A Data Reusage Algorithm Based on Incremental Combination of LMS Filters

Luiz F. O. Chamon, Humberto F. Ferro and Cássio G. Lopes
Dept. of Electronics Systems
University de São Paulo, Brazil
chamon@usp.br, hferro@usp.com, cassio@lps.usp.br

***Abstract*—This work introduces a new data reuse algorithm based on the incremental combination of LMS filters. It is able to outperform the Affine Projection Algorithm (APA) in its standard form, another well-known data reuse adaptive filter. First, the so called true gradient data reuse LMS—sometimes referred to as data reuse LMS—is shown to be a limiting case of the regularized APA. Afterwards, an incremental counterpart of its recursion is inspired by distributed optimization and adaptive networks scenarios. Simulations in different scenarios show the efficiency of the proposed data reuse algorithm, that is able to match and even outperform the APA in the mean-square sense at lower computational complexity.**

I. Introduction

Since the early ages of adaptive filtering, many efforts have been made to derive adaptive algorithms with faster convergence than the classical LMS filter while retaining its low complexity. Among the different solutions proposed, data reuse (DR) has become a very successful technique to accelerate convergence, although at the cost of being inefficient implementation-wise [1].

One of the most celebrated DR adaptive filters (AFs) is the Affine Projection Algorithm (APA) [2]. Although other DR algorithms have been proposed [3]–[6], they are rarely compared to the APA due to its superior performance. For instance, in the case of speech echo cancellation the APA has almost the same performance as a Fast Recursive Least Square (FRLS) filter with roughly 3 times less operations [7].

Another recently introduced technique to improve filtering performance is the combination of AFs. This approach consists of aggregating a pool of AFs through mixing parameters, adaptive or not, and attempting to achieve universality, i.e., making the overall system at least as good—usually in the mean-square sense—as the best filter in the set. Combinations with different step sizes, orders, adaptive algorithms, topologies, and supervising rules can be found in [8]–[14].

In this work, a solution merging combinations of AFs and DR is devised by:

- Extending the concept of combination of AFs to accommodate for DR;
- Showing that the APA is, under certain conditions, equivalent to the true gradient DR LMS (TRUE-LMS) [3], [4];
- Devising an incremental counterpart of the TRUE-LMS;
- Showing through simulations that the proposed algorithm can either match or outperform the regularized APA with lower complexity [15], [16].

Mr. Chamon's work is supported by the Coordination for the Improvement of Education Personnel (CAPES, Brazil) and Mr. Ferro and Mr. Lopes are supported by the National Council for Scientific and Technological Development (CNPq, Brazil).

II. Problem formulation

In a system identification scenario, consider the measurements $d(i) = u_i w^o + v(i)$, where w^o is an $M \times 1$ vector that models the unknown system, u_i is the $1 \times M$ regressor vector that captures samples $u(i)$ of an input signal with variance σ_u^2, and $v(i)$ the realization of an i.i.d. process with variance σ_v^2. An AF can then be described as

$$w_i = w_{i-1} + \mu p, \tag{1}$$

in which w_i is an estimate of w^o at iteration i, μ is a step size, and $p = -B\,\nabla^* J(w_{i-1})$ is the update direction vector, with B any positive definite matrix and $J(w_{i-1})$ the underlying cost function the filter attempts to minimize. The operator * denotes conjugate transposition [16].

Different choices of p lead to different adaptive algorithms, such as

$$w_i = w_{i-1} + \mu u_i^* e(i) \qquad \text{(LMS)} \tag{2}$$

$$w_i = w_{i-1} + \mu \frac{u_i^*}{\|u_i\|^2 + \epsilon} e(i) \qquad \text{(NLMS)}, \tag{3}$$

where $0 < \epsilon \ll 1$ is a regularization factor and $e(i) = d(i) - u_i w_{i-1}$ is the output estimation error [16].

A. Data reuse

Data reuse consists either of using $K > 1$ times a single data pair $\{u_i, d(i)\}$ or operating over a set of past data pairs $\{U_i, d_i\}$, where $U_i = [\ u_i^T \ \cdots \ u_{i-K+1}^T\ ]^T$ is a $K \times M$ regressors matrix and $d_i = [\ d(i) \ \cdots \ d(i-K+1)\ ]^T$ is a $K \times 1$ measurements vector. Such algorithms are particularly appropriate in communication systems, where the rate of signaling is constrained by bandwidth restrictions, or speech applications, where data is inherently intermittent [1], [17]. Moreover, most of these AFs are able to trade off complexity and performance by changing K, which is invaluable in many scenarios [6].

