Convex Combination of Adaptive Filters for a Variable Tap-Length LMS Algorithm

Yonggang Zhang and Jonathon A. Chambers, *Senior Member, IEEE*

Abstract—A convex combination of adaptive filters is utilized to improve the performance of a variable tap-length least-mean-square (LMS) algorithm in a low signal-to-noise environment (SNR $\leq$ 0 dB). As shown by our simulations, the adaptation of the tap-length in the variable tap-length LMS algorithm is highly affected by the parameter choice and the noise level. Combination approaches can improve such adaptation by exploiting advantages of parallel adaptive filters with different parameters. Simulation results support the good properties of the proposed method.

Index Terms—Adaptive filters, convex combination filters, variable tap-length least-mean-square (LMS) algorithm.

I. INTRODUCTION

THE least-mean-square (LMS) algorithm has been extensively used in many applications as a consequence of its simplicity and robustness [1], [2]. When applying the LMS algorithm, the tap-length of the adaptive filter is generally fixed. However, in certain applications, the tap-length of the optimal filter is unknown or even variable. When the tap-length is under-modeled, the mean-square output error is likely to increase, as shown in the analysis in [3] and [4]; thus, a variable tap-length algorithm is needed in such environments.

Methods have been proposed during recent years to search for the optimal filter tap-length [5]–[8], and a summary of the work is given in [9]. The fractional tap-length (FT) method is more robust and has lower computational complexity compared with other methods [9], but its performance can depend on the parameter choice, particularly when the noise level is high. Such sensitivity to parameter selection motivates a search for new approaches to variable tap-length adaptive filtering. As described in [10] and [11], the convex combination of adaptive filters can improve the performance of adaptive schemes. In this letter, this combination will be used for the first time together with the FT method to solve the optimal filter tap-length search problem in a high noise environment, where SNR $\leq$ 0 dB. Simulations will be performed to support the advantages of this new approach.

The remainder of this letter is organized as follows: In Section II, we will formulate the FT variable tap-length algorithm. Convex combination of adaptive filters will be introduced in Section III. The convex combination for the FT method will be formulated in Section IV. Simulations will be given in Section V to support the advantages of this new approach. Section VI concludes this letter.

II. FT VARIABLE TAP-LENGTH LMS ALGORITHM

The FT variable tap-length LMS algorithm is designed to find the optimal filter tap-length. In common with most approaches to derive algorithms for adaptive filters, this problem is transferred to the optimization of some criteria related to the tap-length. For formulation convenience, we denote the steady-state tap-length of the FT algorithm as L; $\mathbf{w}_L$ and $\mathbf{x}_L(n)$ are the corresponding steady-state filter vector and input vector, respectively, and n denotes the discrete time index. Also, we define the segmented steady-state error as [9]

$$e_M^{(L)}(n) = d(n) - \mathbf{w}_{L,1:M}^T \mathbf{x}_{L,1:M}(n) \qquad (1)$$

where $d(n)$ is the desired signal, $1 \leq M \leq L$, $\mathbf{w}_{L,1:M}$, and $\mathbf{x}_{L,1:M}(n)$ are vectors consisting of the first M coefficients of the filter vector $\mathbf{w}_L$ and the input vector $\mathbf{x}_L(n)$, respectively. The mean square of this segmented steady-state error is defined as $\xi_M^{(L)} = E\{(e_M^{(L)}(n))^2\}$. The underlying basis of the FT method is to find the minimum value of L that satisfies [9]

$$\xi_{L-\Delta}^{(L)} - \xi_L^{(L)} \leq \varepsilon \qquad (2)$$

where Δ is a positive integer less than L, and ε is a small positive value determined by the system requirements. The minimum L that satisfies (2) is then chosen as the optimum tap-length. A detailed description of this criterion and another similar criterion can be found in [9].

