Low-complexity Approximation to the Kalman Filter Using Convex Combinations of Adaptive Filters from Different Families

Raffaello Claser and Vitor H. Nascimento
LPS-PSI-EPUSP
University of São Paulo
São Paulo, SP, Brazil
rclaser@usp.br, vitor@lps.usp.br

Abstract—It is known that combinations of the least mean square (LMS) and recursive least squares (RLS) algorithms may achieve a performance in tracking better than what is possible to obtain with either kind of filter individually. In this paper, we consider combinations of LMS and RLS filters and compare their performance under a nonstationary condition with the optimal solution obtained via Kalman filter. We show that combination schemes may have a tracking performance close to that of a Kalman filter, but with lower computational complexity (linear in the filter length instead of quadratic—in the case of the example shown here—or cubic, for general Kalman models).

I. INTRODUCTION

When choosing an adaptive algorithm for a given application, one of the important points to be considered is the algorithm’s ability to track variations in the parameter vector one wishes to estimate [1]. The Kalman filter (KF) has long been shown to be the optimal solution to many tracking and data prediction tasks [2], and 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 following a linear model [2].

In time-varying scenarios, combination schemes offer improved tracking capabilities with respect to the component filters [3]. Nevertheless, when a combination of two adaptive filters of the same family is used, for example two least mean-squares (LMS) with different step sizes, or two recursive least-squares (RLS) with different forgetting factors, the resulting performance will never be better than the performance of each filter using optimum settings for a certain nonstationary condition.

As it was shown in [4], [5], when combining filters from different families, namely LMS and RLS, it is possible to take advantage of the tracking properties from each filter and obtain a structure with better performance than if each filter were implemented individually. Combinations of Kalman Filters were also proposed using different update rules as proposed in [6] and [7].

Assuming we want to estimate a vector with \( M \) parameters, the computational complexity for a convex combination between one LMS and one RLS can be implemented with \( O(M) \) operations (if lattice or Dichotomous Coordinate Descent - DCD algorithms are used) and the Kalman Filter requires \( O(M^2) \) operations (for a first-order random walk state-space model, see (10) below), or \( O(M^3) \) for a general state-space model. This paper describes how close the combination scheme can get to the optimal excess mean square error (EMSE) obtained via Kalman Filter. We show that the performance gain obtained with the Kalman filter is not very large, less than 1dB, even when we do not have exact knowledge of the true covariance matrix of the noise process.

This paper is organized as follows: in Section II we review the LMS and RLS algorithms, as well as their respective combination, and also the Kalman filter equations. Section III presents the data model adopted in this paper. Section IV compares the performance of each algorithm under different conditions, and finally, section V concludes the paper.

II. PROBLEM FORMULATION

Let \( d_n \) be a zero-mean scalar-valued real random variable with variance \( \sigma^2_d \), and let \( u_n \) be a \( 1 \times M \) zero-mean real-valued random input regressor vector with positive-definite covariance matrix denoted by \( R_u = \mathbb{E}\{u_n^* u_n\} \). Then, the solution \( w_n \) of the linear least-mean-squares problem

\[
\min_w \mathbb{E} |d_n - u_n w|^2, \tag{1}
\]

can be approximated recursively as follows [8], [9]

\[
w_n = w_{n-1} + f[d_n, u_n, s_n]. \tag{2}
\]

Different adaptive schemes are characterized by their update functions \( f[\cdot] \) (\( s_n \) represents any other state information that is needed for the update rule). For the LMS and RLS cases, the following equations describe how to update \( w_n \) at each iteration \( n \), respectively [8], [9]:

\[
w_n = w_{n-1} + \mu u_n^T [d_n - u_n w_{n-1}], \tag{3}
\]

\[
w_n = w_{n-1} + \mu R_u^{-1} [d_n - u_n w_{n-1}], \tag{4}
\]

where the step size \( \mu \) and the forgetting factor \( \lambda \) are tuned such that the solution of the resulting algorithm converges to the same value as that of the Kalman filter when the regressor \( u_n \) is a zero-mean realization of an \( M \times 1 \) white Gaussian random vector.

