On the tracking performance of adaptive filters and their combinations

Raffaello Claser and Vítor H. Nascimento
Department of Electronic Systems Eng.
University of São Paulo
São Paulo, Brazil
{rclaser,vitnasci}@usp.br

***Abstract*—Combinations of adaptive filters have attracted attention as a simple solution to improve filter performance, including tracking properties. In this paper, we consider combinations of LMS and RLS filters, and study their performance for tracking time-varying solutions. Modeling the variation of the parameter vector to be estimated as a first order autoregressive (AR) model, we show that a convex combination between one LMS and one RLS filters with their optimum settings may have a tracking performance close to the optimal excess mean-square error (EMSE) and mean-square deviation (MSD) obtained via Kalman filter, but with lower computational complexity (linear in the filter length instead of quadratic — in the case of diagonal matrices in the Kalman model — or cubic, for general Kalman models).**

***Index Terms*—Adaptive filter, convex combination, Kalman filter, tracking**

I. Introduction

When the *a priori* knowledge about the filtering scenario is limited or imprecise, selecting the most adequate filter and adjusting its parameters becomes a challenging task, and erroneous choices can lead to a considerable loss in performance. The Kalman filter (KF) has long been shown to be the optimal solution to many tracking and data prediction tasks [1], in a wide variety of applications ranging from navigation [2], [3] to image processing [4], [5]. This filter is optimal in the sense it minimizes the mean square error of the estimated parameters when all noises involved are Gaussian and the parameter vector to be estimated changes according to a known linear model. Thus, consider a state-space description of the form:

$$\mathbf{x}_i = \mathbf{F}_i\mathbf{x}_{i-1} + \mathbf{G}_i\mathbf{t}_i, \qquad \mathbf{z}_i = \mathbf{H}_i\mathbf{x}_{i-1} + \mathbf{v}_i, \tag{1}$$

where $\mathbf{x}_i$ is the unknown $M \times 1$ state vector, $\mathbf{F}_i$ is the $M \times M$ state-transition matrix, $\mathbf{G}_i$ is the $M \times N$ control-input model, $\mathbf{t}_i$ is an $N \times 1$ real Gaussian random state noise vector with zero mean and covariance matrix $\mathbf{T}_i$, $\mathbf{z}_i$ is the $D \times 1$ observation vector, $\mathbf{H}_i$ is the $D \times M$ measurement matrix, and $\mathbf{v}_i$ is a $D \times 1$ real Gaussian random measurement noise vector with zero mean and covariance matrix $\mathbf{V}_i$. Matrices $\mathbf{F}_i$, $\mathbf{G}_i$, $\mathbf{H}_i$, $\mathbf{T}_i$ and $\mathbf{V}_i$ are assumed to be real and known.

In this case, given observations $\mathbf{z}_i$ that satisfy the state-space model described in (1), an estimate $\hat{\mathbf{x}}_i$ for $\mathbf{x}_i$ can be recursively computed by using the following set of KF equations [6]:

$$\boldsymbol{\Omega}_i = \mathbf{V}_i + \mathbf{H}_i\boldsymbol{\mathcal{P}}_{i-1}\mathbf{H}_i^T, \tag{2a}$$

$$\mathbf{K}_i = (\mathbf{F}_i\boldsymbol{\mathcal{P}}_{i-1}\mathbf{H}_i^T + \mathbf{G}_i\boldsymbol{\mathcal{S}}_i)\boldsymbol{\Omega}_i^{-1}, \tag{2b}$$

$$\mathbf{e}_i = \mathbf{z}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i-1}, \qquad \hat{\mathbf{x}}_i = \mathbf{F}_i\hat{\mathbf{x}}_{i-1} + \mathbf{K}_i\mathbf{e}_i, \tag{2c}$$

$$\boldsymbol{\mathcal{P}}_i = \mathbf{F}_i\boldsymbol{\mathcal{P}}_{i-1}\mathbf{F}_i^T + \mathbf{G}_i\mathbf{T}_i\mathbf{G}_i^T - \mathbf{K}_i\boldsymbol{\Omega}_i\mathbf{K}_i^T, \tag{2d}$$

where $\boldsymbol{\mathcal{P}}_{i-1}$ is the $M \times M$ error covariance matrix, i.e.,

$$\boldsymbol{\mathcal{P}}_{i-1} = \mathbb{E}\{(\mathbf{x}_{i-1} - \hat{\mathbf{x}}_{i-1})(\mathbf{x}_{i-1} - \hat{\mathbf{x}}_{i-1})^T\}, \tag{3}$$

$\mathbb{E}\{\cdot\}$ denotes expectation, $\mathbf{K}_i$ is the $M \times D$ Kalman gain, $\boldsymbol{\mathcal{S}}_i = \mathbb{E}\{\mathbf{t}_i\mathbf{v}_i^T\}$ and $\mathbf{e}_i$ is a $D \times 1$ error vector.

As we can see in equations (2b) and (2d), in general one needs $\mathcal{O}(M^3)$ operations to compute the Kalman gain and the covariance matrix $\boldsymbol{\mathcal{P}}_{i-1}$ or $\mathcal{O}(M^2)$ operations for a first-order random walk state-space model of the form

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{t}_i. \tag{4}$$

Depending on the application, this computational cost may be prohibitive. Variations on the Kalman recursions were proposed to reduce the computational complexity. The Schmidt-Kalman filter [7, Ch. 9] and the Chandrasekhar-Kailath-Morf-Sidhu (CKMS) filter [8, Ch. 11 and 13] are able to reduce the total number of arithmetical operations, but with complexity still $\mathcal{O}(M^2)$. Although the hardware technology for embedded systems is quite powerful, a KF with reduced complexity is important to deal with real-time systems that require high sampling rates and low latencies [9]. In the case of active noise control (ANC) [9] and multi-channel linear-prediction (MCLP) based for blind speech dereverberation [10], the KF tends to outperform most adaptive algorithms in terms of convergence speed and robustness. By enforcing a band-matrix structure for the covariance matrix $\boldsymbol{\mathcal{P}}_i$ in the case of [9] and block-diagonal matrix in the case of [10], the authors developed low complexity, i.e $\mathcal{O}(M)$, approximations for the KF.

The implementation of a Kalman filter also requires knowledge of a reasonable approximation to the state matrices $\mathbf{F}_i$, $\mathbf{G}_i$, $\mathbf{H}_i$, and covariance matrices $\mathbf{T}_i$ and $\mathbf{V}_i$, which may not be available in all applications. Model-free adaptive filters such as the least mean squares (LMS) and recursive least squares (RLS) algorithms [6], [11]–[13] can be implemented in $\mathcal{O}(M)$ complexity and do not require knowledge of the state matrices. Their performance, however is suboptimal with respect to that of the Kalman filter. The purpose of this paper is to study how well adaptive filters and their combinations can approximate the optimum performance of the Kalman filter in tracking time-varying parameter vectors.

As was shown in [14], through the convex combination between one LMS and one RLS filters (implemented with lattice [15] or Dichotomous Coordinate Descent — DCD [16] algorithms), it is possible to estimate the vector $\mathbf{x}_i$ modeled according to (4) with $\mathcal{O}(M)$ operations and with an excess mean square error (EMSE) $\mathbb{E}\{\|\mathbf{H}_i(\mathbf{x}_{i-1}-\hat{\mathbf{x}}_{i-1})\|^2\}$ less than 1dB from the solution obtained via KF. However, since the covariance matrix of $\mathbf{x}_i$ goes to infinity as $i \to \infty$, this model is unstable and does not reflect most practical situations. Given this scenario, this paper studies the tracking behavior of combinations of LMS and RLS filters using a more general model than (4) for the evolution of the optimum parameter vector $\mathbf{x}_i$. Our goal is to describe how close the combination scheme can get to the optimal excess mean-square error (EMSE) and mean-square deviation $\mathbb{E}\{\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2\}$ (MSD) obtained via Kalman filter while keeping the complexity linear in the filter length. In doing so we are able to derive how fast the parameter vector can change so that an adaptive filter can still track the variations.

The contributions of this paper are (the first three contributions are new with respect to [14]):

- Derivation of a theoretical recursion to estimate the covariance matrix $\boldsymbol{\mathcal{P}}_i$ and the corresponding steady-state EMSE and MSD of the Kalman filter considering the data model used in adaptive filtering, in which $\mathbf{H}_i$ is random (an alternate model for the steady-state MSD performance of the Kalman filter was given recently in [17]);
- The derivation of the fastest speed of change of the optimum parameter vector (in a sense to be described later in the text) that an adaptive filter can track. This result enables us to answer when a model-free adaptive filter can be used, and in which situations a model-dependent Kalman filter is necessary. We compare our bound with the nonstationarity degree (NSD) from [18];
- Derivation of a theoretical model for RLS and the convex combination of one LMS and one RLS filters under the autorregressive model (18) for the evolution of the optimum weight vector;
- The proposal of convex combinations of adaptive filters as a low-cost approximation to the Kalman filter, and theoretical expressions to quantify the quality of the approximation for the proposed model.

II. Combinations of Adaptive Filters

Combinations of adaptive filters enable a reduction in the sensitivity of the filter to choices of parameters such as the step-size, forgetting factor or filter length [13]. The idea is to combine the outputs of two (or several) different independently-run adaptive algorithms to achieve better performance than that of a single filter. In general, this approach is more robust than variable parameter schemes [19].

Due to its relative simplicity, the convex combination of adaptive filters was the first combined scheme that attracted attention, but other options are also available [20]–[22]. As shown in [23], it can be proved that the optimum combination is universal, i.e, the optimum combined estimate is at least as good as the best of the component filters in steady-state.

Several applications for combinations of adaptive filters have been proposed, such as acoustic echo cancellation [24], adaptive line enhancement [25], array beamforming [26], and active noise control [27]. Figure 1 illustrates a convex combination structure between two adaptive filters used to approximate a given desired variable $d(i)$ based on an input regressor vector $\mathbf{u}_i$.

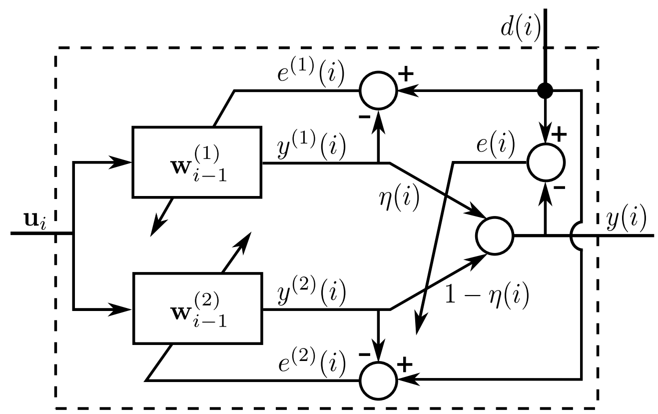


Figure 1. Convex combination of two transversal adaptive filters.

As shown in Figure 1, the output of the convex combination is computed according to

$$y(i) = \eta(i)y^{(1)}(i) + [1-\eta(i)]y^{(2)}(i), \tag{5}$$

where $\eta(i)$ is a mixing parameter that lies in $[0,1]$, $y^{(n)}(i) = \mathbf{u}_i^T\mathbf{w}_{i-1}^{(n)}$, for $n=1,2$, are the outputs of the transversal filters, $\mathbf{u}_i$ is the input regressor vector and $\mathbf{w}_{i-1}^{(n)}$ are the weight vectors of the component filters. The weight vector and the estimation error of the overall filter are given respectively by

$$\mathbf{w}_i = \eta(i)\mathbf{w}_i^{(1)} + [1-\eta(i)]\mathbf{w}_i^{(2)} \tag{6}$$

and

$$e(i) = d(i) - y(i) = \eta(i)e^{(1)}(i) + [1-\eta(i)]e^{(2)}(i). \tag{7}$$

In order to reduce the gradient noise when $\eta(i) \approx 0$ or $\eta(i) \approx 1$ and to ensure that $\eta(i)$ will remain between $[0,1]$, a nonlinear transformation of an auxiliary variable $a(i)$ can be used as follows:

$$\eta(i) = \frac{\text{sgm}[a(i)] - \text{sgm}[-a_+]}{\text{sgm}[a_+] - \text{sgm}[-a_+]}, \tag{8}$$

where $\text{sgm}(a) = \frac{1}{1+\mathrm{e}^{-a}}$, $a(i)$ is restricted to an interval $[-a_+, a_+]$ in order to avoid that the adaptation of $a(i)$ (see (9) below) slows down too much when $\eta(i)$ is close to 0 or to 1 because of the factor $\text{sgm}[a(i)]\{1-\text{sgm}[a(i)]\}$ [28]:

$$\begin{aligned} a(i+1) = a(i) + \\ + \frac{\mu_a}{\epsilon + p(i)} e(i)[y^{(1)}(i) - y^{(2)}(i)]\text{sgm}[a(i)]\{1-\text{sgm}[a(i)]\}. \end{aligned} \tag{9}$$

Here, ϵ is a small positive number and $p(i)$ is a low-pass filtered estimate of the power of $[y^{(1)}(i) - y^{(2)}(i)]$ defined by

$$p(i) = \gamma p(i-1) + (1-\gamma)[y^{(1)}(i) - y^{(2)}(i)]^2, \tag{10}$$

with $0 \ll \gamma < 1$.

The step-size μ_a usually will be chosen in the interval $[0.01, 1]$. As mentioned in [29], a common choice in practice for a_+ is 4.