Gradient-based methods can be used to solve for L on the basis of (2). However, the tap-length that will be used in the adaptive filter structure must be an integer, and this constrains the adaptation of the tap-length. Different approaches have been applied to solve this problem [5]–[9]. In [9], the conception of "pseudo fractional tap-length" denoted by $l_f(n)$ is utilized to make instantaneous tap-length adaptation possible. As explained in [9], $l_f(n)$ is no longer constrained to be an integer, and the true tap-length remains unchanged until the "change" of the fractional tap-length accumulates to some extent. Based on this approach, the FT algorithm can then be formulated as follows:

$$l_f(n+1) = (l_f(n) - \alpha) - \gamma \left[\left(e_{L(n)}^{(L(n))} \right)^2 - \left(e_{L(n)-\Delta}^{(L(n))} \right)^2 \right] \qquad (3)$$

where γ is the step size for the tap-length adaptation, and α is the leakage parameter and should be much less than γ [9]. The tap-length $L(n+1)$ that will be used in the adaptation of the filter weights in the next iteration is decided from the previous fractional tap-length $l_f(n)$

$$L(n+1) = \begin{cases} \lfloor l_f(n) \rfloor, & \text{if } |L(n) - l_f(n)| > \delta \\ L(n), & \text{otherwise} \end{cases} \qquad (4)$$

where $\lfloor . \rfloor$ is the floor operator that rounds down the embraced value to the nearest integer.

Although this FT method performs well under low-noise condition, its performance depends on the choice of the parameters. In a high-noise environment, SNR ≤ 0 dB, fixed parameters that achieve both fast convergence rate and small steady-state mean-square error (MSE) will be difficult to obtain. Next, we will introduce a convex combination of adaptive filters that can improve the performance of the FT method in such high-noise environments.

III. CONVEX COMBINATION OF ADAPTIVE FILTERS

In previous research into a convex combination structure, two filters are adapted separately [10]. The output signals and the output errors of both filters are combined in such a manner that the advantages of both component filters are retained: the rapid convergence from the fast filter and the reduced steady-state error from the slow filter. Here, we assume that the first filter has a fast convergence rate throughout this letter. The output of the overall filter is

$$y(n) = \lambda(n)y_1(n) + (1 - \lambda(n))\, y_2(n) \qquad (5)$$

where $y_i(n) = \mathbf{w}_i^T(n)\mathbf{x}_i(n)$, $i = 1, 2$, $\mathbf{w}_i(n)$ and $\mathbf{x}_i(n)$ are the adaptive filter weight vector and input vector of the ith filter, and $\lambda(n) \in [0, 1]$ is a mixing scalar parameter. The output error of the overall filter is

$$e(n) = \lambda(n)e_1(n) + (1 - \lambda(n))\, e_2(n) \qquad (6)$$

where $e_1(n)$ and $e_2(n)$ are the output errors of the two component filters

$$e_i(n) = d(n) - \mathbf{w}_i^T(n)\mathbf{x}_i(n), \quad i = 1, 2. \qquad (7)$$

The key point of the convex combination of adaptive filters is to control the overall filter by the mixing parameter $\lambda(n)$ according to the performance of the two component filters.

In the convex combination of adaptive filters, the mixing parameter $\lambda(n)$ is adapted to minimize the quadratic error of the overall filter [10]. Rather than adapting $\lambda(n)$ directly, a variable parameter $a(n)$ that defines $\lambda(n)$ via a sigmoidal function is adopted. The sigmoidal function is

$$\lambda(n) = sgm\,(a(n)) = \left(1 + e^{-a(n)}\right)^{-1} \qquad (8)$$

and the update equation for $a(n)$ is given by

$$a(n + 1) = a(n) + \mu_a e(n)\,[y_1(n) - y_2(n)]\,\lambda(n)\,[1 - \lambda(n)] \qquad (9)$$

where μ_a is the step size of the adaptation of $a(n)$ and should be chosen appropriately to obtain a fast and stable convergence of the combination. The parameter $a(n)$ is also restricted to the interval $[-a^+, a^+]$, which limits the permissible range of $\lambda(n)$ to $[1 - \lambda+, \lambda+]$, where $\lambda+ = sgm(a^+)$ is a constant close to unity [10]. A good choice for a^+ is 4, which constrains $\lambda(n)$ to $[0.018, 0.982]$. This value has been used in [10] and also will be used in our simulations.