This work was partially funded by Fapesp grant 2014/04256-2 and CNPq grant 306268/2014-0.
\[ w_n = w_{n-1} + P_n u_n^T [d_n - u_n w_{n-1}] \]

\[ P_n = \beta^{-1} P_{n-1} - \beta^{-2} \frac{P_{n-1} u_n^T u_n P_{n-1}}{1 + \beta^{-1} u_n^T P_{n-1} u_n}, \]

where \( \beta \) is a positive step-size (usually a small value), \( \mu \) is a forgetting factor, and \( P_n \) is an estimate of the inverse of the regressor autocovariance matrix \( R_u \), which can be computed in an efficient way using the matrix inversion lemma, RLS-DCD algorithm, or if lattice algorithms are used, its explicit evaluation may be avoided [3], [8], [10].

Comparing both algorithms, the RLS has faster initial convergence, but has larger computational complexity, and LMS is initially slower, but has smaller computational complexity.\( \star \) Despite the fact that the standard RLS has \( O(M^2) \), fast \( O(M) \) versions of RLS (such as lattice [8] or Dichotomous Coordinate Descent algorithm - DCD [12]) can be applied to reduce the complexity.

According to [4], a convex combination approach is an interesting way to improve adaptive filter performance. In this case, the individual filters are independently adapted using their own error signals, while the combination is adapted by means of a stochastic gradient algorithm.\( \star \) Fig. 1 illustrates a block diagram of this combination.

The output of the overall filter combination is denoted by

\[ y_n = \lambda_n y_{n1}^1 + [1 - \lambda_n] y_{n2}^2, \]

where \( y_{n1}^1 \) and \( y_{n2}^2 \) are the outputs of two transversal filters at time \( n \), i.e., \( y_{ni} = u_n w_{ni}, i = 1, 2 \). \( e_n^1 \) and \( e_n^2 \) are the output filters errors \( (e_{ni} = d_n - y_{ni}, i = 1, 2) \), and \( \lambda_n \) is a mixing scalar parameter that lies between zero and one.

The idea behind the combination is that the best properties of the individual filters \( w_{n1}^1 \) and \( w_{n2}^2 \) can be extracted if \( \lambda_n \) is assigned with the appropriate values at each iteration \( n \).

In all cases mentioned until now, no information about how the optimum solution \( w_n^0 \) evolves with time was used to compute the solution \( w_n \). But, if we have access to such information, a Kalman Filter (KF) can be implemented in order to solve the problem optimally.

The KF is the optimal linear least-mean-squares (l.l.m.s.) solution to the problem of sequentially estimating the states of a dynamical system in which the state evolution and measurement processes are both linear and Gaussian [13]. Thus considering a state-space description of the form:

\[ x_n = F_n x_{n-1} + G_n t_n \]

\[ z_n = H_n x_n + v_n \]

where \( x_n \) is the \( M \times 1 \) state vector, \( F_n \) is the \( M \times M \) state-transition matrix, \( G_n \) is the \( M \times N \) control-input model, \( t_n \) is a \( N \times 1 \) Gaussian random state noise vector with zero-mean and covariance matrix \( T_n \), \( z_n \) is the \( D \times 1 \) observation vector, \( H_n \) is the \( D \times M \) measurement matrix, and \( v_n \) is a \( D \times 1 \) Gaussian random measurement noise vector with zero-mean and covariance matrix \( R_n \).

In this case, given observations \( z_n \) that satisfy the state-space model described in (6) and (7), an approximate solution \( \hat{x}_n \) can be recursively computed by using the following set of KF equations [8]:

\[ R_{e,n} = R_n + H_n P_{n|n-1} H_n^T \]

\[ K_n = (F_n P_{n|n-1} H_n + G_n S_n)^{-1} \]

\[ e_n = z_n - H_n \hat{x}_{n|n-1} \]

\[ \hat{x}_{n+1|n} = F_n \hat{x}_{n|n-1} + K_n e_n \]

\[ P_{n+1|n} = F_n P_{n|n-1} F_n^T + G_n T_n G_n^T - K_n R_{e,n} K_n^T \]

\( K_n \) is the \( M \times D \) Kalman gain, \( S_n \) is the cross-covariance matrix between the noise processes \( (t_n, v_n) \) which is equal to \( E \{ t_n v_n^* \} \) and \( e_n \) is a \( D \times 1 \) error vector. Here, the notation \( n|n-1 \) is used in order to indicate that the estimation is based on the observations from \( z_0 \) through \( z_{n-1} \).