It can be shown that the optimum mixing parameter in steady state is given by [23], [30]:

$$\eta^{\mathrm{o}} = \frac{\zeta^{(2)} - \zeta^{(12)}}{\zeta^{(1)} - 2\zeta^{(12)} + \zeta^{(2)}}, \tag{11}$$

where $\zeta^{(12)}$ is the cross EMSE between both filters in the convex combination which is obtained according to

$$\zeta^{(12)} = \lim_{i\to\infty} \mathbb{E}\{e_a^{(1)}(i)e_a^{(2)}(i)\}, \tag{12}$$

$\zeta^{(n)}$ is the steady-state excess mean-square error (EMSE) for each component filter, given by

$$\zeta^{(n)} = \lim_{i\to\infty} \mathbb{E}\{[e_a^{(n)}(i)]^2\}, \quad \text{for } n = 1, 2 \tag{13}$$

and $e_a^{(n)}(i)$ is the *a priori* error given by

$$e_a^{(n)}(i) = \mathbf{u}_i^T \tilde{\mathbf{w}}_{i-1}^{(n)}, \tag{14}$$

where $\tilde{\mathbf{w}}_{i-1}^{(n)} = \mathbf{w}_{i-1}^{\mathrm{o}} - \mathbf{w}_{i-1}^{(n)}$ and $\mathbf{w}_i^{\mathrm{o}}$ is the optimum weight vector at time i (the adopted model for its time evolution is given in (18) below).

In equations (13) and (12), the *a priori* error $e_a(i)$ of the overall combination structure can be written as a function of the *a priori* errors of the component filters [19], i.e,

$$e_a(i) = \eta(i)e_a^{(1)}(i) + [1 - \eta(i)]e_a^{(2)}(i). \tag{15}$$

As mentioned in [29], the optimum EMSE for the convex combination between two adaptive filters, using the optimum mixing parameter η^{o}, is given by

$$\zeta^{\mathrm{COMB}} = \zeta^{(12)} + \frac{\Delta\zeta^{(1)}\Delta\zeta^{(2)}}{\Delta\zeta^{(1)} + \Delta\zeta^{(2)}}, \tag{16}$$

where $\Delta\zeta^{(n)} = \zeta^{(n)} - \zeta^{(12)}$, for $n = 1, 2$.

This paper is organized as follows: in Section III we derive a general expression for the covariance matrix (23) considering algorithms of the general class given by (20) and in Section IV we derive the EMSE and MSD for the LMS and RLS algorithms, as well as their respective combination and for the Kalman filter, considering the proposed model (18). Section V compares the performance of each algorithm under different conditions, and finally, section VI concludes the paper.

III. Tracking model

Consider a nonstationary data model relating the random sequences $d(i)$ and $\mathbf{u}_i$ through a linear model of the form

$$d(i) = \mathbf{u}_i^T \mathbf{w}_{i-1}^{\mathrm{o}} + v(i), \tag{17}$$

where $v(i)$ is a zero-mean random variable with constant variance $\sigma_v^2 = \mathbb{E}\{v(i)\}^2$ and uncorrelated with $\mathbf{u}_i$. The weight vector $\mathbf{w}_i^{\mathrm{o}}$ is assumed to evolve according to [31]

$$\mathbf{w}_i^{\mathrm{o}} = \theta\mathbf{w}_{i-1}^{\mathrm{o}} + \sqrt{1-\theta^2}\mathbf{q}_i, \tag{18}$$

where θ is a scalar variable in the range $0 < \theta \leq 1$ and $\mathbf{q}_i$ is a stationary random perturbation independent of the (zero mean) initial conditions $\{\mathbf{w}_{-1}^{\mathrm{o}}, \mathbf{w}_{-1}\}$, of $\mathbf{u}_j$ for all j and of $d(j)$ for all $j < i$. In order to keep the power of the output signal $\mathbf{w}_i^{\mathrm{o}}$ independent of θ and be able to study the EMSE behavior of each filter according to the variation of θ, the model (18) was defined in such a way that the power of $\mathbf{w}_i^{\mathrm{o}}$ in the limit is independent of θ, i.e, $\lim_{i\to\infty} \mathbb{E}\|\mathbf{w}_i^{\mathrm{o}}\|^2 = \mathbb{E}\|\mathbf{q}_i\|^2$. The following assumptions are also considered for this model:

- The sequence $\{\mathbf{u}_i\}$ is real and zero mean, and such that $\mathbf{u}_i$ is independent of $\mathbf{u}_j$ for $i \neq j$ (i.e., the sequence is i.i.d. — independent and identically distributed).
- The noise sequence $v(i)$ is i.i.d.
- The autocorrelation matrix $\mathbf{R} = \mathbb{E}\{\mathbf{u}_i\mathbf{u}_i^T\}$ is positive-definite ($\mathbf{R} > 0$).
- The random sequences $\mathbf{u}_i$ and $v(i)$ are jointly Gaussian.
- The sequence $\mathbf{q}_i$ is i.i.d., zero-mean and with positive-definite autocorrelation matrix equal to

$$\mathbb{E}\{\mathbf{q}_i\mathbf{q}_i^T\} = \mathbf{Q}. \tag{19}$$

- The regressor vector $\mathbf{u}_i$ is independent of the weight-error vector $\tilde{\mathbf{w}}_{i-1}$. This condition is an approximation, part of the widely used independence assumptions in adaptive filter theory [11], [13].

We focus on the convex combination of two algorithms of the following general class [13]:

$$\mathbf{w}_i^{(n)} = \mathbf{w}_{i-1}^{(n)} + \rho^{(n)}\mathbf{M}_i^{(n)}\mathbf{u}_i e^{(n)}(i), \tag{20}$$

where the superscript n is associated to the first ($n = 1$) or second ($n = 2$) filter of the combination, $\mathbf{w}_i^{(n)}$ represents the length-M coefficient vector, $\rho^{(n)}$ is a step-size (which is equal to $\mu^{(n)}$ for LMS or 1 for RLS), $\mathbf{u}_i$ is the input regressor vector, $e^{(n)}(i)$ is the estimation error given by $d(i) - \mathbf{u}_i^T\mathbf{w}_{i-1}^{(n)}$ and $\mathbf{M}_i^{(n)}$ is an M-by-M symmetric nonsingular matrix equal to the identity matrix for LMS or, in the RLS case, equal to $\mathbf{P}_i$

$$\mathbf{P}_i = \mathbf{R}_i^{-1} = \left(\nu\mathbf{I} + \sum_{\ell=0}^{i} \lambda^{i-\ell}\mathbf{u}_\ell\mathbf{u}_\ell^T\right)^{-1}, \tag{21}$$

where $\mathbf{I}$ denotes the $M \times M$ identity matrix, λ is the forgetting factor and $\nu\mathbf{I}$ is an initial condition to guarantee invertibility.

It is common in the literature to evaluate the steady-state MSD (ε) and EMSE (ζ) for one adaptive filter as

$$\varepsilon = \lim_{i\to\infty} \mathrm{Tr}\{\mathbf{S}_i\}, \qquad \zeta = \lim_{i\to\infty} \mathrm{Tr}\{\mathbf{R}\mathbf{S}_i\}, \tag{22}$$

where $\mathrm{Tr}\{\mathbf{A}\}$ stands for the trace of matrix $\mathbf{A}$ and

$$\mathbf{S}_i = \mathbb{E}\{\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^T\}. \tag{23}$$

As shown in [23], convex combinations of algorithms of the form (20) are universal in the mean-square error sense, that is, the performance of the combination (if the optimum mixing parameter $\eta(i)$ is used) is always at least as good as that of the best individual filter. Practical algorithms to estimate $\eta(i)$ are described in [19], [23].

As Eweda shows in [32], in problems where it is necessary to track time-varying parameters, either LMS or RLS may present the best tracking performance, depending on the statistics of the regressor and the desired signals. In this case,

[19] shows that there are three possible situations that can occur in steady-state for the combination: the combination will closely follow LMS or RLS if one of these filters significantly outperforms the other, or, if both component filters have similar performance, the combination may actually outperform both of them.

Tracking analyses of convex combinations of the algorithms of the form (20) depend on analytical expressions of the cross-MSD and cross-EMSE. Using independence assumptions, such expressions can be obtained through the evaluation of

$$\varepsilon^{(12)} = \lim_{i\to\infty} \mathrm{Tr}\{\mathbf{S}_i^{(12)}\}, \quad \zeta^{(12)} = \lim_{i\to\infty} \mathrm{Tr}\{\mathbf{R}\mathbf{S}_i^{(12)}\}, \quad (24)$$

where

$$\mathbf{S}_i^{(12)} = \mathbb{E}\{\tilde{\mathbf{w}}_i^{(1)}(\tilde{\mathbf{w}}_i^{(2)})^T\}. \quad (25)$$

It is possible to modify (11) to obtain the value of the mixing parameter $\eta^{\mathrm{o}}_{\mathrm{MSD}}$ that minimizes the MSD of the combination. Unfortunately, estimating $\eta^{\mathrm{o}}_{\mathrm{MSD}}$ online is not feasible, so we define the combination MSD using η^{o} from (11), which is optimum for the MSE and EMSE and suboptimal for the MSD. Subtracting (6) from $\mathbf{w}_i^{\mathrm{o}}$, it can be shown that

$$\varepsilon^{\mathrm{COMB}} = \eta^{\mathrm{o}2}\varepsilon^{(1)} + (1-\eta^{\mathrm{o}})^2\varepsilon^{(2)} + 2\eta^{\mathrm{o}}(1-\eta^{\mathrm{o}})\varepsilon^{(12)}. \quad (26)$$

The main focus of the following analysis is on the tracking behavior of the adaptive filter in steady-state or, in other words, after initial convergence of the coefficients $\mathbf{w}_i^{(n)}$. Although the optimum weights are time-varying, under model (18), the MSD and EMSE approach a steady-state value as seen in Section IV.

We start the analysis by subtracting both sides of (20) from $\mathbf{w}_i^{\mathrm{o}}$, to get

$$\tilde{\mathbf{w}}_i^{(n)} = \mathbf{w}_i^{\mathrm{o}} - \mathbf{w}_{i-1}^{(n)} - \rho^{(n)}\mathbf{M}_i^{(n)}\mathbf{u}_i e^{(n)}(i). \quad (27)$$

Using the linear regression model of (17) for the desired response $d(i)$ and the *a priori* error signal $e_a^{(n)}(i)$ given by (14), the error signal defined as $e^{(n)}(i) = d(i) - \mathbf{u}_i^T\mathbf{w}_{i-1}^{(n)}$ with $n = 1, 2$, can be rewritten as:

$$e_i^{(n)} = e_a^{(n)}(i) + v(i). \quad (28)$$

Replacing the model (18) and the error signal $e_i^{(n)}$ given by (28) into equation (27), we arrive at:

$$\begin{aligned}\tilde{\mathbf{w}}_i^{(n)} &= \theta\mathbf{w}_{i-1}^{\mathrm{o}} + \sqrt{1-\theta^2}\mathbf{q}_i - \mathbf{w}_{i-1}^{(n)} \\ &\quad - \rho^{(n)}\mathbf{M}_i^{(n)}\mathbf{u}_i[e_a^{(n)}(i) + v(i)].\end{aligned} \quad (29)$$

By adding and subtracting $(1-\theta)\mathbf{w}_{i-1}^{\mathrm{o}}$ in (29) and replacing $e_a^{(n)}(i)$ by (14), we get:

$$\begin{aligned}\tilde{\mathbf{w}}_i^{(n)} &= \left[\mathbf{I} - \rho^{(n)}\mathbf{M}_i^{(n)}\mathbf{u}_i\mathbf{u}_i^T\right]\tilde{\mathbf{w}}_{i-1}^{(n)} - \rho^{(n)}\mathbf{M}_i^{(n)}\mathbf{u}_i v(i) \\ &\quad - (1-\theta)\mathbf{w}_{i-1}^{\mathrm{o}} + \sqrt{1-\theta^2}\mathbf{q}_i.\end{aligned} \quad (30)$$

In order to compute the covariance matrices (23) and (25) multiply (30) with $n = \ell$ by its transpose with $n = m$ and take the expectation of both sides. Assuming that $\mathbf{q}_i$ is independent of the initial conditions and of $\mathbf{u}_i$, after some algebra, we get

$$\begin{aligned}\mathbb{E}\{\tilde{\mathbf{w}}_i^{(\ell)}(\tilde{\mathbf{w}}_i^{(m)})^T\} &= \mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\} \\ &- \rho^{(\ell)}\mathbb{E}\{\mathbf{M}_i^{(\ell)}\mathbf{u}_i\mathbf{u}_i^T\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\} \\ &- \rho^{(m)}\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\mathbf{u}_i\mathbf{u}_i^T\mathbf{M}_i^{(m)}\} \\ &- (1-\theta)\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} \\ &+ \rho^{(\ell)}\rho^{(m)}\mathbb{E}\{\mathbf{M}_i^{(\ell)}\mathbf{u}_i\mathbf{u}_i^T\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\mathbf{u}_i\mathbf{u}_i^T\mathbf{M}_i^{(m)}\} \\ &- (1-\theta)\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\} \\ &+ (1-\theta)\rho^{(\ell)}\mathbb{E}\{\mathbf{M}_i^{(\ell)}\mathbf{u}_i\mathbf{u}_i^T\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} \\ &+ \rho^{(\ell)}\rho^{(m)}\sigma_v^2\mathbb{E}\{\mathbf{M}_i^{(\ell)}\mathbf{u}_i\mathbf{u}_i^T\mathbf{M}_i^{(m)}\} \\ &+ (1-\theta)\rho^{(m)}\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\mathbf{u}_i\mathbf{u}_i^T\mathbf{M}_i^{(m)}\} \quad (31) \\ &+ (1-\theta)^2\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} + (1-\theta^2)\mathbb{E}\{\mathbf{q}_i\mathbf{q}_i^T\}.\end{aligned}$$