As shown in [10], the steady-state performance of the convex of combination adaptive filters is better than or as close as desired to its best component filter. If we denote the noise contained in the desired signal $d(n)$ as $t(n)$ and define the excess mean-square error (EMSE) of the overall filter as $J_{ex}(n) = E\{(e(n) - t(n))^2\}$, the advantage of the convex combination structure can be shown as [10]

$$J_{ex}(\infty) \leq \min\,[J_{ex,1}(\infty), J_{ex,2}(\infty)] \qquad (10)$$

where $J_{ex}(\infty)$, $J_{ex,1}(\infty)$, and $J_{ex,2}(\infty)$ denote the steady-state EMSE of the overall filter, the first component filter, and the second component filter, respectively. Note that no assumption is made about the specific nature of the adaptive filter, and thus, (10) is suitable for any adaptive algorithm [10].

Two modifications have been proposed to improve the performance of the original convex combination algorithm [11], and both are used in our simulations. One of the modifications is to take advantage of the fast filter to speed up the convergence of the slow one. It is achieved by modifying the adaptation of the slow component filter at an early stage of the adaptation to approach that of the fast component filter.

Another modification is to improve the convergence of the parameter $a(n)$. It is clear that when both outputs of the two filters are similar, the factor $y_1(n) - y_2(n)$ in (9) will be very small, and the convergence of $a(n)$ will be slow. A momentum term for adapting parameter $a(n)$ is then added to alleviate this problem [11]

$$a(n + 1) = a(n) + \mu_a e(n)\,[y_1(n) - y_2(n)]\,\lambda(n)\,[1 - \lambda(n)] + \rho\,(a(n) - a(n - 1)) \qquad (11)$$

where ρ is a positive constant. In general, 0.5 is a good choice of the parameter ρ, as shown in [11]. Compared with the basic convex combination of adaptive filters, these modifications have improved the convergence rate of the overall filter. Next, we will consider for the first time applying this convex combination of filters in the FT variable tap-length algorithm.

IV. CONVEX COMBINATION FILTERS FOR THE FT ALGORITHM

In the structure of convex combination for the FT algorithm, two component adaptive filters are utilized to implement the FT method separately, but the philosophy for selecting parameters γ, α, and Δ in (3) of each component filter is different. The parameters in the first filter are set to provide quick convergence rate of the tap length, whereas parameters for the second filter are set to provide a smooth curve of the evolution of the tap-length, which results in a small steady-state EMSE.

Although an exact theoretical analysis for the performance of the resulting convex combination FT algorithm is very difficult due to the nonlinearity between the output error and the tap-length, some general guides for the choice of the parameters can be obtained from the examination of (3). In particular, we will consider the following two aspects of the parameter choice problem.

1) *Convergence rate.* There are two ways to increase the convergence rate: increasing the step size γ or increasing the term $(e_{L(n)}^{(L(n))})^2 - (e_{L(n)-\Delta}^{(L(n))})^2$, which can be obtained by increasing the parameter Δ. However, a large γ will result in large fluctuation of the evolution of the tap-length and

potentially cause the algorithm to stall, because when the tap-length approaches to some value that is smaller than the optimal filter length, extra output error will be produced [3], [4]. Also, Δ is limited by the optimal tap-length, and a too-large Δ will result in an overestimation of the optimal filter tap-length due to the criterion described in Section II.

2) *Small steady-state EMSE.* This needs an accurate estimate of the tap-length, which can only be obtained with small parameters γ and Δ. However, a slow convergence will also ensue.

Since the bias brought by the leakage parameter α is not significant and the choice of this parameter is not critical due to the formulation in [9], we will assume that this parameter has been chosen appropriately, as in all simulations in the next section. According to the discussions above, in the first filter, we should use a large step size γ and a large parameter Δ. In the second filter small parameters, γ and Δ should be used. Examples of the parameters γ and Δ will be given in the next section.

