = \lambda(i)w_{1,i-1} + [1-\lambda(i)]w_{2,i-1} \quad (1)$$

The first author was supported by a scholarship from CAPES granted by the Electrical Engineering Graduate Program of University of Sao Paulo, Brazil (code: 33002010045P3)

where $w_{1,i-1}$ and $w_{2,i-1}$ are the individual LMS filters updated independently according to [7]

$$w_{k,i} = w_{k,i-1} + \mu_k u_i^*(d(i) - u_i w_{k,i-1}), \quad k = 1, 2 \quad (2)$$

resembling the usually adopted system identification scenario, in which u_i is a $1 \times M$ row regressor vector that captures samples of an input (white) signal $u(i)$ with variance σ_u^2 and μ_k is the filter step-size. The plant output is modeled as $d(i) = u_i w^o + v(i)$, where $v(i)$ is the white gaussian measurement noise with variance σ_v^2, and w^o is a $M \times 1$ column vector that models the unknown plant. The

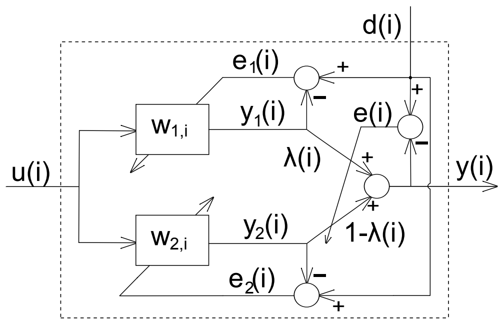

Fig. 1. Adaptive convex combination of two transversal filters

combining factor $\lambda(i)$ plays the role of an activation function, which is chosen to guarantee convexity [1, 2, 4], e.g.,

$$\lambda(i) = \frac{1}{1 + e^{-a(i)}} \quad (3)$$

The parameter $a(i)$ is adapted attempting to minimize the overall estimation error $e(i) = d(i) - u_i w_{i-1}$ in the mean-square sense.Generally a gradient-descent rule is adopted

$$a(i) = a(i-1) + \mu_a e(i)[y_1(i) - y_2(i)]\lambda(i)[1-\lambda(i)] \quad (4)$$

where $y_k(i) = u_i w_{k,i-1}$, $k = 1, 2$, and μ_a is a step-size. Fig. 1 depicts the arrangement. The resulting algorithm is typically known as the convex LMS algorithm, or CLMS for short, and it is well known to present universal behavior. Fig. 2 depicts the excess mean-square error ($EMSE = E|u_i(w^o - w_{i-1})|^2$) curves for a typical example employing $\mu_1 = 0.07$, $\mu_2 = 0.007$, $\mu_a = 1000$, $\sigma_u^2 = 1$ and $\sigma_v^2 = 10^{-3}$. Note how CLMS is able to track the transient response of the faster filter (μ_1) and reach the steady-state performance of the more accurate (slower) filter (μ_2) [1].

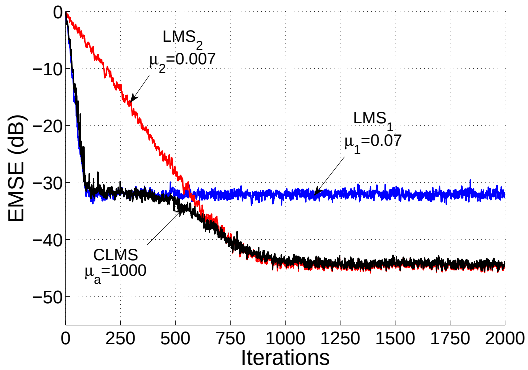

Fig. 2. EMSE of the LMS filters and their Convex Combination averaged over 200 realizations.

3. SERIES-COOPERATIVE STRUCTURE

Despite the clear CLMS advantage over the component filters, the transient of the accurate filter is not relevant and the steady-state of the fast filter is wasted: most of the time one filter is numerically annihilated by the combiner λ. This is caused by the inherent parallel-independent structure, in which the overall system inexorably "awaits" the accurate filter to catch up in order to quickly commute. In stationary environments λ can be interpreted as a switching mechanism. The great advantage offered by that structure is the simple design of the combiner.