To simplify equation (31), the following assumptions will be considered:

- **Assumption 1:** The regressor vector $\mathbf{u}_i$ is independent of the weight-vector $\tilde{\mathbf{w}}_{i-1}^{(n)}$ for $n = 1, 2$;
- **Assumption 2:** Matrix $\mathbf{M}_i^{(n)}$ varies slowly in relation to $\tilde{\mathbf{w}}_{i-1}^{(n)}$. Thus, when $\mathbf{M}_i^{(n)}$ appears inside the expectations of (31), we simply replace it by its mean $\bar{\mathbf{M}}^{(n)}$. For LMS this assumption is not necessary, since $\mathbf{M}_i^{(n)} = \mathbf{I}$. For RLS, considering large enough i, we can approximate $\mathbb{E}\{\mathbf{M}_i^{(n)}\}$ by $\mathbb{E}\{\mathbf{P}_i\} \approx \bar{\mathbf{P}} \triangleq (1-\lambda)\mathbf{R}^{-1}$.
- **Assumption 3:** According to [13] and relation (25), the value of $\mathbb{E}\{\mathbf{M}_i^{(\ell)}\mathbf{u}_i\mathbf{u}_i^T\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\mathbf{u}_i\mathbf{u}_i^T\mathbf{M}_i^{(m)}\}$ can be approximated by $\bar{\mathbf{M}}^{(\ell)}\left\{\mathbf{R}\mathrm{Tr}\{\mathbf{R}\mathbf{S}_{i-1}^{(\ell m)}\} + 2\mathbf{R}\mathbf{S}_{i-1}^{(\ell m)}\mathbf{R}\right\}\bar{\mathbf{M}}^{(m)}$ when the input regressor vector $\mathbf{u}_i$ is Gaussian and real.

By considering the above assumptions and the relation (25), equation (31) simplifies to:

$$\begin{aligned}\mathbf{S}_i^{(\ell m)} &\cong \mathbf{S}_{i-1}^{(\ell m)} - \rho^{(m)}\mathbf{S}_{i-1}^{(\ell m)}\mathbf{R}\bar{\mathbf{M}}^{(m)} - \rho^{(\ell)}\bar{\mathbf{M}}^{(\ell)}\mathbf{R}\mathbf{S}_{i-1}^{(\ell m)} \\ &+ \rho^{(\ell)}\rho^{(m)}\bar{\mathbf{M}}^{(\ell)}\left\{\mathbf{R}\mathrm{Tr}\{\mathbf{R}\mathbf{S}_{i-1}^{(\ell m)}\} + 2\mathbf{R}\mathbf{S}_{i-1}^{(\ell m)}\mathbf{R}\right\}\bar{\mathbf{M}}^{(m)} \\ &+ \rho^{(\ell)}\rho^{(m)}\sigma_v^2\bar{\mathbf{M}}^{(\ell)}\mathbf{R}\bar{\mathbf{M}}^{(m)} \\ &+ (1-\theta)\left[\rho^{(\ell)}\bar{\mathbf{M}}^{(\ell)}\mathbf{R} - \mathbf{I}\right]\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(\ell)}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} \\ &+ (1-\theta)\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\tilde{\mathbf{w}}_{i-1}^{(m)})^T\}\left[\rho^{(m)}\mathbf{R}\bar{\mathbf{M}}^{(m)} - \mathbf{I}\right] \\ &+ (1-\theta)^2\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} + (1-\theta^2)\mathbf{Q}.\end{aligned} \quad (32)$$

As required in (32), the terms $\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\}$ and $\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(n)}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\}$ can be obtained as follows:

$$\begin{aligned}&\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} = \\ &= \mathbb{E}\{(\theta\mathbf{w}_{i-2}^{\mathrm{o}} + \sqrt{1-\theta^2}\mathbf{q}_{i-1})(\theta\mathbf{w}_{i-2}^{\mathrm{o}} + \sqrt{1-\theta^2}\mathbf{q}_{i-1})^T\} \\ &= \theta^2\mathbb{E}\{\mathbf{w}_{i-2}^{\mathrm{o}}(\mathbf{w}_{i-2}^{\mathrm{o}})^T\} + (1-\theta^2)\mathbb{E}\{\mathbf{q}_{i-1}\mathbf{q}_{i-1}^T\} \\ &= \theta^2\mathbb{E}\{\mathbf{w}_{i-2}^{\mathrm{o}}(\mathbf{w}_{i-2}^{\mathrm{o}})^T\} + (1-\theta^2)\mathbf{Q}.\end{aligned} \quad (33)$$

Assuming that the adaptive filter is in steady-state, i.e.:

$$\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} = \mathbb{E}\{\mathbf{w}_{i-2}^{\mathrm{o}}(\mathbf{w}_{i-2}^{\mathrm{o}})^T\}, \quad \text{as} \quad i\to\infty, \quad (34)$$

equation (33) will converge to:

$$\lim_{i\to\infty}\mathbb{E}\{\mathbf{w}_{i-1}^{\mathrm{o}}(\mathbf{w}_{i-1}^{\mathrm{o}})^T\} = \mathbf{Q}. \quad (35)$$

The term $\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(n)}(\mathbf{w}_{i-1}^{\rm o})^T\}$ can be obtained by taking the expectation of the product between equation (30) and the model (18), both at time $i-1$, i.e.:

$$\begin{aligned}\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(n)}(\mathbf{w}_{i-1}^{\rm o})^T\} &= \mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(n)}(\theta\mathbf{w}_{i-2}^{\rm o}+\sqrt{1-\theta^2}\mathbf{q}_{i-1})^T\}\\ &= \theta\mathbb{E}\{\tilde{\mathbf{w}}_{i-2}^{(n)}(\mathbf{w}_{i-2}^{\rm o})^T\}\\ &- \rho^{(n)}\theta\mathbb{E}\{\mathbf{M}_{i-1}^{(n)}\mathbf{u}_{i-1}\mathbf{u}_{i-1}^T\tilde{\mathbf{w}}_{i-2}^{(n)}(\mathbf{w}_{i-2}^{\rm o})^T\}\\ &- \rho^{(n)}\theta\mathbb{E}\{v(i-1)\mathbf{M}_{i-1}^{(n)}\mathbf{u}_{i-1}(\mathbf{w}_{i-2}^{\rm o})^T\}\\ &- \theta(1-\theta)\mathbb{E}\{\mathbf{w}_{i-2}^{\rm o}(\mathbf{w}_{i-2}^{\rm o})^T\}\\ &+ \sqrt{1-\theta^2}\left[\theta\mathbb{E}\{\mathbf{q}_{i-1}(\mathbf{w}_{i-2}^{\rm o})^T\}+\mathbb{E}\{\tilde{\mathbf{w}}_{i-2}^{(n)}\mathbf{q}_{i-1}^T\}\right]\\ &- \rho^{(n)}\sqrt{1-\theta^2}\mathbb{E}\{\mathbf{M}_{i-1}^{(n)}\mathbf{u}_{i-1}\mathbf{u}_{i-1}^T\tilde{\mathbf{w}}_{i-2}^{(n)}\mathbf{q}_{i-1}^T\}\\ &- \rho^{(n)}\sqrt{1-\theta^2}\mathbb{E}\{v(i-1)\mathbf{M}_{i-1}^{(n)}\mathbf{u}_{i-1}\mathbf{q}_{i-1}^T\}\\ &- (1-\theta)\sqrt{1-\theta^2}\mathbb{E}\{\mathbf{w}_{i-2}^{\rm o}\mathbf{q}_{i-1}^T\}\\ &+ (1-\theta^2)\mathbb{E}\{\mathbf{q}_{i-1}\mathbf{q}_{i-1}^T\}.\end{aligned}$$

Since $\mathbf{q}_{i-1}$ and $v(i-1)$ are zero-mean and $\mathbf{u}_{i-1}$ is independent of $\tilde{\mathbf{w}}_{i-2}^{(n)}$ and $\mathbf{w}_{i-2}^{\rm o}$, then

$$\begin{aligned}\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(n)}(\mathbf{w}_{i-1}^{\rm o})^T\} &= \theta\mathbb{E}\{\tilde{\mathbf{w}}_{i-2}^{(n)}(\mathbf{w}_{i-2}^{\rm o})^T\}\\ &- \rho^{(n)}\theta\bar{\mathbf{M}}^{(n)}\mathbf{R}\mathbb{E}\{\tilde{\mathbf{w}}_{i-2}^{(n)}(\mathbf{w}_{i-2}^{\rm o})^T\}\\ &- \theta(1-\theta)\mathbb{E}\{\mathbf{w}_{i-2}^{\rm o}(\mathbf{w}_{i-2}^{\rm o})^T\}+(1-\theta^2)\mathbf{Q}.\end{aligned} \tag{36}$$

Assuming that the filter is operating in steady-state, i.e

$$\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(n)}(\mathbf{w}_{i-1}^{\rm o})^T\} = \mathbb{E}\{\tilde{\mathbf{w}}_{i-2}^{(n)}(\mathbf{w}_{i-2}^{\rm o})^T\}, \quad \text{as} \quad i\to\infty \tag{37}$$

and considering that the eigenvalues of $\theta(\mathbf{I}-\rho^{(n)}\bar{\mathbf{M}}^{(n)}\mathbf{R})$ lie between (-1,1), equation (36) will converge to

$$\lim_{i\to\infty}\mathbb{E}\{\tilde{\mathbf{w}}_{i-1}^{(n)}(\mathbf{w}_{i-1}^{\rm o})^T\} = (1-\theta)\mathbf{\Psi}^{-1}\mathbf{Q}, \tag{38}$$

where $\mathbf{\Psi} = (1-\theta)\mathbf{I}+\rho^{(n)}\theta\bar{\mathbf{M}}^{(n)}\mathbf{R}$.

IV. Theoretical steady-state analysis

Based on the previous results obtained from model (18), the next subsections IV-A to IV-D presents the theoretical steady-state EMSE for four different situations, namely: one individual LMS filter, one individual RLS filter, their convex combination and finally for the Kalman filter.

A. Theoretical steady-state EMSE for LMS

To compute the EMSE for an individual LMS filter, substitute $\ell = m = 1$ in (32), considering $\bar{\mathbf{M}}^{(1)} = \mathbf{I}$ and $\rho^{(1)} = \mu$ and assuming that the filter is operating in steady-state, i.e $i\to\infty$, we get the following recursion for $\mathbf{S}_\infty^{(11)}$ after replacing equations (35) and (38) in (32):

$$\begin{aligned}\mathbf{S}_\infty^{(11)} &\cong \mathbf{S}_\infty^{(11)} - \mu[\mathbf{S}_\infty^{(11)}\mathbf{R}+\mathbf{R}\mathbf{S}_\infty^{(11)}]\\ &+ \mu^2\left\{\mathbf{R}\text{Tr}\{\mathbf{R}\mathbf{S}_\infty^{(11)}\}+2\mathbf{R}\mathbf{S}_\infty^{(11)}\mathbf{R}\right\}+\mu^2\sigma_v^2\mathbf{R}\\ &+ (1-\theta)^2[\mu\mathbf{R}-\mathbf{I}][(1-\theta)\mathbf{I}+\mu\theta\mathbf{R}]^{-1}\mathbf{Q}\\ &+ (1-\theta)^2\mathbf{Q}[(1-\theta)\mathbf{I}+\mu\theta\mathbf{R}]^{-1}[\mu\mathbf{R}-\mathbf{I}]\\ &+ 2(1-\theta)\mathbf{Q}.\end{aligned} \tag{39}$$

To simplify (39) and get an easier expression to deal with, we multiply and divide by θ the second and third lines of (39) and reorganize the terms to obtain

$$\begin{aligned}\mathbf{S}_\infty^{(11)} &\cong \mathbf{S}_\infty^{(11)} - \mu[\mathbf{S}_\infty^{(11)}\mathbf{R}+\mathbf{R}\mathbf{S}_\infty^{(11)}]\\ &+ \mu^2\left\{\mathbf{R}\text{Tr}\{\mathbf{R}\mathbf{S}_\infty^{(11)}\}+2\mathbf{R}\mathbf{S}_\infty^{(11)}\mathbf{R}\right\}+\mu^2\sigma_v^2\mathbf{R}\\ &+ \theta^{-1}(1-\theta)^2[\mu\theta\mathbf{R}-\theta\mathbf{I}][\mathbf{I}+\mu\theta\mathbf{R}-\theta\mathbf{I}]^{-1}\mathbf{Q}\\ &+ \theta^{-1}(1-\theta)^2\mathbf{Q}[\mathbf{I}+\mu\theta\mathbf{R}-\theta\mathbf{I}]^{-1}[\mu\theta\mathbf{R}-\theta\mathbf{I}]\\ &+ 2(1-\theta)\mathbf{Q}.\end{aligned} \tag{40}$$

By using the push-through identity property $\mathbf{A}(\mathbf{I}+\mathbf{A})^{-1} = (\mathbf{I}+\mathbf{A})^{-1}\mathbf{A} = \mathbf{I}-(\mathbf{I}+\mathbf{A})^{-1}$, where $\mathbf{A} = \mu\theta\mathbf{R}-\theta\mathbf{I}$, after some algebra, we obtain the following simplified version of equation (39)

$$\begin{aligned}\mathbf{S}_\infty^{(11)} &\cong \mathbf{S}_\infty^{(11)} - \mu[\mathbf{S}_\infty^{(11)}\mathbf{R}+\mathbf{R}\mathbf{S}_\infty^{(11)}]\\ &+ \mu^2\left\{\mathbf{R}\text{Tr}\{\mathbf{R}\mathbf{S}_\infty^{(11)}\}+2\mathbf{R}\mathbf{S}_\infty^{(11)}\mathbf{R}\right\}+\mu^2\sigma_v^2\mathbf{R}\\ &- \theta^{-1}(1-\theta)^2[(1-\theta)\mathbf{I}+\mu\theta\mathbf{R}]^{-1}\mathbf{Q}\\ &- \theta^{-1}(1-\theta)^2\mathbf{Q}[(1-\theta)\mathbf{I}+\mu\theta\mathbf{R}]^{-1}\\ &+ 2\theta^{-1}(1-\theta)\mathbf{Q}.\end{aligned} \tag{41}$$

Based on (41), instead of computing the steady-state EMSE for LMS in a direct way, we follow the same steps of [13] and define the rotated matrix given by $\bar{\mathbf{S}}_\infty^{(11)} = \mathbf{U}^T\mathbf{S}_\infty^{(11)}\mathbf{U}$, where $\mathbf{U}$ is an orthogonal matrix that diagonalizes $\mathbf{R}$, that is

$$\mathbf{U}^T\mathbf{R}\mathbf{U} = \text{diag}(\lambda_i) \triangleq \mathbf{\Lambda} \tag{42}$$

where diag(λ_i) is a diagonal matrix formed with the eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_M$ of $\mathbf{R}$.