As we can see in equations (8b) and (8e), in general one needs \( O(M^3) \) operations to compute the Kalman gain and the covariance matrix \( P_{n+1|n} \). Depending on the application, this computational cost may be prohibitive. We propose here to use a combination of an LMS and an RLS instead, and show that, in the case of the model usually employed to study tracking of adaptive filters, the optimum solution obtained with the KF is only slightly better than the result obtained with the combination.
III. DATA MODEL

In the sequel we adopt the following assumptions.

- $d_n$ and $u_n$ are related according to the following linear regression model
  \[ d_n = u_n w_n^o + v_n, \]  
  \[ \text{Eqn. (9)} \]

where $v_n$ is i.i.d. noise, independent of $u_n$ and with variance $\sigma_v^2$.

- $E\{u_n\} = 0$ and $E\{d_n\} = 0$.

- $w_n^o$ changes according to a first-order random walk
  \[ w_n^o = w_{n-1}^o + q_n, \]  
  \[ \text{Eqn. (10)} \]

where $q_n$ is independent of $u_n$, with autocovariance $Q_n$.

Note that, due to its simplicity, this is the most usual model used in the literature to study tracking properties of adaptive filters [8]. By comparing the state-space model described in (6) and (7) with the random-walk described in (10) and the linear regression model described in (9), the Kalman model corresponding to (10) and (9) is such that: $x_n$ corresponds to $w_n^o$, $H_n$ corresponds to $u_n$, $t_n$ corresponds to $q_n$, $z_n$ corresponds to $d_n$, and the matrices $F_n$ and $G_n$ are equal to $I_n$, where $I_n$ denotes the $M \times M$ identity matrix. Under these conditions, the number of matrix multiplications of equations (8b) and (8e) are reduced, resulting in a $O(M^2)$ computational complexity for the KF.

As will be shown in section IV, once the matrix $Q_n$ is known, the excess mean-square error (EMSE) obtained via KF can be closely approximated by the combination of two adaptive filters (namely LMS and RLS) and their respectively parameters, step-size $\mu$ and forgetting factor $\beta$, which are chosen optimally according to the equations [8]:

\[ \mu_o = \sqrt{\frac{\text{Tr}(Q_n)}{\sigma_v^2}}, \quad \beta_o = 1 + \frac{\text{Tr}(Q_n R_u)}{\sigma_v^2 M}, \]  
\[ \text{Eqn. (11)} \]

where $\mu_o$ and $\beta_o$ are the optimum tracking parameters, and $\sigma_v^2$ is the variance of the random measurement noise $v_n$.

The advantage is that the combination can be implemented with $O(M)$ complexity, while the Kalman Filter for model (10) requires computational complexity $O(M^2)$. Recall also that, using the method of [4], combination schemes can achieve a performance close to the optimum, even without knowledge of the true value of $Q_n$.

We also define the following error variables that are commonly used to characterize the performance of adaptive filters [8]:

- A priori filter error:
  \[ e_n^a = u_n \tilde{w}_n, \]
  \[ \text{where } \tilde{w}_n = w_n^o - w_n. \]

- Filter error:
  \[ e_n = d_n - u_n w_n = e_n^a + v_n. \]

- Excess Mean Square Error (EMSE):
  \[ \zeta_n = E\{(e_n^2)^2\} = E\{(e_n)^2\} - E\{(v_n)^2\}. \]

During their operation, adaptive filters normally go from a convergence phase, where the expected error decreases, to a steady-state regime in which the error tends towards some asymptotic value [14]. Table I presents the theoretical optimum EMSE expressions for each adaptive filter (with their respectively combination) [3].

<table>
<thead>
<tr>
<th>Alg.</th>
<th>$\zeta_o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS</td>
<td>$\sqrt{\sigma_v^2 \text{Tr}(R_u) \text{Tr}(Q_n)}$</td>
</tr>
<tr>
<td>RLS</td>
<td>$\sqrt{\sigma_v^2 M \text{Tr}(Q_n R_u)}$</td>
</tr>
<tr>
<td>Combination</td>
<td>$\zeta_{12} = \zeta_{12}^L &gt; \zeta_{12}^R$</td>
</tr>
</tbody>
</table>

where $\zeta_{1} = \zeta_{50}^{LMS}$, $\zeta_{2} = \zeta_{50}^{RLS}$ and $\zeta_{12}$ is given by [3]:

\[ \zeta_{12} = \mu_o \beta_o \sigma_v^2 \text{Tr}(\Sigma) + \text{Tr}(Q_n \Sigma), \]  
\[ \text{Eqn. (12)} \]

with $\Sigma = (\beta_o I_n + \mu_o R_u)^{-1} R_u$.