In order to compensate for the aforementioned effect and anticipate the switching time, in the literature ad-hoc weight transfers ($w_1 \rightarrow w_2$) are conditionally performed [8], and further control mechanisms are required, since the accurate filter may be contaminated with the higher gradient noise arising from the fast filter.

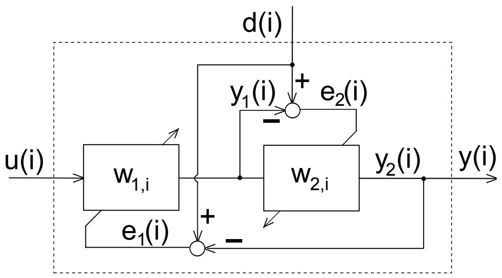

Fig. 3. Proposed topology

3.1. A switching algorithm

The ad-hoc weight transfer procedure may be formally motivated and naturally implemented, without resorting to control mechanisms, inspired by incremental and cooperative structures [6]. Topologically the filters are rearranged in series – see Fig. 3, and λ now continuously and progressively transfer the weights. The resulting algorithm is quite simple and summarized in the sequel:

$$\begin{aligned} w_{1,i} &= w_{i-1} + \mu_1 \lambda(i) u_i^*(d(i) - u_i w_{i-1}) \\ w_{2,i} &= w_{1,i} + \mu_2 (1-\lambda(i)) u_i^*(d(i) - u_i w_{1,i}) \\ w_i &\leftarrow w_{2,i} \end{aligned} \quad (5)$$

Note that the filters are no longer independent, they explicitly cooperate, balanced by λ. Furthermore, the incremental arrangement allows λ to play simultaneously the role of a combiner while decreasing the net step-size at the same time. On the other hand, direct cooperation makes the combiner design quite challenging.

3.2. Enhancing performance: simultaneous operation

The potential of the series-cooperative structure can be further explored if simultaneous operation is implemented. For that, λ can be more efficiently used as follows

$$\begin{aligned} w_{1,i} &= w_{i-1} + \mu_1 \lambda(i) u_i^*(d(i) - u_i w_{i-1}) \\ w_{2,i} &= w_{1,i} + \left[\frac{\mu_1 \lambda(i) + (1-\lambda(i))\mu_2}{1/\gamma} \right] u_i^*(d(i) - u_i w_{1,i}) \\ w_i &\leftarrow w_{2,i} \end{aligned} \quad (6)$$

where $\gamma \in (0, 1]$ is a step-size contracting factor introduced to improve steady-state while keeping transient performance. This is only possible in the series-cooperative arrangement; nevertheless, λ has to be designed so that the AFs operate simultaneously.

4. DESIGN OF THE MIXING PARAMETER

As a matter of fact, algorithms (5) and (6) may be regarded as a resource reallocation of the original CLMS: the same AFs are employed (same complexity), the same combiner and the same signals. Therefore, a direct comparison is fair. The challenge is to design $\lambda(i)$ properly, since the filters are explicitly impacting each other via the incremental procedure.

In this section we illustrate the potential of the new structure and both new algorithms. Initially a deterministic design for λ is introduced to test the new algorithms; in the sequel a simple though effective way to design the mixing parameter automatically is presented.

4.1. Deterministic design

Due to the "switching nature" attributed to λ, it can be chosen similarly to the parallel case

$$\lambda(i) = \frac{1}{1 + e^{s\cdot(i-n)}} \quad (7)$$

where now n is the *activation instant* and s controls the curve *smoothness*. The parameters $[s, n]$ have been tuned carefully to extract the best performance from the new INC-COOP algorithms and the CLMS algorithm, yielding a meaningful comparison.