A recursion for $\bar{\mathbf{S}}_\infty^{(11)}$ can be obtained by multiplying (41) from the left by $\mathbf{U}^T$ and from the right by $\mathbf{U}$. Defining the rotated matrix $\bar{\mathbf{Q}} = \mathbf{U}^T\mathbf{Q}\mathbf{U}$ and recalling that $\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^T$ and $\mathbf{I} = \mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U}$, we get after simplifications

$$\begin{aligned}\bar{\mathbf{S}}_\infty^{(11)} &\cong \bar{\mathbf{S}}_\infty^{(11)} - \mu\{\bar{\mathbf{S}}_\infty^{(11)}\mathbf{\Lambda}+\mathbf{\Lambda}\bar{\mathbf{S}}_\infty^{(11)}\}\\ &+ \mu^2\{\mathbf{\Lambda}\text{Tr}\{\mathbf{\Lambda}\bar{\mathbf{S}}_\infty^{(11)}\}+2\mathbf{\Lambda}\bar{\mathbf{S}}_\infty^{(11)}\mathbf{\Lambda}\}+\mu^2\sigma_v^2\mathbf{\Lambda}\\ &- \theta^{-1}(1-\theta)^2[(1-\theta)\mathbf{I}+\mu\theta\mathbf{\Lambda}]^{-1}\bar{\mathbf{Q}}\\ &- \theta^{-1}(1-\theta)^2\bar{\mathbf{Q}}[(1-\theta)\mathbf{I}+\mu\theta\mathbf{\Lambda}]^{-1}\\ &+ 2\theta^{-1}(1-\theta)\bar{\mathbf{Q}}.\end{aligned} \tag{43}$$

Using the rotated matrix $\bar{\mathbf{S}}_\infty^{(11)}$, the steady-state EMSE can be computed as Tr$\{\mathbf{\Lambda}\bar{\mathbf{S}}_\infty^{(11)}\}$ and so, it depends only on the diagonal entries of $\bar{\mathbf{S}}_\infty^{(11)}$. We can work therefore only with these diagonal entries and define the vectors

$$\bar{\mathbf{s}}_\infty^{(11)} = \text{diag}\{\bar{\mathbf{S}}_\infty^{(11)}\} \text{ and } \boldsymbol{\ell} = \text{diag}\{\mathbf{\Lambda}\}, \tag{44}$$

where diag$\{\mathbf{A}\}$ represents a column vector with the diagonal elements of $\mathbf{A}$.

Given this, by applying the diagonal operator to both sides of equation (43) and taking the limit as $i\to\infty$ we obtain

$$\begin{aligned}\bar{\mathbf{s}}_\infty^{(11)} &= \left[2\mu\mathbf{\Lambda}-\mu^2\boldsymbol{\ell}\boldsymbol{\ell}^T-2\mu^2\mathbf{\Lambda}^2\right]^{-1}\Bigg\{\mu^2\sigma_v^2\boldsymbol{\ell}\\ &+ \frac{2(1-\theta)}{\theta}\left\{\mathbf{I}-\left[\mathbf{I}+\frac{\mu\theta}{1-\theta}\mathbf{\Lambda}\right]^{-1}\right\}\text{diag}\{\bar{\mathbf{Q}}\}\Bigg\}.\end{aligned} \tag{45}$$

The steady-state EMSE and MSD can now be computed as

$$\zeta^{\text{LMS}} = \boldsymbol{\ell}^T \bar{\mathbf{s}}_\infty^{(11)}, \qquad \varepsilon^{\text{LMS}} = \mathbb{1}^T \bar{\mathbf{s}}_\infty^{(11)}, \tag{46}$$

where $\mathbb{1} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$.

Due to the complexity of (45), we present next an approximation valid for sufficiently small μ in order to better describe the qualitative behavior of the LMS filter. We use the analysis below to show that there is a minimum value for θ, $\theta_{\text{min}}^{\text{LMS}}$, below which the filter is no longer able to track variations in $\mathbf{w}_i^{\text{o}}$. Assuming that the term $\mu^2 \left\{ \mathbf{R}\text{Tr}\{\mathbf{R}\mathbf{S}_\infty^{(11)}\} + 2\mathbf{R}\mathbf{S}_\infty^{(11)}\mathbf{R} \right\}$ can be neglected with respect to the three first terms on the right-hand side of (41) and applying the trace operator to both sides of this equation, ζ^{LMS} reduces to

$$\begin{aligned}\zeta^{\text{LMS}} = \text{Tr}\{\mathbf{R}\mathbf{S}_\infty^{(11)}\} \cong \frac{1}{2\mu} \Big\{ & \mu^2\sigma_v^2 \text{Tr}\{\mathbf{R}\} \\ & - 2\theta^{-1}(1-\theta)^2 \text{Tr}\left\{[(1-\theta)\mathbf{I} + \mu\theta\mathbf{R}]^{-1}\mathbf{Q}\right\} \\ & + 2\theta^{-1}(1-\theta)\text{Tr}\{\mathbf{Q}\} \Big\},\end{aligned} \tag{47}$$

where we used the property $\text{Tr}\{\mathbf{AB}\} = \text{Tr}\{\mathbf{BA}\}$.

Applying the matrix inversion lemma [6] to the term $[(1-\theta)\mathbf{I} + \mu\theta\mathbf{R}]^{-1}$ in (47), after some algebra, we obtain the following approximate expression

$$\zeta^{\text{LMS}} \cong \frac{\mu\sigma_v^2 \text{Tr}\{\mathbf{R}\}}{2} + (1-\theta)\text{Tr}\left\{[(1-\theta)\mathbf{I} + \mu\theta\mathbf{R}]^{-1}\mathbf{R}\mathbf{Q}\right\}. \tag{48}$$

The μ^{o} that minimizes (48) can be obtained by setting the first derivative of ζ^{LMS} equal to zero. When $\theta = 1$, ζ^{LMS} reduces to a linear function of μ given by

$$\zeta^{\text{LMS}} \cong \left.\frac{\mu\sigma_v^2 \text{Tr}\{\mathbf{R}\}}{2}\right|_{\theta=1} \tag{49}$$

and so, the optimum step-size that minimizes (49) is $\mu^{\text{o}} = 0$ with the corresponding minimum EMSE $\zeta_{\text{LMS}}^{\text{o}} = 0$[1].

In order to compute the first derivative of (48) for $0 \leq \theta < 1$, we rewrite the terms of (48) as a sum of scalars. To do this, we use the same decomposition for matrix $\mathbf{R}$ described in (42) and the same matrix $\bar{\mathbf{Q}}$ used in (43) to obtain

$$\begin{aligned}\zeta^{\text{LMS}} &\cong \frac{\mu\sigma_v^2 \text{Tr}\{\mathbf{\Lambda}\}}{2} + (1-\theta)\text{Tr}\left\{[(1-\theta)\mathbf{I} + \mu\theta\mathbf{\Lambda})]^{-1}\mathbf{\Lambda}\bar{\mathbf{Q}})\right\} \\ &= \sum_{i=1}^{M} \frac{\mu\sigma_v^2\lambda_{ii}}{2} + \sum_{i=1}^{M} \frac{(1-\theta)\lambda_{ii}\bar{q}_{ii}}{1-\theta+\mu\theta\lambda_{ii}},\end{aligned} \tag{50}$$

where $\bar{q}_{ii}$ are the diagonal elements of $\bar{\mathbf{Q}}$.

By taking the first derivative of (50) and setting it equal to 0, after simplifications, we find the following expression to compute the optimum step-size $\mu = \mu^{\text{o}}$ that minimizes the steady-state EMSE for LMS

$$\sum_{i=1}^{M} \left\{ \frac{\theta(1-\theta)\lambda_{ii}^2\bar{q}_{ii}}{(1-\theta+\mu\theta\lambda_{ii})^2} - \frac{\sigma_v^2\lambda_{ii}}{2} \right\} = 0. \tag{51}$$

[1]The reader may find this result puzzling, since the tracking model usually seen in the literature includes a term proportional to μ^{-1}, resulting in a positive optimum step size [6], [13]. Our result stems from the factor $\sqrt{1-\theta^2}$ in our model (18), which we included to keep the power of $\mathbf{w}_i^{\text{o}}$ independent of θ. Note that for $\theta = 1$, the optimum step size is $\mu^{\text{o}} = 0$ because the model (18) implies that $\mathbb{E}\{\mathbf{w}_i^{\text{o}}\} = \mathbf{0}$, and the initial condition in our model also satisfies $\mathbb{E}\{\mathbf{w}_{-1}\} = \mathbf{0}$.

Since in general a closed solution for μ in (51) is not possible (if the λ_{ii} are distinct, (51) would reduce to finding the roots of a $2M$-degree polynomial in μ), we seek solutions for $\theta \in [0, 1[$ numerically. Given this, let us start by naming $f(\mu)$ as the left-hand side of (51), i.e:

$$f(\mu) = \sum_{i=1}^{M} \left\{ \frac{\theta(1-\theta)\lambda_{ii}^2\bar{q}_{ii}}{(1-\theta+\mu\theta\lambda_{ii})^2} - \frac{\sigma_v^2\lambda_{ii}}{2} \right\}.$$

Since we are seeking conditions under which $f(\mu) = 0$ has a positive solution, we may start our analysis by finding the corresponding boundaries of $f(\mu)$ when $\mu \to 0$ and $\mu \to \infty$, which are:

$$\lim_{\mu\to 0} f(\mu) = \frac{\theta\text{Tr}\{\mathbf{\Lambda}^2\bar{\mathbf{Q}}\}}{1-\theta} - \frac{\sigma_v^2\text{Tr}\{\mathbf{\Lambda}\}}{2}, \qquad \text{and}$$

$$\lim_{\mu\to\infty} f(\mu) = -\sum_{i=1}^{M} \frac{\sigma_v^2\lambda_{ii}}{2} = -\frac{\sigma_v^2\text{Tr}\{\mathbf{\Lambda}\}}{2}.$$

Since the lower bound $\lim_{\mu\to\infty} f(\mu)$ is always negative ($\mathbf{R}$ was assumed to be a positive-definite matrix), the upper bound $\lim_{\mu\to 0} f(\mu)$ must be greater than or equal to 0 to ensure a solution for $f(\mu) = 0$. In other words,

$$\frac{\theta\text{Tr}\{\mathbf{\Lambda}^2\bar{\mathbf{Q}}\}}{1-\theta} - \frac{\sigma_v^2\text{Tr}\{\mathbf{\Lambda}\}}{2} \geq 0. \tag{52}$$

By rewriting $\text{Tr}\{\mathbf{\Lambda}^2\bar{\mathbf{Q}}\}$ as $\text{Tr}\{\mathbf{R}^2\mathbf{Q}\}$ and $\text{Tr}\{\mathbf{\Lambda}\}$ as $\text{Tr}\{\mathbf{R}\}$, after solving inequality (52) for θ, we obtain an expression for the minimum θ of LMS — $\theta_{\text{min}}^{\text{LMS}}$ — that guarantees an intersection point $\mu = \mu^{\text{o}} \geq 0$ between $f(\mu)$ and the μ-axis. This expression is given by

$$\theta_{\text{min}}^{\text{LMS}} = \frac{\gamma}{\gamma+2}, \tag{53}$$

where $\gamma = \sigma_v^2\text{Tr}\{\mathbf{R}\}/\text{Tr}\{\mathbf{R}^2\mathbf{Q}\}$. For $\theta < \theta_{\text{min}}^{\text{LMS}}$ it can be shown that $\mu^{\text{o}} = 0$, otherwise $\mu^{\text{o}} > 0$.

