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

Raffaello Claser and Vítor 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].

This work was partially funded by Fapesp grant 2014/04256-2 and CNPq grant 306268/2014-0.

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 $\mathcal{O}(M)$ operations (if lattice or Dichotomous Coordinate Descent - DCD algorithms are used) and the Kalman Filter requires $\mathcal{O}(M^2)$ operations (for a first-order random walk state-space model, see (10) below), or $\mathcal{O}(M^3)$ 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 respectively 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 σ_d^2, and let $\mathbf{u}_n$ be a $1 \times M$ zero-mean real-valued random input regressor vector with positive-definite covariance matrix denoted by $\mathbf{R}_u = \mathbf{E}\{\mathbf{u}_n^T \mathbf{u}_n\}$. Then, the solution $\mathbf{w}_n^o$ of the linear least-mean-squares problem

$$\min_w = \mathbf{E}\ |d_n - \mathbf{u}_n \mathbf{w}|^2, \tag{1}$$

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

$$\mathbf{w}_n = \mathbf{w}_{n-1} + f[d_n, \mathbf{u}_n, \mathbf{s}_n]. \tag{2}$$

Different adaptive schemes are characterized by their update functions $f[\cdot]$ ($\mathbf{s}_n$ 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 $\mathbf{w}_n$ at each iteration n, respectively [8], [9]:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mu \mathbf{u}_n^T [d_n - \mathbf{u}_n \mathbf{w}_{n-1}], \tag{3}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{P}_n\mathbf{u}_n^T[d_n - \mathbf{u}_n\mathbf{w}_{n-1}], \tag{4a}$$

$$\mathbf{P}_n = \beta^{-1}\mathbf{P}_{n-1} - \beta^{-2}\frac{\mathbf{P}_{n-1}\mathbf{u}_n^T\mathbf{u}_n\mathbf{P}_{n-1}}{1+\beta^{-1}\mathbf{u}_n\mathbf{P}_{n-1}\mathbf{u}_n^T}, \tag{4b}$$

where μ is a positive step-size (usually a small value), β is a forgetting factor, and $\mathbf{P}_n$ is an estimate of the inverse of the regressor autocovariance matrix $\mathbf{R}_\mathbf{u}$, 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 and is more robust [8], [9], [11]. Despite the fact that the standard RLS has $\mathcal{O}(M^2)$, fast $\mathcal{O}(M)$ 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 in order to minimize the error of the overall structure. Fig. 1 illustrates a block diagram of this combination.

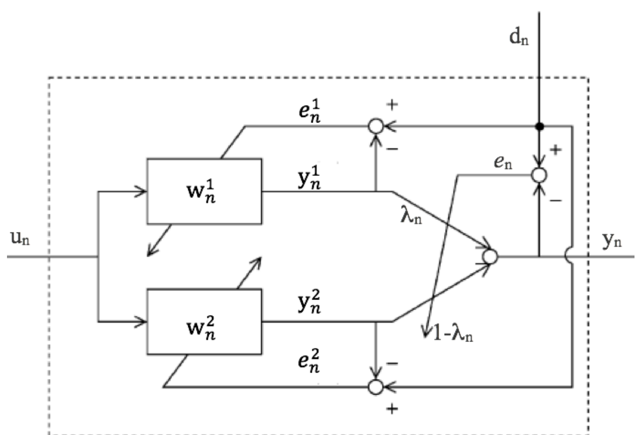

Figure 1. Block diagram of an adaptive convex combination of two transversal filters.

The output of the overall filter combination is denoted by

$$y_n = \lambda_n y_n^1 + [1-\lambda_n]y_n^2, \tag{5}$$

where y_n^1 and y_n^2 are the outputs of two transversal filters at time n, i.e, $y_n^i = \mathbf{u}_n\mathbf{w}_n^i, i = 1, 2$, e_n^1 and e_n^2 are the output filters errors ($e_n^i = d_n - y_n^i, i = 1, 2$), and λ_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 $\mathbf{w}_n^1$ and $\mathbf{w}_n^2$ can be extracted if λ_n is assigned with the appropriate values at each iteration n.