IV. SIMULATIONS

As shown in [15], when the LMS and RLS filters are combined following the same approach, an interesting result is obtained. Assuming the tracking model (10) and considering the optimum LMS and RLS filters with adaptation parameters given by expressions described in (11) [14], LMS will outperform RLS if $Q_n$ is proportional to the autocorrelation matrix of the input signal, $R_u$, and the opposite will occur when $Q_n \propto R_u^{-1}$.

Consider an example where $Q_n$ is a mixture of $R_u$ and $R_u^{-1}$ given by [14]:

\[ Q_n = 10^{-5} \left[ \alpha \frac{R_u}{\text{Tr}(R_u)} + (1 - \alpha) \frac{R_u^{-1}}{\text{Tr}(R_u^{-1})} \right], \]  
\[ \text{Eqn. (13)} \]

where $\alpha \in (0, 1)$.

Then, as can be seen in Fig. 2, the steady-state EMSE that can be achieved by combining both types of filters (LMS and RLS) with optimum settings, is within less than 1dB from the optimum EMSE obtained via KF. However, the computational cost is reduced from $O(M^2)$ to $O(M)$ operations if a lattice or DCD implementation are used for RLS. Other settings considered for this simulation were: $M = 7$, $\sigma_v^2 = 10^{-2}$, $R_u$ a Toeplitz matrix with first row given by

\[ \frac{1}{7} \begin{bmatrix} 1 & 0.8 & 0.8^2 & \ldots & 0.8^6 \end{bmatrix}, \]

$F_n = G_n = I_n$, $H_n = u_n$, $z_n = d_n$ and $R_n = \sigma_v^2$. For this experiment, 1000 simulations were performed in order to obtain ensemble average EMSE curves for each filter. In this example, $\mu$ and $\beta$ were chosen optimally according to (11).
Equation
Initialization: For $L=1$, $Q_n = \begin{bmatrix} X \end{bmatrix}$, $Q_n$ smoothly changes between $R_n$ and $R_n^{-1}$.

Even if an $M \times M$ positive-definite random perturbation (with spectral norm $10\%$ of the original $Q_n$) is added to the covariance matrix $Q_n$ at each realization (according to Table II), the combination still has an EMSE close to that of the KF (see Fig. 3). Note that the step size and forgetting factor were still chosen according to (11), but using the nominal value for $Q_n$, so the filters are not operating in the optimal condition anymore.

Table II

<table>
<thead>
<tr>
<th>Step</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Initialization: $Q_n = \text{Equation (13)}$</td>
</tr>
<tr>
<td>02</td>
<td>For realizations $L = 1, \ldots, 1000$</td>
</tr>
<tr>
<td>03</td>
<td>$X = \text{rand}(M)^t 10^{-6}$ % Auxiliary random variable</td>
</tr>
<tr>
<td>04</td>
<td>$[U, S, V] = \text{svd}(X)$ % Singular Value Decomp. of $X$</td>
</tr>
<tr>
<td>05</td>
<td>$\Delta = U \text{abs}(S)^t U$ % $M \times M$ random perturbation</td>
</tr>
<tr>
<td>06</td>
<td>$Q_n = Q_n + \Delta$ % $Q_n$ with random perturbation</td>
</tr>
</tbody>
</table>

V. 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 of LMS and RLS filters and compared the resulting EMSE with the optimal case obtained via Kalman Filter.

As it was shown, by using a convex combination between LMS and RLS, it is possible to achieve a steady-state EMSE performance close to the optimal case obtained via Kalman Filter, when a non-stationary environment is considered. The advantage arises from the fact that the combination can be implemented with $O(M)$ complexity, while it takes at least $O(M^2)$ operations to compute the corresponding Kalman Filter. Similar performance was still obtained, even without precise knowledge of the true value of $Q_n$.

References