Consider the system identification scenario and let $w^o = \frac{1}{\sqrt{10}}[1, 1, \cdots, 1]$, ($\|w^o\| = 1$) and $M = 10$, the signal variances are $\sigma_u^2 = 1$ and $\sigma_v^2 = 10^{-3}$. All the LMS component filters used in the combinations (CLMS, INC-COOP1 and INC-COOP2) have step-sizes $\mu_1 = 0.07$ and $\mu_2 = 0.007$. For the CLMS, $\mu_a = 1000$. Fig. 4 depicts the $\lambda(i)$'s (top) and the EMSE (bottom) employed in the pilot experiment. For CLMS we have $[s = 0.012, n = 550]$ and for both INC-COOPs $[s = 0.015, n = 120]$. INC-COOP$_2$ uses $\gamma = 0.1$.

Note in Fig.4–bottom the superior performance of the INC-COOP algorithms. INC-COOP1 is able to promptly switch filters earlier, avoiding the stagnation experienced by the CLMS algorithm (which awaits the crossing point), and reproducing the steady-state performance of the accurate filter (LMS$_2$). Furthermore, the simultaneous operation imposed by INC-COOP$_2$ yields faster convergence

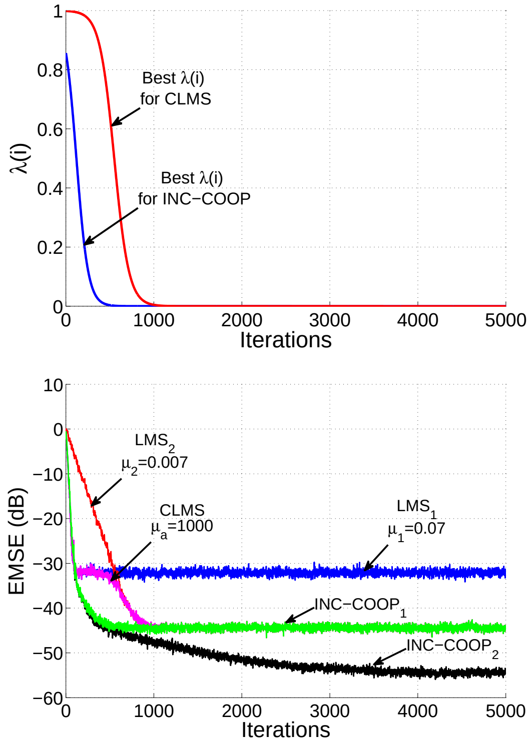

Fig. 4. Top - Time evolution of the deterministic $\lambda(i)$. Bottom - the correspondent EMSE averaged over 200 realizations.

and *smaller* error for the *same* λ. Note that in *non-stationary* environments (currently under study) all algorithms are valid candidates.

4.2. A simple design for the mixing parameter

A simple rule for parameter adjustment in adaptive filtering is to low-pass filter a quantity $q(i)$ that captures the learning status and feed it back into the adaptive process, namely

$$a(i) = \alpha \cdot a(i-1) + \beta \cdot q(i) \tag{8}$$

in which $0 < \alpha < 1$ and $q(i)$ is a chosen figure of merit related to adaptation performance. Such approach has been successfully adopted across several areas in adaptive filtering, such as step-size design [9], regularization control [10] and robust filtering [11].

Heuristically, experience across the several fields in adaptive filtering aforementioned shows that $0.95 < \alpha < 0.99$ renders a good learning evolution for a wide Signal-to-Noise Ratio (SNR) range. For a quick design, one can assign $\beta = (1-\alpha)$. Depending on the metric selected for $q(i)$, $\beta < (1-\alpha)$ compensates for low SNR (say $\beta = 0.1 \cdot (1-\alpha)$). Detailed analysis is required to show the impact of such parameters on system performance (future work).

Here the output overall quadratic error is chosen $q(i) = e^2(i)$ to train the INC-COOP algorithms, where $e(i) = d(i) - u_i w_{2,i-1}$. Since $e^2(i)$ approaches zero, a slight bias is required in λ for the INC-COOP case

$$\lambda_s(i) = \frac{2}{1+e^{-a(i)}} - 1 \tag{9}$$

so that the full excursion $\lambda_s \in [0,1]$ is guaranteed.