This result can be interpreted as follows: since the steady-state power of $\mathbf{w}_i^{\text{o}}$ is independent of θ, and (18) is a low-pass filter for $0 < \theta < 1$ with a bandwidth that increases as θ decreases, we conclude that for $\theta \leq \theta_{\text{min}}^{\text{LMS}}$, the filter no longer can track the variations of the weight vector. The optimum solution $\mu^{\text{o}} = 0$ in this case results from the assumption that $\mathbb{E}\{\mathbf{w}_i^{\text{o}}\} = \mathbb{E}\{\mathbf{w}_{-1}\} = \mathbf{0}$. Section V-A discusses the steady-state EMSE of the filter when $\mathbb{E}\{\mathbf{w}_i^{\text{o}}\} \neq \mathbf{0}$ and $\theta \leq \theta_{\text{min}}^{\text{LMS}}$.

The nonstationarity degree (NSD) of [18] also provides a bound for the tracking capabilities of an adaptive filter. Under our notation, a filter is unable to track a time-varying vector if $\text{NSD} = (1-\theta^2)\,\text{Tr}\{\mathbf{RQ}\}/\sigma_v^2 > 1$. We provide a comparison between our result and the NSD in Section V-B.

B. Theoretical steady-state EMSE for RLS

Similar to the LMS filter, to compute the steady-state EMSE and MSD for an individual RLS filter, we use $\ell = m = 2$, $\bar{\mathbf{M}}^{(2)} = \bar{\mathbf{P}} = (1-\lambda)\mathbf{R}^{-1}$, $\rho^{(2)} = 1$ and the same orthogonal transformation $\mathbf{U}$ in (32), obtaining

$$\begin{aligned}\bar{\mathbf{s}}_\infty^{(22)} \cong & \left[2\lambda\mathbf{I} - (1-\lambda)\boldsymbol{\ell}^{\cdot -1}\boldsymbol{\ell}^T\right]^{-1} \\ & \left\{(1-\lambda)\sigma_v^2\boldsymbol{\ell}^{\cdot -1} + \frac{2(1-\theta)}{1-\theta\lambda}\text{diag}\{\bar{\mathbf{Q}}\}\right\},\end{aligned} \tag{54}$$

where $\boldsymbol{\ell}^{\cdot -1}$ denotes element-wise inversion. The RLS steady-state EMSE and MSD are then

$$\zeta^{\text{RLS}} = \boldsymbol{\ell}^T \bar{\mathbf{s}}_\infty^{(22)}, \qquad \varepsilon^{\text{RLS}} = \mathbb{1}^T \bar{\mathbf{s}}_\infty^{(22)}. \tag{55}$$

Similarly to the qualitative analysis of the LMS filter developed in Sec. IV-A, for λ close to 1, the term $(1-\lambda)\boldsymbol{\ell}^{\cdot -1}\boldsymbol{\ell}^T$ can be disregarded with respect to $2\lambda\mathbf{I}$ in (54) and so the EMSE simplifies to

$$\zeta^{\text{RLS}} \cong \frac{(1-\lambda)\sigma_v^2 M}{2} + \frac{(1-\theta)\text{Tr}\{\mathbf{RQ}\}}{1-\theta\lambda}. \tag{56}$$

The λ^{o} that minimizes (56) can be obtained by setting the first derivative of ζ^{RLS} equal to zero. When $\theta = 1$, ζ^{RLS} reduces to a linear function of $(1-\lambda)$ given by

$$\zeta^{\text{RLS}} \cong \left.\frac{(1-\lambda)\sigma_v^2 M}{2}\right|_{\theta=1} \tag{57}$$

and so, the optimum forgetting factor that minimizes (57) is $\lambda^{\text{o}} = 1$ with the corresponding minimum EMSE $\zeta^{\text{o}}_{\text{RLS}} = 0$ (see the footnote in the previous section).

When $0 \le \theta < 1$, the first derivative of (56) is equal to

$$\frac{\partial \zeta^{\text{RLS}}}{\partial \lambda} \cong \frac{-\sigma_v^2 M}{2} + \frac{\theta(1-\theta)\text{Tr}\{\mathbf{RQ}\}}{(1-\theta\lambda)^2}. \tag{58}$$

Setting (58) equal to 0 and solving for λ leads to the following expression for λ^{o}:

$$\lambda^{\text{o}} \cong \min\left\{\frac{1-\sqrt{\Gamma}}{\theta}, 1\right\}, \tag{59}$$

where $\Gamma = 2\theta(1-\theta)\text{Tr}\{\mathbf{RQ}\}/(\sigma_v^2 M)$.

After simplification, the minimum θ for the RLS filter — $\theta^{\text{RLS}}_{\text{min}}$ — that guarantees $\lambda^{\text{o}} < 1$ in (59) is given by

$$\theta^{\text{RLS}}_{\text{min}} = \frac{1}{2\alpha^2 + 1}, \tag{60}$$

where $\alpha = \sqrt{\text{Tr}\{\mathbf{RQ}\}/(\sigma_v^2 M)}$. The comments at the end of the previous section also apply here.

a) Remark:: More precise values for θ_{min} than (53) and (60) can be obtained numerically directly by minimizing the full expressions for ζ^{LMS} (46) and ζ^{RLS} (55). We discuss this further in Section V.

C. Theoretical steady-state EMSE for combination between LMS and RLS

To compute the steady-state EMSE of the convex combination between LMS and RLS filters, we assume the LMS as $\ell = 1$ and the RLS as $m = 2$, with the corresponding parameters $\rho^{(1)} = \mu$, $\bar{\mathbf{M}}^{(1)} = \mathbf{I}$, $\rho^{(2)} = 1$ and $\bar{\mathbf{M}}^{(2)} = \bar{\mathbf{P}} = (1-\lambda)\mathbf{R}^{-1}$. Following the same steps as for the LMS and RLS filters, we obtain the diagonal $\bar{\mathbf{s}}_\infty^{(12)}$ of the steady-state transformed cross-covariance matrix $\mathbf{U}^T\mathbf{S}_\infty^{(12)}\mathbf{U}$ as

$$\begin{aligned}\bar{\mathbf{s}}_\infty^{(12)} = &\left\{(1-\lambda)\mathbf{I} + \mu\boldsymbol{\Lambda} - \mu(1-\lambda)\left(\mathbb{1}\boldsymbol{\ell}^T + 2\boldsymbol{\Lambda}\right)\right\}^{-1} \\ &\left\{\mu(1-\lambda)\sigma_v^2\mathbb{1} - \frac{(1-\theta)^2}{\theta}[(1-\theta)\mathbf{I} + \mu\theta\boldsymbol{\Lambda}]^{-1}\text{diag}\{\bar{\mathbf{Q}}\}\right. \\ &\left. + \frac{(1-\theta)(1-2\theta\lambda+\theta)}{\theta(1-\theta\lambda)}\text{diag}\{\bar{\mathbf{Q}}\}\right\},\end{aligned} \tag{61}$$

and again

$$\zeta^{(12)} = \boldsymbol{\ell}^T \bar{\mathbf{s}}_\infty^{(12)}, \qquad \varepsilon^{(12)} = \mathbb{1}^T \bar{\mathbf{s}}_\infty^{(12)}. \tag{62}$$

For sufficiently small μ and λ close to 1, the term $\mu(1-\lambda)\left(\mathbb{1}\boldsymbol{\ell}^T + 2\boldsymbol{\Lambda}\right)$ can be neglected with respect to the three first terms on the right-hand side of (61), and the following approximation is obtained for the cross-EMSE:

$$\begin{aligned}\zeta^{(12)} = \text{Tr}\{\mathbf{RS}_\infty^{(12)}\} &\cong \mu\sigma_v^2(1-\lambda)\text{Tr}\{\boldsymbol{\Gamma}\} \\ &- \frac{(1-\theta)^2\text{Tr}\left\{\boldsymbol{\Gamma}[(1-\theta)\mathbf{I} + \mu\theta\mathbf{R}]^{-1}\mathbf{Q}\right\}}{\theta} \\ &+ \frac{(1-\theta)(1-2\theta\lambda+\theta)\text{Tr}\{\boldsymbol{\Gamma}\mathbf{Q}\}}{\theta(1-\theta\lambda)},\end{aligned} \tag{63}$$

where $\boldsymbol{\Gamma} = \mathbf{R}[\mu\mathbf{R} + (1-\lambda)\mathbf{I}]^{-1}$.

D. Theoretical steady-state EMSE for Kalman filter

The straightforward approach to develop a theoretical model for the steady-state EMSE of the Kalman filter would be to use the nonstationary data model described in (17). However, this results in a recursion for the covariance matrix that is costly to solve. Instead, we define a modified data model (see (64) below) that can be put in a DARE (Discrete-Time Algebraic Riccati Equation) format and therefore can be solved efficiently using iterative procedures (see [33], [34]).

Given this, assume that instead of the linear model for $d(i)$ presented in (17), we have access to the following nonstationary data model

$$\mathbf{u}_i d(i) = \mathbf{u}_i\mathbf{u}_i^T\mathbf{w}^{\text{o}}_{i-1} + \mathbf{u}_i v(i), \tag{64}$$

where the optimum solution $\mathbf{w}_i^{\text{o}}$ is still given by (18).

Comparing the state-space model (1)–(2) with (18)–(64), the Kalman model corresponding to the nonstationary data model described in (18) and (64) is presented in Table I.

Table I
RELATION BETWEEN THE KALMAN FILTER AND THE ADAPTIVE FILTER VARIABLES.

KF	-	AF	KF	-	AF
$\mathbf{x}_i$	$\leftrightarrow$	$\mathbf{w}_i^{\text{o}}$	$\hat{\mathbf{x}}_i$	$\leftrightarrow$	$\mathbf{w}_i$
$\mathbf{z}_i$	$\leftrightarrow$	$\mathbf{u}_i d(i)$	$\mathbf{H}_i$	$\leftrightarrow$	$\mathbf{u}_i\mathbf{u}_i^T$
$\mathbf{F}_i$	$\leftrightarrow$	$\theta\mathbf{I}$	$\mathbf{G}_i$	$\leftrightarrow$	$\sqrt{1-\theta^2}\mathbf{I}$
$\mathbf{v}_i$	$\leftrightarrow$	$\mathbf{u}_i v(i)$	$\mathbf{t}_i$	$\leftrightarrow$	$\mathbf{q}_i$
$\mathbf{V}_i$	$\leftrightarrow$	$\sigma_v^2\mathbf{R}$	$\mathbf{T}_i$	$\leftrightarrow$	$\mathbf{Q}$

Since the Kalman model assumes $\mathbf{H}_i$ to be deterministic and $\mathbf{u}_i$ is random, using the equivalences from Table I in (2d) and defining $\tilde{\mathbf{w}}_i = \mathbf{w}_i^o - \mathbf{w}_i$, we can use (2d) to write

$$\begin{aligned}\mathbb{E}\{\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^T|\mathbf{u}_0,\ldots,\mathbf{u}_i\} = \boldsymbol{\mathcal{P}}_i = \theta^2\boldsymbol{\mathcal{P}}_{i-1} + (1-\theta^2)\mathbf{Q} \\ - \theta^2\boldsymbol{\mathcal{P}}_{i-1}\mathbf{u}_i\mathbf{u}_i^T[\sigma_v^2\mathbf{R} + \mathbf{u}_i\mathbf{u}_i^T\boldsymbol{\mathcal{P}}_{i-1}\mathbf{u}_i\mathbf{u}_i^T]^{-1}\mathbf{u}_i\mathbf{u}_i^T\boldsymbol{\mathcal{P}}_{i-1},\end{aligned} \tag{65}$$

where the cross-covariance matrix $\boldsymbol{\mathcal{S}}_i = \mathbf{0}$ since $v(i)$ and $\mathbf{q}_i$ are independent and zero mean. Now take the expectation with respect to $\mathbf{u}_0,\ldots,\mathbf{u}_i$, assuming that $\mathbb{E}\{\mathbf{u}_i\mathbf{u}_i^T[\mathbf{u}_i\mathbf{u}_i^T\sigma_v^2 + \mathbf{u}_i\mathbf{u}_i^T\bar{\boldsymbol{\mathcal{P}}}_{i-1}\mathbf{u}_i\mathbf{u}_i^T]^{-1}\mathbf{u}_i\mathbf{u}_i^T\} \approx \mathbf{R}[\sigma_v^2\mathbf{R} + \mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_{i-1}\mathbf{R}]^{-1}\mathbf{R}$ (an approximation that tends to be better for long filters [13]), obtaining the Riccati recursion

$$\begin{aligned}\bar{\boldsymbol{\mathcal{P}}}_i = \mathbb{E}\{\boldsymbol{\mathcal{P}}_i\} &\cong \theta^2\bar{\boldsymbol{\mathcal{P}}}_{i-1} + (1-\theta^2)\mathbf{Q} \\ &- \theta^2\bar{\boldsymbol{\mathcal{P}}}_{i-1}\mathbf{R}[\sigma_v^2\mathbf{R}+\mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_{i-1}\mathbf{R}]^{-1}\mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_{i-1}.\end{aligned} \tag{66}$$

The Kalman EMSE and MSD are obtained using

$$\zeta^{\text{KAL}} = \text{Tr}\{\mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_\infty\}, \qquad \varepsilon^{\text{KAL}} = \text{Tr}\{\bar{\boldsymbol{\mathcal{P}}}_\infty\}, \tag{67}$$

where $\bar{\boldsymbol{\mathcal{P}}}_\infty = \lim_{i\to\infty}\bar{\boldsymbol{\mathcal{P}}}_{i-1}$ is the solution of the discrete-time algebraic Riccati equation obtained by substituting $\bar{\boldsymbol{\mathcal{P}}}_i$ and $\bar{\boldsymbol{\mathcal{P}}}_{i-1}$ by $\bar{\boldsymbol{\mathcal{P}}}_\infty$ in (66).