Both modifications of the basic convex combination structure that have been introduced in the previous section are adopted in our simulations. In the first modification, a simple criterion is set up to decide how and when we should modify the adaptation of the fractional tap-length of the slow component filter

$$l_{f2}(n+1) = \phi l_{f2}(n) + (1-\phi) l_{f1}(n), \text{ if } \lambda(n) > t \quad (12)$$

where l_{fi} denotes the fractional tap-length of the ith filter, t is a threshold between 0.5 and 1, and ϕ is a weight parameter close to but less than unity. This criterion is easy to understand: when $\lambda(n)$ is larger than t, the first filter performs better than the second filter, and the fractional tap-length of the second filter should be modified to approach to that of the first filter, to speed up the convergence rate of the second filter. In general, a value between 0.6 and 0.8 is a good choice for t, and the value 0.99 is a good choice for ϕ, which gives both fast and stable convergence for the fractional tap-length of the second filter. The second modification is described by (11).

If we denote the initial tap lengths of both component filters as L_{ini}, the step size for the adaptation of the weights vector of both component filters as μ, the implementation of this convex combination method can then be summarized as follows.

Initialization: $\alpha_1, \alpha_2, \gamma_1, \gamma_2, \Delta_1, \Delta_2, \mu, \mu_a, \delta, \rho, t, \phi, L_{ini}, a^+$. All these values should be set according to the system requirement, and an example of the set of these parameters can be seen in the simulations in Section V.

Update at each iteration:
1) Update the filter coefficients of both component filters by using the LMS or related algorithm with the current tap-lengths.
2) Adapt the fractional tap-lengths of both component filters, respectively, according to (3).
3) Calculate parameters $a(n)$ and $\lambda(n)$ according to (8) and (11).
4) Modify the fractional tap-length of the slow component filter according to (12).
5) Update the current tap-lengths of both filters according to (4).

As will be confirmed in the later simulations, the first component filter provides a good tracking ability of the tap-length for the overall filter. On the other hand, a smooth curve of tap-length and small EMSE is obtained from the second component filter. Through appropriate update of the mixing parameter λ, both

advantages of these two filters are extracted, and a better performance can thereby be obtained. Similarly to (10), the overall EMSE performance of the convex combination of filters is better than or as close as desired to the best component filter.

V. SIMULATION RESULTS

Now, we will give a simulation to support the advantages of the proposed convex combination approach. The setup of this simulation is similar to that in [9]. In the simulations in [9], a low-noise environment where SNR is 20 dB is used. Since the performance of the proposed approach is comparable with the FT method in a low-noise level, a high-noise level environment where the SNR is 0 dB is used in our simulation, to highlight the advantages of the proposed approach. The normalized LMS (NLMS) algorithm [1], [2] is also used in this simulation.

The input signal $x(n)$ is obtained by passing white Gaussian noise through a spectral shaping filter with a z-domain transfer function of $H(z) = 0.35 + z^{-1} + 0.35 z^{-2}$. Similar to that in [9], the following two unknown systems are tested:

$$\mathbf{h}_1 = \sum_{k=1}^{80} a_k z^{-k}, \quad \mathbf{h}_2 = \sum_{k=1}^{30} b_k z^{-k} \quad (13)$$

where a_k and b_k are chosen from a white Gaussian random sequence with zero mean and unit variance. The desired signal is obtained by filtering the input signal with $\mathbf{h}_1$ or $\mathbf{h}_2$ in different time intervals

$$d(n) = \mathbf{w}^T(n)\mathbf{x}(n) \quad (14)$$

where $\mathbf{w}(n) = \mathbf{h}_1$ for $n < 10,000$ or $n \geq 20,000$ and $\mathbf{w}(n) = \mathbf{h}_2$ for $10,000 \leq n < 20,000$.