In all cases mentioned until now, no information about how the optimum solution $\mathbf{w}_n^o$ evolves with time was used to compute the solution $\mathbf{w}_n$. 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:

$$\mathbf{x}_n = \mathbf{F}_n\mathbf{x}_{n-1} + \mathbf{G}_n\mathbf{t}_n \tag{6}$$

$$\mathbf{z}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{v}_n \tag{7}$$

where $\mathbf{x}_n$ is the $M\times 1$ state vector, $\mathbf{F}_n$ is the $M\times M$ state-transition matrix, $\mathbf{G}_n$ is the $M\times N$ control-input model, $\mathbf{t}_n$ is a $N\times 1$ Gaussian random state noise vector with zero-mean and covariance matrix $\mathbf{T}_n$, $\mathbf{z}_n$ is the $D\times 1$ observation vector, $\mathbf{H}_n$ is the $D\times M$ measurement matrix, and $\mathbf{v}_n$ is a $D\times 1$ Gaussian random measurement noise vector with zero-mean and covariance matrix $\mathbf{R}_n$.

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

$$\mathbf{R}_{\mathbf{e},\mathbf{n}} = \mathbf{R}_n + \mathbf{H}_n\mathbf{P}_{n|n-1}\mathbf{H}_n^* \tag{8a}$$

$$\mathbf{K}_n = (\mathbf{F}_n\mathbf{P}_{n|n-1}\mathbf{H}_n^* + \mathbf{G}_n\mathbf{S}_n)\mathbf{R}_{\mathbf{e},\mathbf{n}}^{-\mathbf{1}} \tag{8b}$$

$$\mathbf{e}_n = \mathbf{z}_n - \mathbf{H}_n\hat{\mathbf{x}}_{n|n-1} \tag{8c}$$

$$\hat{\mathbf{x}}_{n+1|n} = \mathbf{F}_n\hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n\mathbf{e}_n \tag{8d}$$

$$\mathbf{P}_{n+1|n} = \mathbf{F}_n\mathbf{P}_{n|n-1}\mathbf{F}_n^* + \mathbf{G}_n\mathbf{T}_n\mathbf{G}_n^* - \mathbf{K}_n\mathbf{R}_{\mathbf{e},\mathbf{n}}\mathbf{K}_n^* \tag{8e}$$

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

$$\mathbf{P}_{n|n-1} = \mathbf{E}\{(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1})(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1})^T\},$$

$\mathbf{K}_n$ is the $M\times D$ Kalman gain, $\mathbf{S}_n$ is the cross-covariance matrix between the noise processes ($\mathbf{t}_n$ and $\mathbf{v}_n$) which is equal to $\mathbf{E}\{\mathbf{t}_n\mathbf{v}_n^*\}$ and $\mathbf{e}_n$ 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 $\mathbf{z}_0$ through $\mathbf{z}_{n-1}$.

As we can see in equations (8b) and (8e), in general one needs $\mathcal{O}(M^3)$ operations to compute the Kalman gain and the covariance matrix $\mathbf{P}_{n+1|n}$. 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.

ISBN 978-0-9928626-7-1 © EURASIP 2017

III. Data Model

In the sequel we adopt the following assumptions.

- d_n and $\mathbf{u}_n$ are related according to the following linear regression model
$$d_n = \mathbf{u}_n \mathbf{w}_n^o + v_n, \quad (9)$$
where v_n is i.i.d. noise, independent of $\mathbf{u}_n$ and with variance σ_v^2.
- $\mathbf{E}\{\mathbf{u}_n\} = 0$ and $\mathbf{E}\{d_n\} = 0$.
- $\mathbf{w}_n^o$ changes according to a first-order random walk
$$\mathbf{w}_n^o = \mathbf{w}_{n-1}^o + \mathbf{q}_n, \quad (10)$$
where $\mathbf{q}_n$ is independent of $\mathbf{u}_n$, with autocovariance $\mathbf{Q}_n$.

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: $\mathbf{x}_n$ corresponds to $\mathbf{w}_n^o$, $\mathbf{H}_n$ corresponds to $\mathbf{u}_n$, $\mathbf{t}_n$ corresponds to $\mathbf{q}_n$, $\mathbf{z}_n$ corresponds to d_n and the matrices $\mathbf{F}_n$ and $\mathbf{G}_n$ are equal to $\mathbf{I}_n$, where $\mathbf{I}_n$ 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 $\mathcal{O}(M^2)$ computational complexity for the KF.