The analytical expressions of the steady-state ζ-EMSE for LMS, RLS, their convex combination and for the Kalman filter are summarized in Table II for the general case $\mu > 0$ and $0 \ll \lambda < 1$ and in Table III for sufficiently small μ and λ close to 1. The results for the MSD are summarized in Table IV.

Table II

ANALYTICAL EXPRESSIONS FOR THE STEADY-STATE ζ-EMSE CONSIDERING LMS, RLS, THEIR CONVEX COMBINATION AND FOR THE KALMAN FILTER.

ζ-EMSE
$\zeta^{\text{LMS}} = \boldsymbol{\ell}^T\left[2\mu\boldsymbol{\Lambda} - \mu^2\boldsymbol{\ell}\boldsymbol{\ell}^T - 2\mu^2\boldsymbol{\Lambda}^2\right]^{-1}\left\{\mu^2\sigma_v^2\boldsymbol{\ell} + \frac{2(1-\theta)}{\theta}\left\{\mathbf{I} - \left[\mathbf{I} + \frac{\mu\theta}{1-\theta}\boldsymbol{\Lambda}\right]^{-1}\right\}\text{diag}\{\bar{\mathbf{Q}}\}\right\}$, with $\boldsymbol{\Lambda} = \mathbf{U}^T\mathbf{R}\mathbf{U}$, $\bar{\mathbf{Q}} = \mathbf{U}^T\mathbf{Q}\mathbf{U}$ and $\boldsymbol{\ell} = \text{diag}\{\boldsymbol{\Lambda}\}$.
$\zeta^{\text{RLS}} = \frac{(1-\lambda)\sigma_v^2 M}{(M+2)\lambda - M} + \frac{2(1-\theta)\text{Tr}\{\mathbf{R}\mathbf{Q}\}}{[(M+2)\lambda - M](1-\theta\lambda)}$
$\zeta^{\text{COMB}} = \frac{\zeta^{\text{LMS}}\zeta^{\text{RLS}} - (\zeta^{(12)})^2}{\zeta^{\text{LMS}} - 2\zeta^{(12)} + \zeta^{\text{RLS}}}$, where $\zeta^{(12)} = \boldsymbol{\ell}^T\left\{(1-\lambda)\mathbf{I} + \mu\boldsymbol{\Lambda} - \mu(1-\lambda)\left(\mathbb{1}\boldsymbol{\ell}^T + 2\boldsymbol{\Lambda}\right)\right\}^{-1}\left\{ + \mu(1-\lambda)\sigma_v^2\mathbb{1} - \frac{(1-\theta)^2}{\theta}[(1-\theta)\mathbf{I} + \mu\theta\boldsymbol{\Lambda}]^{-1}\text{diag}\{\bar{\mathbf{Q}}\} + \frac{(1-\theta)(1-2\theta\lambda+\theta)}{\theta(1-\theta\lambda)}\text{diag}\{\bar{\mathbf{Q}}\}\right\}$, with $\boldsymbol{\Lambda} = \mathbf{U}^T\mathbf{R}\mathbf{U}$, $\bar{\mathbf{Q}} = \mathbf{U}^T\mathbf{Q}\mathbf{U}$ and $\boldsymbol{\ell} = \text{diag}\{\boldsymbol{\Lambda}\}$.
$\zeta^{\text{KAL}} = \text{Tr}\{\mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_\infty\}$, where $\bar{\boldsymbol{\mathcal{P}}}_\infty$ is the solution of $\bar{\boldsymbol{\mathcal{P}}}_\infty \cong \theta^2\bar{\boldsymbol{\mathcal{P}}}_\infty + (1-\theta^2)\mathbf{Q} - \theta^2\bar{\boldsymbol{\mathcal{P}}}_\infty\mathbf{R}[\sigma_v^2\mathbf{R} + \mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_\infty\mathbf{R}]^{-1}\mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_\infty$.

Table III

ANALYTICAL EXPRESSIONS FOR THE STEADY-STATE ζ-EMSE CONSIDERING LMS, RLS AND THEIR CONVEX COMBINATION FOR SUFFICIENTLY SMALL μ AND λ CLOSE TO 1.

ζ-EMSE
$\zeta^{\text{LMS}} \cong \frac{\mu\sigma_v^2\text{Tr}\{\mathbf{R}\}}{2} + (1-\theta)\text{Tr}\left\{[(1-\theta)\mathbf{I} + \mu\theta\mathbf{R}]^{-1}\mathbf{R}\mathbf{Q}\right\}$
$\zeta^{\text{RLS}} \cong \frac{(1-\lambda)\sigma_v^2 M}{2} + \frac{(1-\theta)\text{Tr}\{\mathbf{R}\mathbf{Q}\}}{1-\theta\lambda}$
$\zeta^{\text{COMB}} = \frac{\zeta^{\text{LMS}}\zeta^{\text{RLS}} - (\zeta^{(12)})^2}{\zeta^{\text{LMS}} - 2\zeta^{(12)} + \zeta^{\text{RLS}}}$, where $\zeta^{(12)} \cong \mu\sigma_v^2(1-\lambda)\text{Tr}\{\boldsymbol{\Gamma}\} - \frac{(1-\theta)^2\text{Tr}\left\{\boldsymbol{\Gamma}[(1-\theta)\mathbf{I} + \mu\theta\mathbf{R}]^{-1}\mathbf{Q}\right\}}{\theta} + \frac{(1-\theta)(1-2\theta\lambda+\theta)\text{Tr}\{\boldsymbol{\Gamma}\mathbf{Q}\}}{\theta(1-\theta\lambda)}$, with $\boldsymbol{\Gamma} = \mathbf{R}[\mu\mathbf{R} + (1-\lambda)\mathbf{I}]^{-1}$.

Table IV

ANALYTICAL EXPRESSIONS FOR THE STEADY-STATE ε-MSD CONSIDERING LMS, RLS, THEIR CONVEX COMBINATION AND FOR THE KALMAN FILTER.

ε-MSD
$\varepsilon^{\text{LMS}} = \mathbb{1}^T\left[2\mu\boldsymbol{\Lambda} - \mu^2\boldsymbol{\ell}\boldsymbol{\ell}^T - 2\mu^2\boldsymbol{\Lambda}^2\right]^{-1}\left\{\mu^2\sigma_v^2\boldsymbol{\ell} + \frac{2(1-\theta)}{\theta}\left\{\mathbf{I} - \left[\mathbf{I} + \frac{\mu\theta}{1-\theta}\boldsymbol{\Lambda}\right]^{-1}\right\}\text{diag}\{\bar{\mathbf{Q}}\}\right\}$, with $\boldsymbol{\Lambda} = \mathbf{U}^T\mathbf{R}\mathbf{U}$, $\bar{\mathbf{Q}} = \mathbf{U}^T\mathbf{Q}\mathbf{U}$ and $\boldsymbol{\ell} = \text{diag}\{\boldsymbol{\Lambda}\}$.
$\varepsilon^{\text{RLS}} = \mathbb{1}^T\left[2\lambda\mathbf{I} - (1-\lambda)\bar{\boldsymbol{\ell}}\boldsymbol{\ell}^T\right]^{-1}\left\{(1-\lambda)\sigma_v^2\bar{\boldsymbol{\ell}} + \frac{2(1-\theta)}{1-\theta\lambda}\text{diag}\{\bar{\mathbf{Q}}\}\right\}$, with $\bar{\mathbf{Q}} = \mathbf{U}^T\mathbf{Q}\mathbf{U}$, $\boldsymbol{\ell} = \text{diag}\{\boldsymbol{\Lambda}\}$ and $\bar{\boldsymbol{\ell}} = \text{diag}\{\boldsymbol{\Lambda}^{-1}\}$.
$\varepsilon^{\text{COMB}} = \eta^{o2}\epsilon^{(1)} + (1-\eta^o)^2\epsilon^{(2)} + 2\eta^o(1-\eta^o)\epsilon^{(12)}$, where $\eta^o = \frac{\zeta^{(2)} - \zeta^{(12)}}{\zeta^{(1)} - 2\zeta^{(12)} + \zeta^{(2)}}$, $\varepsilon^{(12)} = \mathbb{1}^T\left\{(1-\lambda)\mathbf{I} + \mu\boldsymbol{\Lambda} - \mu(1-\lambda)\left(\mathbb{1}\boldsymbol{\ell}^T + 2\boldsymbol{\Lambda}\right)\right\}^{-1}\left\{ + \mu(1-\lambda)\sigma_v^2\mathbb{1} - \frac{(1-\theta)^2}{\theta}[(1-\theta)\mathbf{I} + \mu\theta\boldsymbol{\Lambda}]^{-1}\text{diag}\{\bar{\mathbf{Q}}\} + \frac{(1-\theta)(1-2\theta\lambda+\theta)}{\theta(1-\theta\lambda)}\text{diag}\{\bar{\mathbf{Q}}\}\right\}$, with $\boldsymbol{\Lambda} = \mathbf{U}^T\mathbf{R}\mathbf{U}$, $\bar{\mathbf{Q}} = \mathbf{U}^T\mathbf{Q}\mathbf{U}$ and $\boldsymbol{\ell} = \text{diag}\{\boldsymbol{\Lambda}\}$.
$\varepsilon^{\text{KAL}} = \text{Tr}\{\bar{\boldsymbol{\mathcal{P}}}_\infty\}$, where $\bar{\boldsymbol{\mathcal{P}}}_\infty$ is the solution of $\bar{\boldsymbol{\mathcal{P}}}_\infty \cong \theta^2\bar{\boldsymbol{\mathcal{P}}}_\infty + (1-\theta^2)\mathbf{Q} - \theta^2\bar{\boldsymbol{\mathcal{P}}}_\infty\mathbf{R}[\sigma_v^2\mathbf{R} + \mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_\infty\mathbf{R}]^{-1}\mathbf{R}\bar{\boldsymbol{\mathcal{P}}}_\infty$.

V. SIMULATIONS

In order to verify the tracking behavior of the proposed model, we consider a system identification application to compare the performance of the well known LMS and RLS filters and their convex combination with the corresponding optimum solution obtained via Kalman filter.

The unknown plant $\mathbf{w}_i^o$, of length $M = 7$, was initialized with random values in the interval $[-1, 1]$. The solution is then changed at each iteration according to the random-walk model (18). Following [19], the covariance matrix for $\mathbf{q}_i$ is given by

$$\mathbf{Q} = \delta\left[\beta\frac{\mathbf{R}}{\text{Tr}\{\mathbf{R}\}} + (1-\beta)\frac{\mathbf{R}^{-1}}{\text{Tr}\{\mathbf{R}^{-1}\}}\right], \tag{68}$$

where constant δ has been selected to be $\delta = 5\cdot 10^{-2}$, so that $\text{Tr}\{\mathbf{Q}\} = \delta$, and $\beta \in [0, 1]$ is a control parameter that allows to trade off between a situation with $\mathbf{Q} \propto \mathbf{R}$ (for $\beta = 1$, this is the situation in which LMS outperforms RLS according to [32]) and $\mathbf{Q} \propto \mathbf{R}^{-1}$ ($\beta = 0$, the case in which RLS outperforms LMS).

The regressor $\mathbf{u}_i$ is obtained from a process $u(i)$ as $\mathbf{u}_i = [u(i)\, u(i-1) \ldots u(i-M+1)]^T$, where $u(i)$ is generated with a first-order autoregressive (AR) model, whose transfer function is $\sigma_u \sqrt{(1-b^2)}/(1-bz^{-1})$ with $b = 0.8$. This model is fed with an i.i.d. Gaussian random process with variance $\sigma_u^2 = \frac{1}{7}$, so that $\mathrm{Tr}\{\mathbf{R}\} = 1$. The output additive noise $v(i)$ is i.i.d. Gaussian with zero mean and variance $\sigma_v^2 = 10^{-2}$.

Regarding the adjustment for the combinations, we used convex combinations with fixed step-size $\mu_a = 0.25$ and the auxiliary variable $a(i)$ restricted to the interval $[-4, 4]$, while the optimum step-size and forgetting factor of the constituent filters were numerically obtained through the theoretical steady-state EMSE general equations (46) and (55) (that is, we find the optimum values without resorting to the approximations for $\mu \approx 0$ and $\lambda \approx 1$).

To begin with, let us start the tracking analysis considering the parameter $\beta = 0.05$ in (68). In this case, according to (53) and (60), the minimum θ that can be used for the LMS and RLS filters in order to have an optimum step-size $\mu^{\mathrm{o}} > 0$ and optimum forgetting factor $0 \ll \lambda^{\mathrm{o}} < 1$ is $\theta_{\min}^{\mathrm{LMS}} \approx 0.84$ and $\theta_{\min}^{\mathrm{RLS}} \approx 0.92$. The bounds obtained numerically from the general equations (46) and (55) are $\theta_{\min}^{\mathrm{LMS}} \approx 0.88$ and $\theta_{\min}^{\mathrm{RLS}} \approx 0.94$. Figure 2 shows the steady-state EMSEs estimated from the ensemble-average learning curve obtained from 600 independent runs for 30.000 iterations of the algorithms $\mu^{\mathrm{o}} - \mathrm{LMS}$, $\lambda^{\mathrm{o}} - \mathrm{RLS}$, their combination and the corresponding Kalman filter when $0 < \theta < 1$. Figure 3 compares the tracking performance between the simulated case and the theoretical steady-state EMSE equations provided in table II when θ lies in the range $[\min\{\theta_{\min}^{\mathrm{LMS}}, \theta_{\min}^{\mathrm{RLS}}\}, 1]$.