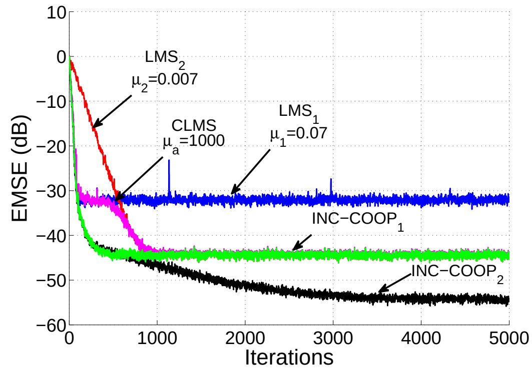

Fig. 5. EMSEs of the combinations using adaptive $\lambda(i)$ averaged over 200 realizations.

5. SIMULATIONS

The low-pass filter (8) and the $\lambda_s(i)$ (9) are implemented with $\alpha = 0.98$ and $a(-1) = 10$ in all the simulations to evaluate the performance of the INC-COOP algorithms as compared to the CLMS algorithm. Simulations are carried out in the system identification scenario mentioned in subsection 4.1.

5.1. High SNR

Algorithms have been tested at $\sigma_v^2 = 10^{-3}$ (SNR=30dB). Fig. 5 shows the EMSEs curves of the INC-COOP algorithms and CLMS averaged over 200 realizations. Both algorithms proposed present a transient response at least as fast as CLMS, allied with the *same* steady-state EMSE as that of the accurate LMS component filter (μ_2). In particular, INC-COOP$_2$ goes beyond and achieve a steady-state EMSE level of approximately 10dB *lower* than CLMS.

5.2. Low SNR

Additionally, simulations have been performed at SNR=$\{10, 5, 3\}$ dB, with σ_v^2 tuned correspondingly. With no change in the combinations parameters, it can be seen in Fig. 6 that the INC-COOP algorithms perform better than the CLMS for this specific situation. Moreover, INC-COOP$_2$ presents the best performance among the three combinations, significantly lower than the CLMS. Also, the INC-COOP curves present lower variance than CLMS and are less susceptible to sparks. Note that $\beta = (1-\alpha)$ was used in SNR=10dB. For SNR=$\{5, 3\}$dB, $\beta = 0.1(1-\alpha)$.

6. CONCLUSION

This work introduced a new framework for combination of adaptive filters. Motivated in a simple though meaningful scenario, the new technique is able to naturally circumvent the stagnation effect without sacrificing steady-state performance. This is achieved with no extra complexity. The same effect in the parallel-independent case is alleviated *only partially* and relies on extra weight transfer control mechanisms.

Future work includes deriving new training techniques for $\lambda(i)$, the use of different algorithms for the component filters and study in non-stationary environments. Furthermore, mean-square analysis and affine combinations [12] will also be considered.

7. REFERENCES

[1] J. Arenas-Garcia, 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, No. 3, pp. 1078–1090, 2006.

[2] Cassio G. Lopes, E. H. Satorius and Ali H. Sayed “Adaptive carrier tracking for direct-to-earth Mars communications”, *Proc. IEEE 40th Asilomar Conference on Signals, Systems and Computers*, pp.1042-1046, Pacific Grove, CA, 2006.

[3] Cassio G. Lopes, E. H. Satorius, P. Eastbrook, and A. H. Sayed, “Efficient adaptive carrier tracking for Mars to Earth communications during entry, descent, and landing”, *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 517-521, Madison, WI, August 2007..

[4] Cassio 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 Transactions on Aerospace and Electronic Systems* Oct. 2010.

[5] Magno T. M. Silva and Vitor H. Nascimento “Improving the Tracking Capability of Adaptive Filters via Convex Combination”, *IEEE Transactions on Signal Processing*, vol. 56, No. 7, pp. 3137–3149, 2008.

