COMBINATION OF ADAPTIVE FILTERS WITH COEFFICIENTS FEEDBACK

Luiz F. O. Chamon, Wilder B. Lopes and C´assio G. Lopes

Universidade de S˜ao Paulo
Dept. of Electronic Systems
S˜ao Paulo, SP - Brazil
chamon@usp.br, wilder@usp.br, cassio@lps.usp.br

ABSTRACT
In parallel combinations of adaptive filters, the component filters are usually run independently to be later on combined, leading to a stagnation phase before reaching a lower error. Conditional transfers of coefficients between the filters have been introduced in an attempt to address this issue. The present work proposes a more natural way of accelerating the convergence to steady-state, using a cyclic feedback of the overall weights to all component filters, instead of a unidirectional conditional transfer. It is shown that, depending on the cycle length, the resulting recursion is equivalent to either: (i) the independent combination, (ii) a variable step size adaptive filter, or (iii) a new hybrid algorithm. Comments on the universality of the approach are presented along with a technique to design the cycle length. Comparisons in stationary and non-stationary system identification scenarios demonstrate the superior performance of this new combination method.

Index Terms— Convex combination, adaptive filters, coefficients feedback

1. INTRODUCTION
Combinations of adaptive filters (AFs) have been introduced as a way to improve filtering performance when the accurate design of a single filter is difficult [1–4]. This approach consists of aggregating a set of AFs using a supervisor which attempts to achieve universality, i.e., making the overall system at least as good—usually in the mean-square sense—as the best filter in the set. Combinations with different step sizes, orders, adaptive algorithms and supervising rules can be found in [3–8]. In such schemes, the component filters run independently while an adaptive mixing parameter merges their weights. This structure presents a well-known convergence stagnation while the mean-square-error (MSE) of the accurate filter does not become smaller than that of the fast one. Different topologies [1] and conditional transfers of coefficients [3, 9] have been proposed to overcome this problem.

In this paper the stagnation issue will be addressed by:
• Embedding a cyclic coefficients feedback in the combination (Fig. 1);
• Showing that for two least-mean-square filters (LMS), the cycle length (L) can turn the combination into either a convex LMS (CLMS) [3], a variable step size (VSS) algorithm or a hybrid of both;
• Proposing a method to design L;
• Comparing through simulations the performance of the new algorithm to existing methods in the literature [1, 3, 9] for different scenarios.

2. PROBLEM FORMULATION
Under the widely adopted system identification scenario, the combination problem consists of finding a mixture of AFs weights which leads to an estimate at least as good as any of the component filters individually. Explicitly, given a pool of AFs

\[ w_{n,i} = w_{n,i-1} + \mu_n p_n \]  

in which \( w_{n,i} \) is the \( M \times 1 \) weights estimate at iteration \( i \) of the \( n^{th} \) filter, \( \mu_n \) is its step size and \( p_n = -B_n \nabla \ J(w_{n,i-1}) \), with \( B_n \) any positive definite matrix, \( J(w_{n,i-1}) \) the underlying cost function the filter attempts to minimize and \( ^\dagger \) denoting the conjugate transpose, a combination is defined as

\[ w_{i-1} = \sum_{n=1}^{N} \lambda_n(i) w_{n,i-1} \]  

The mixing parameter \( \lambda_n(i) \) aggregates the estimates according to supervising rules and constraints defined by the combination strategy adopted. Generally, the overall MSE \( E[e(i)^2] \) is minimized, under the constrain \( \sum \lambda_n(i) = 1 \). The overall error \( e(i) = d(i) - y(i) \) is defined over the combination output \( y(i) = u_i w_{i-1} \), where \( u_i \) is the \( 1 \times M \) regressor vector that captures samples of an input signal with variance \( \sigma_u^2 \), and \( d(i) = u_i w^o + v(i) \) is the desired signal, in which \( v(i) \) is an independent and identically distributed (i.i.d.) noise with variance \( \sigma_v^2 \), and \( w^o \) is an \( M \times 1 \) vector that models the unknown plant. For the individual filters, \( e_n(i) = d(i) - y_n(i) \) and \( y_n(i) = u_i w_{n,i-1} \).

3To account for different filters’ orders, \( M = \max_n M_n \), filling up the vectors with zeros whenever necessary to match the dimensions [5, 6].
For \( N = 2 \) and \( M_1 = M_2 = M \), an approach which has gained a lot of attention in the last years is the convex combination of two independent AFs [3,7], in which (2) reduces to
\[
 w_{i+1} = \lambda(i) w_{1,i+1} + (1 - \lambda(i)) w_{2,i+1},
\]
and \( \lambda \in [0, 1] \). In order to guarantee this last constraint and reduce gradient noise when \( \lambda \approx 0 \) or \( \lambda \approx 1 \), an activation function is used\(^2\)
\[
\lambda(i) \triangleq \frac{1}{1 + e^{-a(i-1)}},
\]
where the auxiliary variable \( a(i) \) is adapted as [3]
\[
a(i) = a(i-1) + \mu_a e(i) [y_2(i) - y_1(i)] \lambda(i)[1 - \lambda(i)],
\]
and limited to \([-a^+, a^+]\) so as not to stall the adaptation.