Independent, zero-mean Gaussian noise is then added to the unknown system output to provide an SNR of 0 dB. The common parameters are set the same for all of the following experiments: $\mu = 0.1$, $\mu_a = 1$, $\delta = 2$, $\rho = 0.5$, $t = 0.8$, $\phi = 0.99$, $L_{ini} = 20$, and $a^+ = 4$. Also, the tap-lengths during the adaptation are constrained to be no less than L_{ini}. Two sets of parameters are tested with the FT method and the convex combination approach:
A) $\alpha = 0.08$, $\gamma = 4$, $\Delta = 15$;
B) $\alpha = 0.01$, $\gamma = 0.5$, $\Delta = 4$;
C) convex combination of A and B.

Note that the parameters of simulation A are set to obtain the best performance for the FT method, and the parameters of both simulations A and B are set to obtain the best performance for the overall filter in simulation C.

Fig. 1 shows the learning curves of the tap-lengths of simulations A, B, and C. Note that the learning curve of the tap-length of the first component filter in simulation C is the same as that in simulation A. It is clear to see in Fig. 1 that the set of parameters in simulation A is good for tracking the variability of the tap-length. However, the fluctuation of the tap-lengths is large due to the high level interference signal. The parameter set of simulation B is good for the interval where the optimal tap-length is 30. However, it is too small to estimate the channel length during intervals with optimal tap-length of 80. Both component filters in simulation C have good tracking abilities, and the second component filter has a very smooth tap-length curve. Furthermore, nearly all the estimations of the optimal filter tap-lengths in both
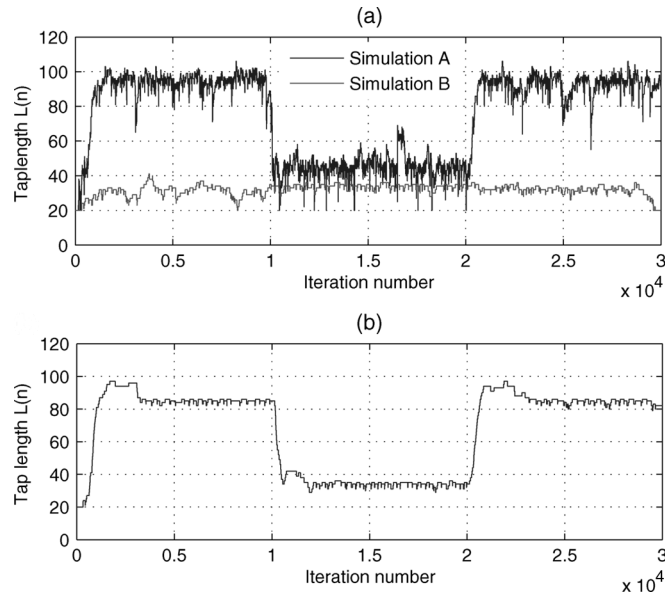
Fig. 1. Learning curves of tap-lengths of simulations A, B, and C. (a) Learning curves of tap-lengths of simulation A and B. (b) Learning curve of tap-lengths of the second filter of simulation C. (Color version available online at http://ieeexplore.ieee.org.)
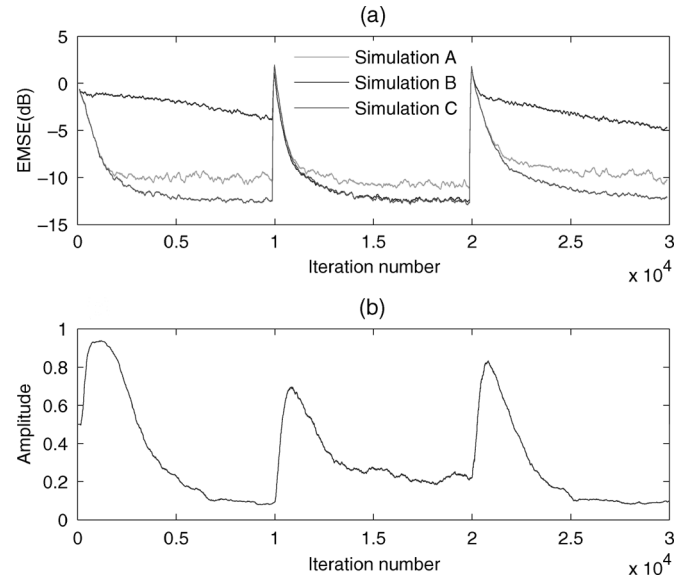