As will be shown in section IV, once the matrix $\mathbf{Q}_n$ 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 μ and forgetting factor β, which are chosen optimally according to the equations [8]:

$$\mu_o = \sqrt{\frac{Tr\{\mathbf{Q}_n\}}{\sigma_v^2 Tr\{\mathbf{R}_u\}}} \quad , \quad \beta_o = 1 - \sqrt{\frac{Tr\{\mathbf{Q}_n \mathbf{R}_u\}}{\sigma_v^2 M}}, \quad (11)$$

where μ_o and β_o are the optimum tracking parameters, and σ_v^2 is the variance of the random measurement noise $\mathbf{v}_n$.

The advantage is that the combination can be implemented with $\mathcal{O}(M)$ complexity, while the Kalman Filter for model (10) requires computational complexity $\mathcal{O}(M^2)$. 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 $\mathbf{Q}_n$.

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

- A priori filter error:
$$e_n^a = \mathbf{u}_n \tilde{\mathbf{w}}_n,$$
where $\tilde{\mathbf{w}}_n = \mathbf{w}_n^o - \mathbf{w}_n$.
- Filter error:
$$e_n = d_n - \mathbf{u}_n \mathbf{w}_n = e_n^a + v_n.$$
- Excess Mean Square Error (EMSE):
$$\zeta_n = \mathbf{E}\{(e_n^a)^2\} = \mathbf{E}\{(e_n)^2\} - \mathbf{E}\{(v_n)^2\}.$$

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 I
OPTIMUM STEADY-STATE EMSES (ζ_o) FOR LMS, RLS AND THEIR COMBINATION.

Alg.	ζ_o
LMS	$\sqrt{\sigma_v^2 Tr\{\mathbf{R}_u\} Tr\{\mathbf{Q}_n\}}$
RLS	$\sqrt{\sigma_v^2 M Tr\{\mathbf{Q}_n \mathbf{R}_u\}}$
Combination	$\frac{\zeta_1 \zeta_2 - \zeta_{12}^2}{\zeta_1 - 2\zeta_{12} + \zeta_2}$

where $\zeta_1 = \zeta_o^{LMS}$, $\zeta_2 = \zeta_o^{RLS}$ and ζ_{12} is given by [3]:

$$\zeta_{12} = \mu_o \beta_o \sigma_v^2 Tr(\mathbf{\Sigma}) + Tr\{\mathbf{Q}_n \mathbf{\Sigma}\}, \quad (12)$$

with $\mathbf{\Sigma} = (\beta_o \mathbf{I}_n + \mu_o \mathbf{R}_u)^{-1} \mathbf{R}_u$.

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 $\mathbf{Q}_n$ is proportional to the autocorrelation matrix of the input signal, $\mathbf{R}_u$, and the opposite will occur when $\mathbf{Q}_n \propto \mathbf{R}_u^{-1}$.

Consider an example where $\mathbf{Q}_n$ is a mixture of $\mathbf{R}_u$ and $\mathbf{R}_u^{-1}$ given by [14]:

$$\mathbf{Q}_n = 10^{-5} \left[\alpha \frac{\mathbf{R}_u}{Tr(\mathbf{R}_u)} + (1-\alpha) \frac{\mathbf{R}_u^{-1}}{Tr(\mathbf{R}_u^{-1})} \right] , \quad (13)$$

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 $\mathcal{O}(M^2)$ to $\mathcal{O}(M)$ operations (if a lattice or DCD implementation are used for RLS). Other settings considered for this simulation were: $M = 7$, $\sigma_v^2 = 10^{-2}$, $\mathbf{R}_u$ a Toeplitz matrix with first row given by
$$\frac{1}{7} \begin{bmatrix} 1 & 0.8 & 0.8^2 & \dots & 0.8^6 \end{bmatrix},$$
$\mathbf{F}_n = \mathbf{G}_n = \mathbf{I}_n$, $\mathbf{H}_n = \mathbf{u}_n$, $\mathbf{z}_n = d_n$ and $\mathbf{R}_n = \sigma_v^2$. For this experiment, 1000 simulations were performed in order to obtain ensemble average EMSE curves for each filter. In this example, μ and β were chosen optimally according to (11).

