

# Project Report

Kailong Wang

September 20, 2023

## Abstract

This is the course report of ECE525 Optimal Signal Processing. The topic covered is the Combinations of Adaptive Filters—Performance and convergence properties.

## 1 Introduction

Adaptive filters form the foundation of numerous signal processing applications, with research focusing on enhancing their performance and properties. Novel adaptive structures aim to improve performance trade-offs, such as convergence rate and steady-state performance, as well as integrating a priori knowledge into filter learning mechanisms. However, when a priori knowledge is limited or imprecise, selecting the ideal filter structure and parameters can be challenging, potentially resulting in suboptimal performance.

One solution to this challenge is the combination of multiple adaptive filters with varying characteristics. This approach mirrors the divide and conquer principle used in ensemble learning within the machine-learning community. The focus of this report is on the review [1] of key ideas and principles behind combination schemes, emphasizing design rules to derive the deterministic bounds that apply to individual sequences.

### 1.1 Problem Formulation and Notation

The adaptive filter in the MSE sense satisfy a linear regression model of the form

$$d(n) = u(n)^T w_o(n) + v(n), \quad (1)$$

where  $w_o(n)$  represents the (possibly) time-varying optimal solution, and  $v(n)$  is a noise sequence (usually WSS), which is under the assumption of independent and identically distributed, and independent of  $u(m)$ , for all  $n, m$ . The estimation of the optimal solution at time  $n$  is usually solved by recursion of the form

$$w(n+1) = f[w(n), d(n), u(n), s(n)], \quad (2)$$

where  $f[\cdot]$  is the update function of the adaptive schemes and  $s(n)$  represents any other state information that required to update the filter.

The following error variables are considered in evaluating and designing the adaptive filters:

- Weight error:  $\tilde{w}(n) = w_o(n) - w(n)$
- A priori filter error:  $e_a(n) = u(n)^T \tilde{w}(n)$
- Filter error:  $e(n) = d(n) - u(n)^T w(n) = e_a(n) + v(n)$
- Mean squared error (MSE):  $\text{MSE}(n) = \mathbb{E}[e^2(n)]$
- Excess MSE (EMSE):  $\xi(n) = \mathbb{E}[e_a^2(n)] = \text{MSE}(n) - \mathbb{E}[v^2(n)]$
- Mean squared deviation:  $\text{MSD}(n) = \mathbb{E}[\|\tilde{w}(n)\|_2^2]$ .

During the operation, adaptive filters normally converge to a steady-state regime in which the MSE decreases to some asymptotic value. Thus, for steady-state performance, the steady-state MSE, EMSE, and MSD are defined as their limiting values as  $n$  increases. For instance, the steady-state EMSE is defined as

$$\xi(\infty) = \lim_{n \rightarrow \infty} \mathbb{E}[e_a^2(n)]. \quad (3)$$

## 1.2 A Basic Combination of Two Adaptive Filters

The most simple combination scheme incorporates two adaptive filters. Both adaptive filters have access to the same input and reference signals and produce their individual estimates of the optimum weight vector,  $w_o(n)$ . The goal of the combination layer is to learn which filter component is performing better dynamically at any particular time, assigning them weights to optimize the overall performance. The affine and convex combinations of the two filters are defined as

$$y(n) = \lambda(n)y_1(n) + (1 - \lambda(n))y_2(n), \quad (4)$$

where  $y_n = w^T(n)w_i(n)$ ,  $i = 1, 2$ ,  $\lambda(n) \in [0, 1]$  is the convex combination parameter, and  $w_i(n)$  is the weight vector of each individual filter. Accordingly, the estimated weight vector  $w(n)$ , the error  $(e(n))$ , and the a priori error  $e_a(n)$  of the combination scheme have the similar form. For the purpose of the performance, it is crucial to design mechanisms to learn the  $\lambda(n)$ . The section 2.3 will discuss the estimation of the combination parameter.

## 1.3 Optimum mixing parameter and combination performance

Here we derive the expression for the optimal mixing parameter in the affine and convex cases, in the sense of minimizing the MSE of the combination. The EMSE of the eq. (4) is

$$\xi(n) = \mathbb{E}[e_a^2(n)] = \lambda^2(n)\xi_1(n) + (1 - \lambda(n))^2\xi_2(n) + 2\lambda(n)[1 - \lambda(n)]\xi_{12}(n), \quad (5)$$

where the individual EMSE  $\xi_i(n) = \mathbb{E}[e_{a,i}^2(n)]$  and the cross-EMSE  $\xi_{12}(n) = \mathbb{E}[e_{a,1}(n)e_{a,2}(n)]$ . Equation (5) is a quadratic function of  $\lambda(n)$ , and its minimum is obtained at affine case

$$\lambda_{\text{aff}}(n) = \frac{\xi_2(n) - \xi_{12}(n)}{\xi_1(n) + \xi_2(n) - 2\xi_{12}(n)} = \frac{\Delta\xi_2(n)}{\Delta\xi_1(n) + \Delta\xi_2(n)}, \quad (6)$$

where  $\Delta\xi_i(n) = \xi_i(n) - \xi_{12}(n)$ ,  $i = 1, 2$ . Or at convex case

$$\lambda_{\text{cvx}}(n) = \left. \frac{\Delta\xi_2(n)}{\Delta\xi_1(n) + \Delta\xi_2(n)} \right|_0^1, \quad (7)$$

where the vertical line denotes truncation to the indicated values. With these expressions, some theoretical performance of optimally combined adaptive filters can be derived and it is included in the [1].