[6] Cassio G. Lopes and Ali H. Sayed “Distributed adaptive incremental strategies: Formulation and performance analysis”, *Proc. IEEE ICASSP*, Toulouse, France, vol. 3, pp. 584-587, May 2006.

[7] Ali H. Sayed, *Adaptive Filters*, Wiley Inter-Science, 2008.

[8] Magno T. M. Silva, Vitor H. Nascimento and J. Arenas-Garcia “A Transient Analysis for the Convex Combination of two Adaptive Filters with Transfer of Coefficients”, *Proc. IEEE ICASSP*, Dallas, TX, May 2010.

[9] Cassio G. Lopes and Jose Carlos M. Bermudez “Evaluation and Design of Variable Step-Size Adaptive Algorithms”, *Proc. IEEE ICASSP* Salt Lake City, UT, vol. 6, pp. 3845 - 3848, May 2001.

[10] Manolis C. Tsakiris and Cassio G. Lopes “An Alternative Criterion for Regularization in Recursive Least-Squares Problems”, *to appear in proceedings of the 7th International Symposium on Wireless Communication Systems*, York, U.K., September 2010, IEEE Press.

[11] Manolis C. Tsakiris and Cassio G. Lopes “A robust affine projection algorithm with feedback compensation of the condition number”, *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)* pp. 2057-2060, Tapei, Taiwan, May 2009, IEEE Press.

[12] Neil J. Bershad, Jose Carlos M. Bermudez and Jean-Yves Tourneret “An Affine Combination of Two LMS Adaptive Filters - Transient Mean-Square Analysis,” *IEEE Transactions on Signal Processing*, vol. 56, No. 5, pp. 1853–1864, 2008.

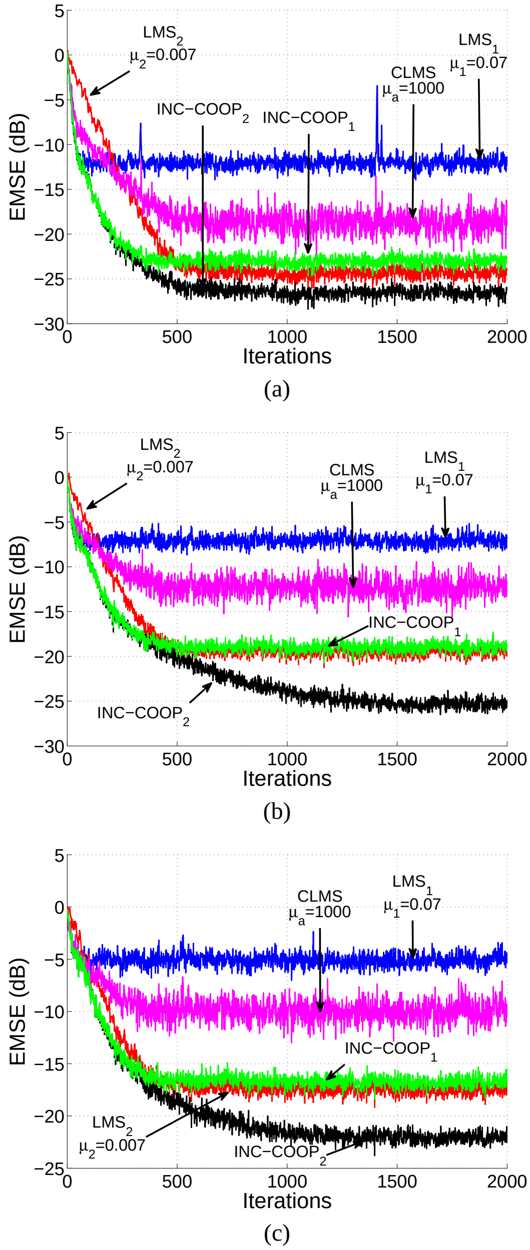

Fig. 6. EMSEs of the combinations for (a)SNR=10dB, (b)SNR=5dB, (c)SNR=3dB. The curves are averaged over 200 realizations.