A detailed study of the convex combination strategy using two LMS filters with different step sizes is carried out in [3]. Fig. 2 depicts the excess mean-square-error (EMSE = \( E[|u_i(w^* - w_i)|^2] \)) for a typical example employing \( \mu_1 = 0.07 \), \( \mu_2 = 0.007 \), \( \mu_a = 1000 \), \( \sigma^2_e = 1 \), and \( \sigma^2_w = 10^{-8} \), averaged over 200 realizations.

![Fig. 2. EMSE of LMS filters and their convex combination](image)

Despite the clear advantage of the CLMS over its components, the combination is most of the time ignoring one of the AFs, leading to a convergence stagnation as soon as the faster filter reaches steady-state. The combination must then hold until the accurate filter catches up to commute due to their parallel-independent operation.

### 2.1. Conditional transfer of coefficients

Transfer of coefficients has been studied for the two LMS case in section 2.1. Conditional transfer of coefficients inspired by incremental-cooperative strategies, is proposed in [1].

2.2. Parallel structure with cyclic feedback

In order to improve the results obtained with the method in section 2.1, while avoiding its implementation issues, a cyclic feedback approach is proposed as an alternative to transferring coefficients. In this approach, the overall estimate (2) is fed back to the component AFs every \( L \) iterations—the cycle length—, providing all of them with the best weights estimate available at that time (Fig. 1). Therefore, \( w_{n,i-1} \) in (1) becomes
\[
w_{n,i-1} = \delta(i - kL)w_{n,i-1} + (1 - \delta(i - kL))w_{n,i-1},
\]
where \( w_{i-1} \) is given by (2), \( \delta(i) \) is the Kronecker delta, and \( k \in \mathbb{N} \).

Conceptually, this new algorithm is more natural than the one proposed in Section 2.1, given that it provides all filters with the global weights, explicitly chosen to minimize a function of the overall error—e.g. MSE. Additionally, the feedback is neither directional nor limited to any two filters, so that all components can take advantage of the overall coefficients. Still, in the theoretical front, Section 3 will show that this method bridges a gap between combinations of AFs of the same kind and VSS techniques. In terms of implementation, this structure also brings advantages since it depends uniquely on counters and allows efficient interruption-based algorithms. Lastly, simulations will show that it can outperform other combinations in different scenarios (Section 4).

### 3. A BRIEF ON ANALYSIS

In order to illustrate the idea, we restrict the sections to follow to the combination of two LMS for white real-valued Gaussian inputs. With \( n = 1,2, M_1 = M_2 = M \), the complete algorithm is
\[
\lambda(i) = \frac{1}{1 + e^{-a(i-1)}},
\]
\[
a(i) = a(i-1) + \mu_a e(i) [y_2(i) - y_1(i)] \lambda(i)[1 - \lambda(i)],
\]
\[
w_{n+1} = \delta(i - kL)w_{n,i-1} + (1 - \delta(i - kL))w_{n,i-1}
\]
\[
w_{n,i} = \lambda(i) w_{n,i} + \mu_a u_i(d(i) - u_i w_{n,i-1}).
\]

Algorithm (7) shows that depending on the choice of the cycle length it is possible to (i) recover the CLMS from [3]; (ii) devise a VSS algorithm; or (iii) formulate a new algorithm:

(i) For \( L \to \infty \) in (6), \( \delta(i - kL) = 0 \) for all finite \( i \) and the equation reduces to \( w_{n,i-1} = w_{n,i-1} \). This makes the component filters operate independently and is equivalent to eliminating the feedback branch in Fig. 1, leading to the CLMS [3].

(ii) However, if \( L = 1 \), (6) becomes \( w_{n,i-1} = w_{n,i-1} \), as if the feedback loop in Fig. 1 was continuously enabled. Using this result in (7) and substituting in (3) yields
\[
w_i = w_{i-1} + \sum_i (i) u_i(d(i) - u_i w_{i-1}),
\]
with \( \sum(i) = \lambda(i+1) \mu_1 + [1 - \lambda(i+1)] \mu_2 \). Note that \( \sum(i) \) is an iteration dependent convex association of \( \mu_1 \) and \( \mu_2 \), leading to a VSS variant of the LMS recursion.

(iii) Finally, choosing any finite \( L > 1 \) gives rise to a new adaptive algorithm that will present a hybrid operation, acting as in (ii) for \( i = kL \) and as in (i) for all other \( i \).
3.1. Comments on universality

A celebrated result for the convex combination is that it may achieve (nearly) universality in steady-state [3]. Since the fCLMS of case (iii) will, at each instant, be identical to either the CLMS or the VSS algorithm (8), suffices to show that the latter is universal to make (iii) will, at each instant, be identical to either the CLMS or the VSS (nearly) universality in steady-state [3]. Since the fCLMS of case 3.2. Design of the cycle length

Section 3.1 showed that the combination converges to the MSE of the best AF regardless the choice of L. However, due to the nature of the adaptation of λ, it may take a long time, especially for small L. This phenomenon is easier understood by looking at (4) and observing that the recursion on a depends upon [y1(i) − y2(i)] = w1(w1,i−1 − w2,i−1). Although w1,i−1 ≠ w2,i−1, their difference can be expected to be small after a feedback, when e1(i) = e2(i), resulting in slower updates of λ. However, it is unwise to take L to be as large as desired because then the component filters will not be able to take advantage of the overall estimation.