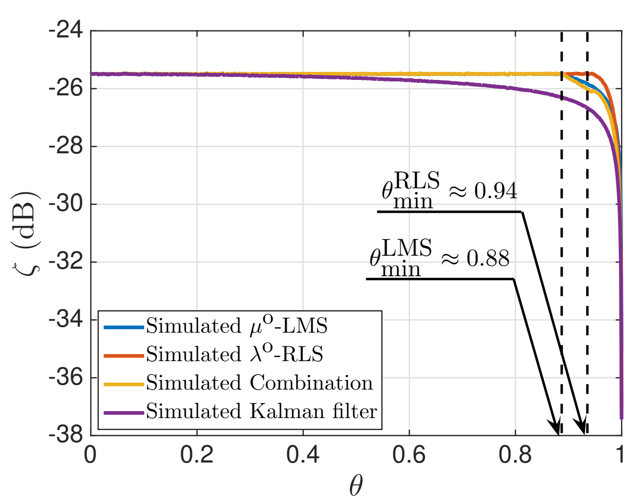


Figure 2. Simulated steady-state EMSEs curves for $\mu^{\mathrm{o}}-$LMS, $\lambda^{\mathrm{o}}-$RLS, their convex combination and the corresponding Kalman filter when $0 < \theta < 1$.

As can be seen in Figures 2 and 3, for $\theta \approx 1$, the steady-state EMSE achieved by combining both LMS and RLS filters with optimum settings is close to the optimum EMSE obtained

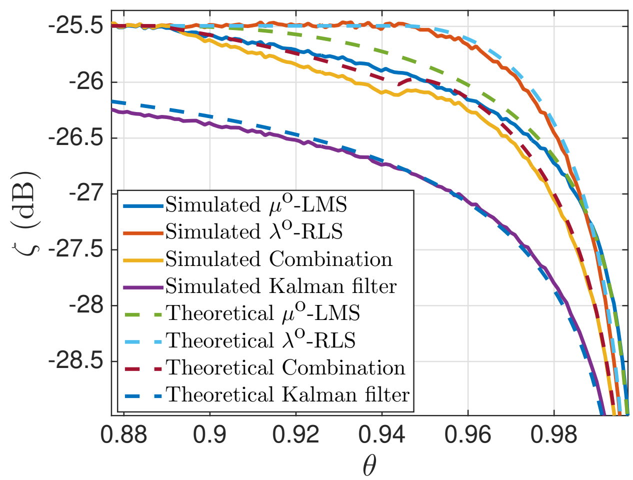


Figure 3. Tracking comparison between the theoretical steady-state EMSE equations from table II and the simulated case for $\min\{\theta_{\min}^{\mathrm{LMS}}, \theta_{\min}^{\mathrm{RLS}}\} < \theta < 1$.

via Kalman filter. As θ decreases, the difference between the combination and the KF curves starts to increase, reaching its maximum value for $\theta \approx \theta_{\min}^{\mathrm{LMS}}$ which is approximately 0.80 dB. For $\theta < \min\{\theta_{\min}^{\mathrm{LMS}}, \theta_{\min}^{\mathrm{RLS}}\}$, the adaptive filters LMS and RLS as well as their combination, are no longer tracking the variations of $\mathbf{w}_i^{\mathrm{o}}$ since $\mu^{\mathrm{o}} \approx 0$ and $\lambda^{\mathrm{o}} = 1$. For $\theta < 0.2$, the resulting performance for all filters is approximately the same. The values of μ^{o} and λ^{o} as θ changes from $\min\{\theta_{\min}^{\mathrm{LMS}}, \theta_{\min}^{\mathrm{RLS}}\}$ to 1 are shown in figure 4.

To see the effect of assumptions of small step-sizes and λ close to one, in figure 4 we also plot the approximated curves obtained through expressions (51) and (59). As can be seen in this figure, for $\theta < 1$, the error between the optimum parameters estimated numerically from the general equations provided in table II and by using equations (51) and (59) is at most $\Delta_{\mu^{\mathrm{o}}} \approx 0.05$ for μ^{o} and $\Delta_{\lambda^{\mathrm{o}}} \approx 0.01$ for λ^{o}.

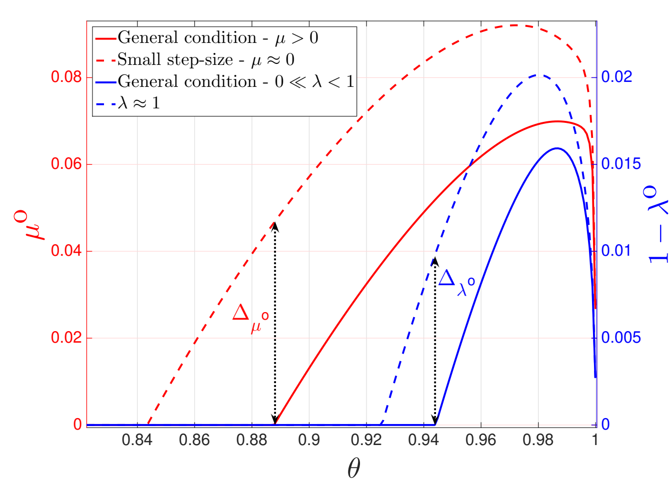


Figure 4. Variation of the optimum parameters for $\min\{\theta_{\min}^{\mathrm{LMS}}, \theta_{\min}^{\mathrm{RLS}}\} < \theta < 1$ considering the general and the approximated equations provided in tables II and III.

By computing the ratio $\theta_{\min}^{\mathrm{LMS}}/\theta_{\min}^{\mathrm{RLS}}$, we are able to compare the θ range allowed for each filter and see if they have the same range ($\theta_{\text{min}}^{\text{LMS}} = \theta_{\text{min}}^{\text{RLS}}$) or if one filter is more restrictive than the other ($\theta_{\text{min}}^{\text{LMS}} > \theta_{\text{min}}^{\text{RLS}}$ or $\theta_{\text{min}}^{\text{LMS}} < \theta_{\text{min}}^{\text{RLS}}$). The ratio between (53) and (60) is given by

$$\frac{\theta_{\text{min}}^{\text{LMS}}}{\theta_{\text{min}}^{\text{RLS}}} = \frac{\text{Tr}\{\mathbf{R}\}\{2\text{Tr}\{\mathbf{RQ}\} + \sigma_v^2 M\}}{M\{\sigma_v^2 \text{Tr}\{\mathbf{R}\} + 2\text{Tr}\{\mathbf{R}^2\mathbf{Q}\}\}}. \tag{69}$$

With this expression, we can see that the value of θ_{min} for LMS can be smaller or larger than that for RLS, depending on the values of $\mathbf{R}$ and $\mathbf{Q}$. As a simple example, consider $\sigma_v^2 = 0.01$, $\mathbf{R} = \text{diag}(1, 2)$. If $\mathbf{Q} = \text{diag}(0.01, 0.001)$, then $\frac{\theta_{\text{min}}^{\text{LMS}}}{\theta_{\text{min}}^{\text{RLS}}} = 1.14$, while, for $\mathbf{Q} = \text{diag}(0.001, 0.01)$ we have $\frac{\theta_{\text{min}}^{\text{LMS}}}{\theta_{\text{min}}^{\text{RLS}}} = 0.830$.

By using the theoretical steady-state EMSE equations summarized in table II, Figure 5 compares the tracking behavior between the LMS, RLS, their combination and the Kalman filter when $\beta \in [0, 1]$ and $\theta = 0.99$. For this simulation we kept the same matrices $\mathbf{R}$ and $\mathbf{Q}$ as well as the parameters σ_v^2 and M used to obtain figure 2.

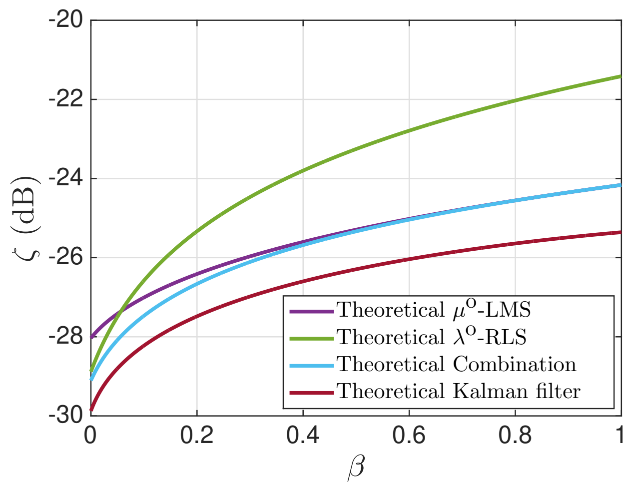


Figure 5. EMSE of μ^{o}-LMS, λ^{o}-RLS, their convex combination and the Kalman filter when $\mathbf{Q}$ smoothly changes between $\mathbf{R}$ and $\mathbf{R}^{-1}$.

As shown in Figure 5, depending on the value of β, the optimal EMSE for RLS filters can be larger or smaller than the optimal EMSE for LMS filters, but always lower bounded by the optimal EMSE provided by the Kalman filter. For the general equations in which $\mu > 0$ and $0 \ll \lambda < 1$, the optimal steady-state EMSE curve for the combination is approximately 1dB larger than the optimal KF. However, compared to the $\mathcal{O}(M^2)$ complexity of the KF, if lattice [15] or DCD [16] implementations are used for RLS, the computational cost of the combination is reduced from $\mathcal{O}(M^2)$ to $\mathcal{O}(M)$. Figure 6 plots the MSD as a function of β, under the same conditions. Note that we plotted two curves for the combination — one (green) using the value of the mixing parameter that is optimal for the MSD, and the second (yellow) using the optimum mixing parameter for the EMSE, which the filter can estimate in practice. It can be seen that both are close, but the practical curve is slightly worse than the best filter at a few points.

In order to see the influence of the filter length M in the optimal steady-state EMSE of each filter, Figure 7 shows

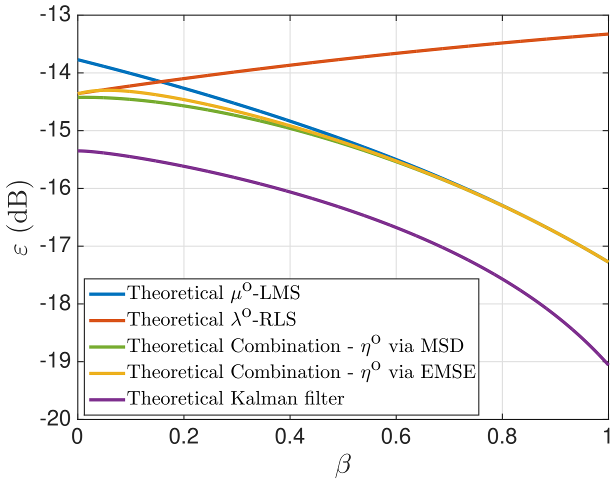


Figure 6. MSD of μ^{o}-LMS, λ^{o}-RLS, their convex combination and the Kalman filter when $\mathbf{Q}$ smoothly changes between $\mathbf{R}$ and $\mathbf{R}^{-1}$.

the tracking behavior for the theoretical and simulated μ^{o}-LMS, λ^{o}-RLS, their convex combination and for the KF when M changes from 1 to 200. For this simulation we used the general equations provided in Table II and compared with the simulated case using the following parameters $\sigma_v^2 = 10^{-2}$, $\text{Tr}\{\mathbf{Q}\} = 5 \cdot 10^{-2}$, $\beta = 0.05$, $\sigma_u^2 = \frac{1}{M}$ to keep $\text{Tr}\{\mathbf{R}\} = 1$ and $\theta = 0.9999$ to cover all possible values of $\theta_{\text{min}}^{\text{LMS}}$ and $\theta_{\text{min}}^{\text{RLS}}$. In addition to the simulated case, we used convex combinations with fixed step-size $\mu_a = 0.25$, the auxiliary variable $a(i)$ restricted to the interval $[-4, 4]$ and the optimal steady-state EMSEs were estimated from the ensemble-average learning curve obtained from 50 independent runs and 10^6 iterations for each algorithm. Due to the long processing time, only a few points were plotted for the simulated curves.

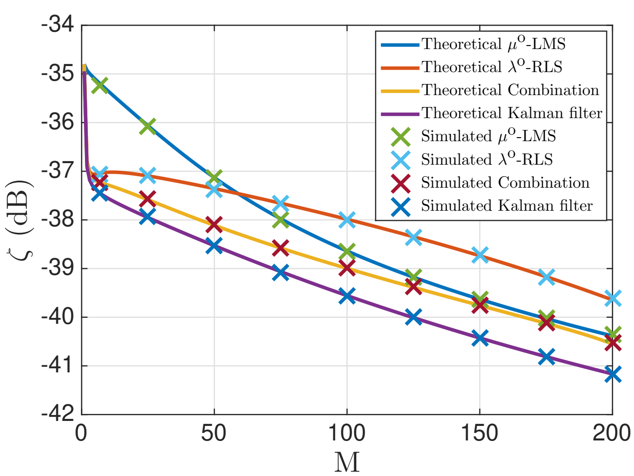


Figure 7. Theoretical and simulated steady-state EMSE for μ^{o}-LMS, λ^{o}-RLS, their convex combination and the Kalman filter for $1 \leq M \leq 200$.

As depicted in figure 7, for the defined range M, the optimal steady-state EMSE for the KF is at most 0.6 dB smaller than the combination when $M > 100$, despite the computational cost for the KF increasing at a rate of $\mathcal{O}(M^2)$ while the combination increases linearly with the filter length.