Fig. 2. Learning curves of EMSE of all simulations and the mixing parameter in simulation C. (a) Learning curves of EMSE of simulations A, B, and C. (b) Learning curve of the mixing parameter in simulation C. (Color version available online at http://ieeexplore.ieee.org.)

component filters of simulation C are larger than the real optimal filter tap-lengths, but the estimations of the second filter with smaller parameter Δ are closer to the real optimal filter tap-lengths.

To make the comparison clear, the learning curves of EMSE rather than MSE are shown in Fig. 2(a). All EMSE curves are obtained by averaging the results over 100 Monte Carlo trials of the same experiment. Note that in Fig. 1, the results are obtained by one realization for the experiment, to show the fluctuation of the tap-lengths in the filter with large parameters. It is clear to see in Fig. 2(a) that the EMSE of simulation B is large at the intervals where the optimal filter length is 80, because the set of its parameters cannot estimate the associated tap-length. The EMSE of simulation A is also large over the interval where the optimal filter length is 30 because of the large fluctuations in the tap-length estimate. Simulation C performs better than both simulations A and B due to the combination approach, and it is robust to system variation.

The evolution of the mixing parameter $\lambda(n)$ in simulation C is shown in Fig. 2(b). This curve is also obtained over 100 Monte Carlo trials of the same experiment. From this figure, we can see that the parameter $\lambda(n)$ increases toward unity initially, to provide a good tracking performance for the overall filter, and then converges to a small value to obtain a small steady-state EMSE from the second component filter. This behavior is repeated in the different regions of the figure. The evolution of this parameter clearly matches the requirements of the convex combination.

VI. CONCLUSION

In this letter, a convex combination structure is utilized to improve the performance of the FT variable tap-length algorithm. Simulation results show that this new approach performs better as compared with the original single-filter-based method. It provides a robust method for a variable tap-length algorithm in a high-noise environment and can be potentially used in many adaptive filter applications.

REFERENCES

[1] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. New York: Wiley, 1998.
[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
[3] Y. Gu, K. Tang, H. Cui, and W. Du, "Convergence analysis of a deficient-length LMS filter and optimal-length sequence to model exponential decay impulse response," *IEEE Signal Process. Lett.*, vol. 10, no. 1, pp. 4–7, Jan. 2003.
[4] K. Mayyas, "Performance analysis of the deficient length LMS adaptive algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2727–2734, Aug. 2005.
[5] F. Riero-Palou, J. M. Noras, and D. G. M. Cruickshank, "Linear equalisers with dynamic and automatic length selection," *Electron. Lett.*, vol. 37, no. 25, pp. 1553–1554, Dec. 2001.
[6] Y. Gu, K. Tang, H. Cui, and W. Du, "LMS algorithm with gradient descent filter length," *IEEE Signal Process. Lett.*, vol. 11, no. 3, pp. 305–307, Mar. 2004.
[7] Y. Gong and C. F. N. Cowan, "A novel variable tap-length algorithm for linear adaptive filters," in *Proc. ICASSP*, Montreal, QC, Canada, Jan. 2004.
[8] ——, "Structure adaptation of linear MMSE adaptive filters," *Proc. Inst. Elect. Eng., Vis., Image, Signal Process.*, vol. 151, no. 4, pp. 271–277, Aug. 2004.
[9] ——, "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2400–2407, Jul. 2005.
[10] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Steady-state performance of convex combinations of adaptive filters," in *Proc. ICASSP*, Philadelphia, PA, Mar. 2005.
[11] J. Arenas-García, V. Gómez-Verdejo, and A. R. Figueiras-Vidal, "New algorithms for improved adaptive convex combination of LMS transversal filters," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 6, pp. 2239–2249, Dec. 2005.