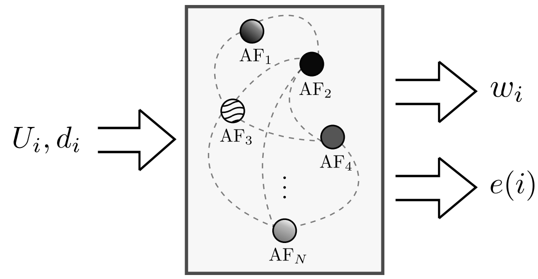

Figure 1. Combination of AFs with DR (dashed lines imply information exchange)

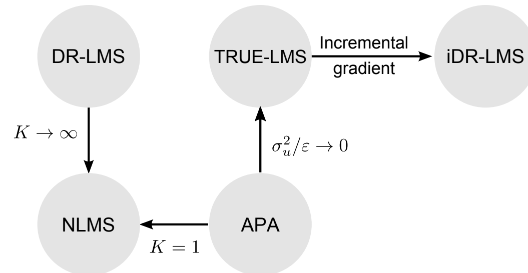

Figure 2. Diagram of relations between DR algorithms

The first DR algorithm, the data reuse LMS (DR-LMS), was introduced in [18] and operated K times over a same data pair. Explicitly,

$$\begin{aligned} w_{0,i} &= w_{i-1} \\ w_{k,i} &= w_{k-1,i} + \mu u_i^* [d(i) - u_i w_{k-1,i}] \\ w_i &= w_{K,i}. \end{aligned} \tag{4}$$

Another algorithm, also commonly referred to as DR-LMS, extends the LMS filter for data sets of the type $\{U_i, d_i\}$, leading to

$$w_i = w_{i-1} + \mu U_i^* e_i, \tag{5}$$

where $e_i = d_i - U_i w_{i-1}$ [3]. In order to avoid confusion with (4), this algorithm will be renamed true gradient data reuse LMS (TRUE-LMS), in reference to true gradient distributed optimization techniques (see Section IV).

Last, the APA's standard form, an acclaimed DR algorithm, is given by [16]

$$w_i = w_{i-1} + \mu U_i^* \left(\epsilon I + U_i U_i^*\right)^{-1} e_i. \tag{6}$$

B. Combination of AFs

In combinations, a pool of N AFs, called the *components*, are aggregated so as to provide a better overall algorithm, in a particular way, than any of the individual filters. So far in the literature, all components operated over the same data pair $\{u_i, d(i)\}$ [9]–[14], contrasting with the definitions introduced hereafter, which allow for a more general use of the available data (Fig. 1).

First, the components are identified by introducing an index $n = 1, \ldots, N$, as in $w_{n,i}$, the n-th component coefficients[2]. For the sake of simplicity, the following presentation will be restricted to LMS component filters for a set of mixing parameters $\{\eta_n(i)\}$, although it can be easily extended to different AFs. Then, parallel and incremental combinations are written as

Definition 1. *Parallel combinations* [9]–[11]

$$\begin{aligned} w_{n,i-1} &= \delta(i - rL)\, w_{i-1} + (1 - \delta(i - rL))\, w_{n,i-1} \\ w_{n,i} &= w_{n,i-1} + \mu_n u_{n,i}^* [d_n(i) - u_{n,i}\, w_{n,i-1}] \\ w_i &= \sum_{n=1}^{N} \eta_n(i)\, w_{n,i} \end{aligned} \tag{7}$$

where $\delta(\cdot)$ is Kronecker's delta and $r \in \mathbb{N}$.

Definition 2. *Incremental combinations* [12]

$$\begin{aligned} w_{0,i} &= w_{i-1} \\ w_{n,i} &= w_{n-1,i} + \eta_n(i) \mu_n u_{n,i}^* [d_n(i) - u_{n,i}\, w_{n-1,i}] \\ w_i &= w_{N,i} \end{aligned} \tag{8}$$

Note that for $u_{n,i} = u_i$ and $d_n(i) = d(i)$ the commonly used combinations are recovered [9]–[12]. Moreover, as $L \to +\infty$, the coefficients feedback [9] is bypassed, resulting in a parallel-independent combination [10], [11].

An important consequence of this novel approach is that DR algorithms are easily cast in this framework, bringing a new interpretation for these AFs. For example, taking the parallel combination (7) with $N = K$, $L = 1$, $\mu_n = \mu$, $u_{n,i} = u_{i-n+1}$, $d_n(i) = d(i - n + 1)$, and $\eta_n(i) = 1/N$ leads to the TRUE-LMS recursion in (5). Similarly, the incremental combination from (8) with $N = K$, $\mu_n = \mu$, $u_{n,i} = u_i$, $d_n(i) = d(i)$, and $\eta_n(i) = 1$ is identical to the DR-LMS (4).