A method is then proposed to design the cycle length L based on the idea that, in a stationary scenario, the overall w should be fed back as soon as the faster filter stops converging, therefore supplying the slower filter with the best estimate the combination can provide at that time. To do so, a linear approximate model for the MSE convergence in dB is adopted (see Fig. 3). This model can then be used to find an approximate point where the faster filter slows down, leading to an estimate for L.

Starting from the weighted variance relation with independence for AFs of the form

\[ w_i = w_{i-1} + \mu u_i' g(u_i) \gamma_{i-1} e(i) \]

with the model (11) is tested in a high SNR/small step size scenario—\( \mu = 0.01, \sigma^2_v = 10^{-4} \)—and in the one presented above—\( \mu = \mu_1, \sigma^2_v = 10^{-2} \)—concurring with the simulations results (Fig. 3). For the adopted framework only, assessments of the convergence time for different cycle lengths were made (Fig. 4), showing the accuracy and robustness of the design method from section 3.2.

Assuming small step sizes and high SNR, a coarse approximation for (10) is \( \text{MSE}_{\text{abs}}(i) = \gamma \text{E}[w^\alpha]^2 + \mu \sigma^2_v \sigma^2_o M \sum_{k=0}^{\infty} \gamma^k r(i) = a_i + b \) with \( a = 10 \log \gamma \) and \( b = 10 \log(\sigma^2_v - \sigma^2_o) \)

With (11), L is estimated via \( r(L) = 10 \log \text{MSE}(\infty) \), where \( \text{MSE}(\infty) = 2\sigma^2_v (1 - \mu \sigma^2_v)/(2 - \mu (M + 2) \sigma^2_v) \) [10, Ch. 16]. It will be shown in Section 4, the algorithm is robust with respect to L, which supports the aforementioned approximations. Moreover, it means that no accurate estimates for \( \sigma^2_v \) are required.

4. SIMULATIONS

This section presents simulations to compare the behavior of CLMS, CLMS with transfer of coefficients (CLMS2), series topology and fCLMS. The approach taken here is to design the fCLMS once and then test it under different scenarios. All parameters were chosen as in [9]: \( \sigma^2_v = 1, \sigma^2_o = 10^{-2} \) (SNR = 20dB), \( \mu_1 = 0.05, \mu_2 = 0.005, \mu_{\text{CLMS}} = 600, \mu_{\text{CLMS2}} = 200, \mu_{\text{CLMS2}} = 100, \alpha = 0.9535, \alpha^* = 4, \) and \( M = 7 \). For the series topology, the INC-COOP with simple design was used with the same configurations as [1], but with the step sizes above. All figures show curves averaged over 1000 runs with a normalized randomly generated \( w^o \).

The model (11) is tested in a high SNR/small step size scenario—\( \mu = 0.01, \sigma^2_v = 10^{-4} \)—and in the one presented above—\( \mu = \mu_1, \sigma^2_v = 10^{-2} \)—concurring with the simulations results (Fig. 3). For the adopted framework only, assessments of the convergence time for different cycle lengths were made (Fig. 4), showing the accuracy and robustness of the design method from section 3.2.

\[ \text{MSE}_{\text{abs}}(i) = 10 \log \gamma + 10 \log(\sigma^2_v - \sigma^2_o) \]

3For stable LMS filters [10] 0 << \gamma < 1.

4.
out for clarity since it performs worst due to the lack of a more robust adaptation rule. In the second plot, the Mean Square Deviation (MSD(i) = E[(w_i^o - w_{i+1}^o)^2]) is showed for a random-walk model w_i^o = w_{i+1}^o + q(i), with q a M x 1 i.i.d. zero-mean Gaussian vector with covariance matrix σ^2_q I and σ^2_q = 10^{-4}. In this case, all AFs performed equally, except for the series topology.

In summary, the key point is the fCLMS robustness: it is able to either outperform or match existing algorithms in different scenarios with one simple design. Performance could be improved in each scenario with specific tuning.

5. CONCLUSION

A novel scheme to overcome the stagnation problem of parallel-independent combinations was proposed by embedding cyclic coefficients feedbacks in the topology, a more natural way of reusing the AFs weights than conditional transfers. A case study for two LMS filters was used to illustrate this method, where the structure was showed to be equivalent to a CLMS, a VSS algorithm or a hybrid AF, depending on the cycle length L. A method to design this parameter was developed based on a transient model for LMS and validated. Simulations showed that the new algorithm can outperform CLMS, transfer of coefficients and series topology under different scenarios.

6. REFERENCES