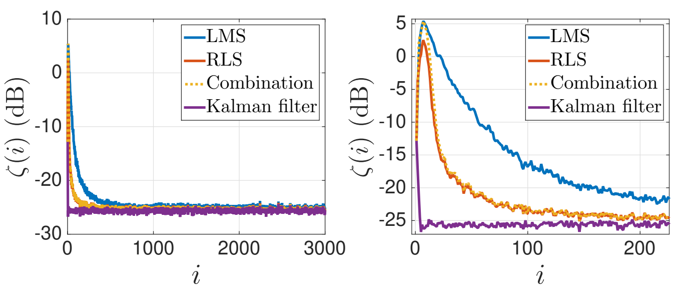
A. DC analysis of the steady-state EMSE

All simulations performed until now considered a zero mean plant $\mathbf{w}_i^{\mathrm{o}}$, whose values are updated at each iteration according to model (18), and the parameter θ was always kept greater than $\min\{\theta_{\text{min}}^{\text{LMS}}, \theta_{\text{min}}^{\text{RLS}}\}$ since we were dealing with tracking problems. Let us assume now that $\theta < \min\{\theta_{\text{min}}^{\text{LMS}}, \theta_{\text{min}}^{\text{RLS}}\}$ and the model (18) has a DC component added to the optimum vector $\mathbf{w}_i^{\mathrm{o}}$, i.e

$$\mathbf{w}_i^{\mathrm{o}} = \theta \mathbf{w}_{i-1}^{\mathrm{o}} + \sqrt{1-\theta^2}\mathbf{q}_i + \bar{\mathbf{w}}^{\mathrm{o}}, \tag{70}$$

where $\bar{\mathbf{w}}^{\mathrm{o}} = \mathbb{E}\{\mathbf{w}_i^{\mathrm{o}}\}$ is an $M \times 1$ vector.

The simulation in Figure 8 compares the performance of the Kalman filter, RLS with $\lambda = 1$ and LMS with decreasing μ (see below) considering a nonzero DC component. Figure 8(b) shows a zoom of the initial iterations of the EMSEs learning curves. Note that the lag between the responses is due to the fact that the Kalman filter knows the exact value of the DC component, while the adaptive filters need to learn it.



(a) (b)

Figure 8. Simulated EMSEs learning-curves for $\theta = 0.5$ and $\bar{\mathbf{w}}^{\mathrm{o}} = \mathbf{1}$, focusing on: a) steady-state regime and b) transient behavior. To have a better visualization of the result, we used doted lines to represent the combination and the KF curves.

We can see that the adaptive filters, although not able to track the variations of $\mathbf{w}_i^{\mathrm{o}}$ (since $\mu^{\mathrm{o}} = 0$ and $\lambda^{\mathrm{o}} = 1$), are still able to converge to the DC component $\bar{\mathbf{w}}^{\mathrm{o}}$. For this simulation we plot the EMSEs estimated from the ensemble-average learning curve of 600 independent runs for 3000 iterations of the algorithms LMS, RLS, their convex combination and the KF considering $\theta = 0.5$ and the same matrices $\mathbf{R}$ and $\mathbf{Q}$ as well as the parameters σ_v^2, β and M used to obtain figure 2.

The following assumptions had to be made in order to ensure convergence to the optimum solution for both LMS algorithm and for the KF:

- LMS: the required $\mu^{\mathrm{o}} = 0$ when $\theta < \theta_{\text{min}}^{\text{LMS}}$ can not be used, since the filter will not be able to adapt and converge to the mean value of $\mathbf{w}_i^{\mathrm{o}}$. For this reason, we assumed variable step-size $\mu(i) = 1/i$ to ensure equivalence to the RLS case with $\lambda = 1$ [35]. To speed up the convergence rate, we used $\mu(i) = 1/\sqrt{i}$ for $i < 1000$.
- KF: we added the *a priori* information of the DC value to the estimate $\hat{\mathbf{x}}_i$ since the KF equations assumes that $\mathbf{x}_i$ is zero mean.

B. Comparison between θ_{min} and the nonstationarity degree

The nonstationarity degree (NSD), defined in [18], is another measure of when an adaptive filter is able to track a time-varying weight vector. It compares the performance (measured by the steady-state EMSE) of an adaptive filter to estimate $\mathbf{u}_i^T \mathbf{w}_{i-1}^{\mathrm{o}}$ with the estimate provided by $d(i)$ itself. Let $\mathcal{F} = \{\mathbf{w}_{i-1} \in \mathbb{R}^M : \mathbf{w}_{i-1}$ depends only on $d(j), \mathbf{u}_j$ for $j < i\}$ be the set of possible a-priori estimates. The NSD is defined as

$$\text{NSD} = \frac{\min_{\mathbf{w}_{i-1}\in\mathcal{F}} \mathbb{E}\left\{[\mathbf{u}_i^T(\mathbf{w}_{i-1} - \mathbf{w}_{i-1}^{\mathrm{o}})]^2\right\}}{\sigma_v^2}. \tag{71}$$

If $\text{NSD} > 1$, [18] argues that an adaptive filter is not able to track $\mathbf{w}_i^{\mathrm{o}}$. Under our model for $\mathbf{w}_i^{\mathrm{o}}$ (18), the NSD is given by

$$\text{NSD} = (1-\theta^2)\frac{\text{Tr}\{\mathbf{RQ}\}}{\sigma_v^2}. \tag{72}$$

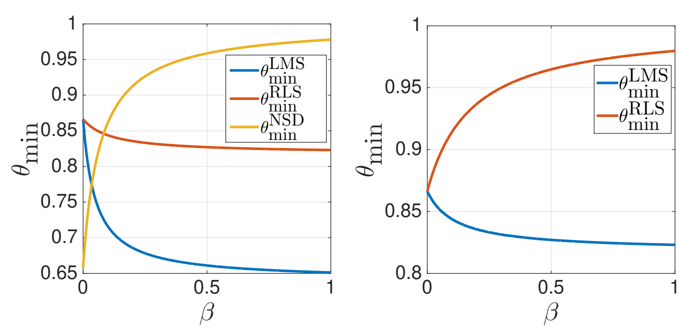
Imposing the condition $\text{NSD} = 1$, we obtain

$$\theta_{\text{min}}^{\text{NSD}} = \sqrt{\max\left\{0, 1 - \frac{\sigma_v^2}{\text{Tr}\{\mathbf{RQ}\}}\right\}}. \tag{73}$$

If $\theta < \theta_{\text{min}}^{\text{NSD}}$, all a-priori estimates would have an EMSE larger than σ_v^2, and therefore if the goal of the adaptive filter were to estimate $\mathbf{u}_i^T\mathbf{w}_{i-1}^{\mathrm{o}}$, no adaptive filter would do better than simply using $d(i)$ as an estimate (note that $\theta_{\text{min}}^{\text{NSD}}$ tends to 1 as $\sigma_v^2 \to 0$).

The minimum values for θ defined in this paper complement the NSD in two ways: (a) we can also define a condition for tracking based on the MSD, while the NSD considers only the EMSE; (b) our estimates for θ_{min} take into account the DC part of the weight vector, as described in Section V-A — that is, a situation in which an adaptive filter is useful to estimate the DC part of the weight vector, but not to track variations around that DC value.

Figure 9 compares the values of θ_{min} as derived in this paper, for both EMSE and MSD, with $\theta_{\text{min}}^{\text{NSD}}$, considering the same case as in Figure 5, but with $\sigma_v^2 = 10^{-3}$.



(a) (b)

Figure 9. Values of θ_{min} considering the EMSE and NSD (a), and the MSD (b).

It can be seen on the left panel that the values of θ_{min} obtained in this paper and through the NSD are quite different: the condition we use to define when a filter is no longer able to track is different than the condition used for the NSD.

VI. Conclusion

Combination approaches are an effective way to improve the performance of adaptive filters. In this paper we have studied the tracking performance of combinations between LMS and RLS filters and compared the resulting EMSE with the optimal case obtained via Kalman filter.

We have shown that, assuming a first order AR model with finite autocovariance matrix to describe the time variation of the unknown parameter vector, it is possible to achieve through convex combination between LMS and RLS filters a steady-state EMSE and MSD performance close to the optimal case obtained via Kalman Filter, as long as the pole of the AR model is greater than a minimum value. This remains true even if the unknown parameter vector has a DC component. The minimum value for the pole of the AR model provides a model for how fast a plant can vary so that an adaptive filter can still track it.

The advantage of our approach using combinations arises from the fact that the combination can be implemented with complexity $\mathcal{O}(M)$, while it takes at least $\mathcal{O}(M^2)$ operations to compute the corresponding Kalman filter for the model considered here.

References

[1] P. Zarchan, *Progress In Astronautics and Aeronautics: Fundamentals of Kalman Filtering: A Practical Approach*. Aiaa, 2005, vol. 208.

[2] M. Barczyk and A. F. Lynch, "Invariant observer design for a helicopter UAV aided inertial navigation system," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 3, pp. 791–806, 2012.

[3] C. H. Lim, T. S. Lim, and V. C. Koo, "Design and development of a real-time GPS-aided SINU system," *International Journal of Advanced Robotic Systems*, vol. 9, no. 5, p. 194, 2012.

[4] S. Salti, A. Lanza, and L. Di Stefano, "Synergistic change detection and tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 4, pp. 609–622, 2014.

[5] G. Bresson, T. Féraud, R. Aufrère, P. Checchin, and R. Chapuis, "Real-time monocular SLAM with low memory requirements," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1827–1839, 2015.

[6] A. H. Sayed, *Adaptive filters*. John Wiley & Sons, 2011.

[7] M. S. Grewal and A. P. Andrews, *Kalman filtering: theory and practice using MATLAB*. Wiley, 2011.

[8] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. NJ: Prentice Hall, 2000.

[9] J. Fabry, S. Liebich, P. Vary, and P. Jax, "Active noise control with reduced-complexity Kalman filter," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018, pp. 166–170.

[10] T. Dietzen, S. Doclo, A. Spriet, W. Tirry, M. Moonen, and T. van Waterschoot, "Low-complexity Kalman filter for multi-channel linear-prediction-based blind speech dereverberation," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2017, pp. 284–288.

[11] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2008.

[12] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 3rd ed. Springer, 2008.

[13] V. H. Nascimento and M. T. M. Silva, "Adaptive filters," in *Academic Press Library in Signal Processing:*, R. Chellappa and S. Theodoridis, Eds. Chennai: Academic Press, 2014, vol. 1, Signal Processing Theory and Machine Learning, pp. 619—761.

[14] R. Claser and V. H. Nascimento, "Low-complexity approximation to the Kalman filter using convex combinations of adaptive filters from different families," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 2630–2633.

[15] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

[16] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3150–3161, 2008.

[17] J. Fabry, S. Kühl, and P. Jax, "On the Steady State Performance of the Kalman Filter Applied to Acoustical Systems," *IEEE Signal Processing Letters*, vol. 27, pp. 1854–1858, 2020.

[18] O. Macchi and M. Turki, "The non stationarity degree: Can an adaptive filter be worse than no processing?" *IFAC Proceedings Volumes*, vol. 25, no. 14, pp. 517–522, 1992.

[19] M. T. Silva and V. 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.

[20] 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, no. 5, pp. 1853–1864, May 2008.

[21] L. F. O. Chamon, W. B. Lopes, and C. G. Lopes, "Combination of adaptive filters with coefficients feedback," in *Proc. Int. Simp. Acoustics, Speech and Sig. Proc. 2012*, 2012, pp. 3785–3788.

[22] S. S. Kozat, A. T. Erdogan, A. C. Singer, and A. H. Sayed, "Steady-state MSE performance analysis of mixture approaches to adaptive filtering," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4050–4063, Aug. 2010.

[23] 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.

[24] J. Arenas-García and A. R. Figueiras-Vidal, "Adaptive combination of proportionate filters for sparse echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1087–1098, 2009.

[25] T. Trump, "A combination of two NLMS filters in an adaptive line enhancer," in *Digital Signal Processing (DSP), 2011 17th International Conference on*. IEEE, 2011, pp. 1–6.

[26] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Combined adaptive beamforming schemes for nonstationary interfering noise reduction," *Signal Processing*, vol. 93, no. 12, pp. 3306–3318, 2013.

[27] N. V. George and A. Gonzalez, "Convex combination of nonlinear adaptive filters for active noise control," *Applied Acoustics*, vol. 76, pp. 157–161, 2014.

[28] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-Garcia, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 3301–3304.

[29] V. H. Nascimento, M. T. Silva, L. A. Azpicueta-Ruiz, and J. Arenas-García, "On the tracking performance of combinations of least mean squares and recursive least squares adaptive filters," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3710–3713.

[30] R. Candido, M. T. Silva, and V. H. Nascimento, "Affine combinations of adaptive filters," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*. IEEE, 2008, pp. 236–240.

[31] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[32] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 2937–2944, 1994.

[33] L.-Z. Lu and W.-W. Lin, "An iterative algorithm for the solution of the discrete-time algebraic Riccati equation," *Linear Algebra and Its Applications*, vol. 188, pp. 465–488, 1993.

[34] Y. Feng, B. D. Anderson, and W. Chen, "Solving discrete algebraic Riccati equations: A new recursive method," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*. IEEE, 2009, pp. 1720–1724.

[35] H. J. Kushner and G. G. Yin, "Applications to learning, state dependent noise, and queueing," in *Stochastic Approximation Algorithms and Applications*. Springer, 1997, pp. 25–46.