## 2 Cross-EMSE of LMS and RLS filters

The cross-EMSE plays an important role in clarifying the performance regime of a combination of two adaptive filters. In this section, we review some main results for LMS and RLS filters.

### 2.1 Combination of two LMS filters with different step sizes

The combination of two filters with different adaptation speeds is a common practice in adaptive filtering. Let's take a look at the combination of two LMS filters as an example. Using the tracking model

$$w_o(n) = w_o(n-1) + q(n), \quad (8)$$

where  $q(n)$  is the weight change vector at every step, which is assumed to have zero-mean and covariance  $Q = \mathbb{E}[q(n)q^T(n)]$ , the theoretical steady-state EMSE depends on  $\text{tr}(Q)$ . To facilitate the comparison among filters, we will recur to the normalized squared deviation (NSD)

$$\text{NSD}(n) = \frac{\xi_i(\infty)}{\xi_o(\infty)}. \quad (9)$$

With several examples provided in the [1], we concluded that

- The tracking ability of the adaptive implementation can be improved by using the combination structure. The paper further discuss the superior performance of the combination regime comparing with the variable step-size (VSS) algorithm.
- The combination of two LMS filters with different step sizes cannot outperform the best individual filter. Thus, with enough prior knowledge, the use of individual filter is preferable.

## 2.2 Improving the tracking performance of LMS and RLS filters

Next we see the combination of LMS and RLS filter. As explained in the paper, LMS will outperform RLS if  $Q$  is proportional to the autocorrelation matrix of the input signal,  $R$ , and the opposite will occur when  $Q \propto R^{-1}$ . Consider a synthetic example where  $Q$  is a mixture of  $R$  and  $R^{-1}$ ,

$$Q = 10^{-5} \left[ \alpha \frac{R}{\text{tr}(R)} + (1 - \alpha) \frac{R^{-1}}{\text{tr}(R^{-1})} \right], \quad (10)$$

the optimal EMSE for RLS filters can be larger or smaller than the optimal EMSE for LMS filters. I hypothesis that the solution space of LMS and RLS are not overlapping so the joint solution space of the combination of LMS and RLS is larger than the individual solution space. Thus, the combination of LMS and RLS can outperform the best individual filter, which is the same as the computational learning scenario.

## 2.3 Estimating the combination parameter

Since the optimum linear combiner is unrealizable, many practical algorithms have been proposed to adjust the mixing parameter in convex and affine combinations. Rewrite the eq. (4) as

$$y(n) = y_2(n) + \lambda(n)[y_1(n) - y_2(n)], \quad (11)$$

we can reinterpret the adaptation of  $\lambda(n)$  as a “second layer” adaptive filter of length one, so that in principle any adaptive rule can be used for adjusting the mixing parameter. However, the time-varying property of  $[y_1(n) - y_2(n)]$  makes the optimization of such a filter challenger than usual. In the affine case, the optimal solution of the combined filter may stay away from the optimum EMSE even with carefully selected parameters. When there is no enough prior knowledge, the convex combination schemes recur to activation functions to keep the mixing parameter in the range of interest, which simplifies the selection of the step size. A **power normalized updating rule** is proposed in the paper to obtain a more robust scheme.

## 3 Combination of several adaptive filters

We have considered the combinations of two adaptive filters. Intuitively, combining  $K$  adaptive filters makes it possible to further increase the robustness and versatility of combination schemes. For instance

- More filters can introduce new step size values, which can be used to improve the tracking ability of the combination scheme.
- More filters can alleviate several comprises simultaneously.

We can select the combination of any two elements at once and incorporates the rest one by one. In terms of the mathematical expression,

$$y(n) = \lambda_{21}(n)\{\lambda_{11}(n)y_1(n) + [1 - \lambda_{11}(n)]y_2(n)\} + [1 - \lambda_{21}(n)]\{\lambda_{12}(n)y_3(n) + [1 - \lambda_{12}(n)]y_4(n)\}. \quad (12)$$

Or in a one-layer schemes

$$y(n) = \sum_{k=1}^{K-1} \lambda_k(n)y_k(n) + \left[ 1 - \sum_{k=1}^{K-1} \lambda_k(n) \right] y_K(n). \quad (13)$$

One useful characteristic of hierarchical schemes is its ability to selects the best combination parameters of both competing models.

## 4 Calculation Cost and Applications

The rest of the paper introduce the strategies to reduce the computation cost in designing the combination schemes. For example, use a low-cost filter as companion to a high-cost one not only improve the overall performance but also reduce the computational cost of the high-cost one, which yield a systematic computation cost reduction. The paper also introduces several applications of the combined filters. For example, by track the changes in the modality of different kinds of signals, we can determine whether a signal is generated through a linear or a nonlinear system, which is not viable through the single filter. The implementations are also discussed in designing adaptive blind equalization, sparse system identification, acoustic echo cancellation.

## References

- [1] Jeronimo Arenas-Garcia et al. “Combinations of Adaptive Filters: Performance and convergence properties”. In: *IEEE Signal Processing Magazine* 33.1 (Jan. 2016), pp. 120–140. DOI: 10.1109/msp.2015.2481746. URL: <https://doi.org/10.1109/msp.2015.2481746>.