ISBN 978-0-9928626-7-1 © EURASIP 2017

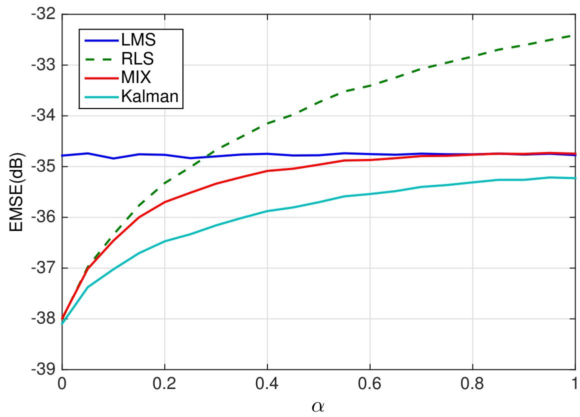

Figure 2. Simulated steady-state EMSE for LMS, RLS, their convex combination and KF, when $\mathbf{Q}_n$ smoothly changes between $\mathbf{R}_u$ and $\mathbf{R}_u^{-1}$.

Even if an $M \times M$ positive-definite random perturbation (with spectral norm 10% of the original $\mathbf{Q}_n$) is added to the covariance matrix $\mathbf{Q}_n$ 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 $\mathbf{Q}_n$, so the filters are not operating in the optimal condition anymore.

Table II
COVARIANCE MATRIX $\mathbf{Q}_n$ WITH RANDOM PERTURBATION.

Step	Equation
01	Initialization: $\mathbf{Q}_n$ = Equation (13)
02	For realizations L = 1, . . . , 1000
03	$\mathbf{X} = randn(M)^*10^{-6}$ % Auxiliary random variable
04	$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathbf{X})$ % Singular Value Decomp. of X
05	$\mathbf{\Delta} = \mathbf{U}^* abs(\mathbf{S})^* \mathbf{U}'$ % $M \times M$ random perturbation
06	$\mathbf{Q}_n = \mathbf{Q}_n + \mathbf{\Delta}$ % $\mathbf{Q}_n$ with random perturbation

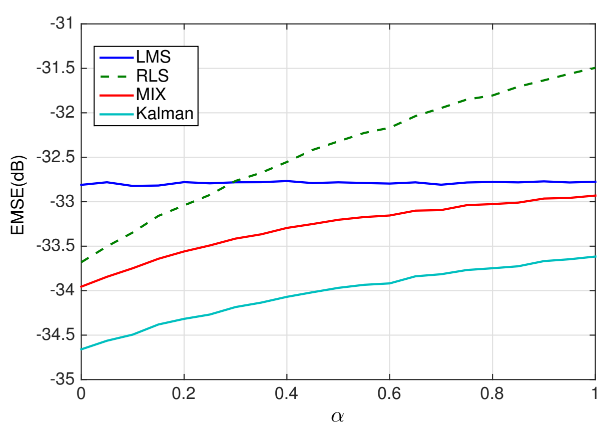

Figure 3. Simulated steady-state EMSE for LMS, RLS, their convex combination and KF, when a random perturbation is added to $\mathbf{Q}_n$.

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 $\mathcal{O}(M)$ complexity, while it takes at least $\mathcal{O}(M^2)$ operations to compute the corresponding Kalman Filter. Similar performance was still obtained, even without precise knowledge of the true value of $\mathbf{Q}_n$.

References

[1] 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.

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

[3] 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.

[4] J. Arenas-García, 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.

[5] 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.

[6] M. Niedzwiecki, "Identification of nonstationary stochastic systems using parallel estimation schemes," *IEEE Transactions on Automatic Control*, vol. 35, no. 3, pp. 329–334, 1990.

[7] ——, "Multiple-model approach to finite memory adaptive filtering," *IEEE Transactions on signal processing*, vol. 40, no. 2, pp. 470–473, 1992.

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

[9] 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.

[10] F. G. A. Neto, V. H. Nascimento, and Y. V. Zakharov, "Low-complexity widely linear RLS filter using dcd iterations."

[11] B. Hassibi, A. H. Sayed, and T. Kailath, "$\mathcal{H}_\infty$ optimality of the LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 267–280, 1996.

[12] 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.

[13] N. Kovvali, M. Banavar, and A. Spanias, "An introduction to Kalman filtering with Matlab examples," *Synthesis Lectures on Signal Processing*, vol. 6, no. 2, pp. 1–81, 2013.

[14] J. Arenas-Garcia, L. A. Azpicueta-Ruiz, M. T. Silva, V. H. Nascimento, and A. H. Sayed, "Combinations of adaptive filters: performance and convergence properties," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 120–140, 2016.

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

ISBN 978-0-9928626-7-1 © EURASIP 2017