III. APA and TRUE-LMS

Other relations involving DR algorithms, like the one described in the previous section, have already been found (Fig. 2). In [17], the NLMS was proved to be a limiting case ($K \to \infty$) of the DR-LMS. A straightforward association also exists between the APA and the NLMS: both algorithms are equivalent for $K = 1$ [16].

In the sequel, a new relation is derived showing that, when over-regularized and/or under vanishing inputs, the APA recursion converges to that of the TRUE-LMS in (5). In the APA update from (6), writing the step size as $\mu = \mu' \epsilon$, $\mu' > 0$, in the limiting condition $\sigma_u^2/\epsilon \to 0$ one has

$$\begin{aligned} &\lim_{\sigma_u^2/\epsilon \to 0} (\mu' \epsilon)\, U_i^* (\epsilon I + U_i U_i^*)^{-1} e_i = \\ &\mu'\, U_i^* \left[\lim_{\sigma_u^2/\epsilon \to 0} \epsilon (\epsilon I + U_i U_i^*)^{-1} \right] e_i = \\ &\mu'\, U_i^* \left[I + \lim_{\sigma_u^2/\epsilon \to 0} \epsilon^{-1} U_i U_i^* \right]^{-1} e_i. \end{aligned} \tag{9}$$

The following theorem establishes the conditions under which the limit in (9) is null and the TRUE-LMS update is recovered.

Definition 3. $f(x)$ is said to be *superlinear* if

$$\lim_{x \to 0^\pm} \frac{f(x)}{|x|} = 0.$$

2 Note that in this work k indexes data, while n indexes component filters.

Theorem 1. *Let the $\{u(i)\}$ collected into the vector u_i be realizations of a zero-mean complex wide-sense stationary random variable (RV) $\mathbf{u}$ with correlation function $\rho(l) = \mathrm{E}\,\mathbf{u}^*(i)\mathbf{u}(i+l)$, $\rho(0) = \sigma_u^2$. Over any path where σ_u^2/ϵ is a superlinear function of $1/M$, on has*

$$\lim_{\sigma_u^2/\epsilon\to 0} \epsilon^{-1} U_i U_i^* = 0 \text{ (a.s.).} \tag{10}$$

Proof: Multiplying (10) by M/M evaluated at the limiting condition $M \to \infty$ gives

$$\lim_{(M,\sigma_u^2/\epsilon)\to(+\infty,0)} \frac{U_i U_i^*}{M} M\epsilon^{-1}$$

The strong law of large numbers [19] guarantees that

$$\lim_{M\to+\infty} \frac{U_i U_i^*}{M} = \begin{bmatrix} \sigma_u^2 & \cdots & \rho(M-1) \\ \vdots & \ddots & \vdots \\ \rho(M-1) & \cdots & \sigma_u^2 \end{bmatrix} \text{ (a.s.),}$$

which makes (10) equal to

$$\lim_{(M,\sigma_u^2/\epsilon)\to(+\infty,0)} \begin{bmatrix} \sigma_u^2 & \cdots & \rho(M-1) \\ \vdots & \ddots & \vdots \\ \rho(M-1) & \cdots & \sigma_u^2 \end{bmatrix} M\epsilon^{-1}.$$

For the diagonal elements, choosing $\sigma_u^2/\epsilon = f(1/M)$ yields

$$\lim_{(M,\sigma_u^2/\epsilon)\to(+\infty,0)} M\frac{\sigma_u^2}{\epsilon} = \lim_{M\to+\infty} \frac{f(1/M)}{1/M} = \lim_{x\to 0^+} \frac{f(x)}{x} = 0$$

for any superlinear function f.

Moreover, it is a property of the correlation function that $|\rho(l)| < \sigma_u^2, \forall\, l \neq 0$, i.e., the magnitude of the off-diagonal elements of the covariance matrix is upper bounded by σ_u^2 [19]. Therefore, the above convergence condition applies not only to the diagonal elements but to the whole matrix. □

Applying Theorem 1 to (9) and rewriting the complete recursions leads to

$$w_i = w_{i-1} + \mu'\, U_i^* e_i,$$

which is indeed the TRUE-LMS algorithm from (5).

IV. The incremental DR-LMS

Although related to the APA, the TRUE-LMS is outperformed by the latter in most scenarios [3], [20]. Nonetheless, the connection derived in Section III suggests the possibility of improving the TRUE-LMS algorithm so that its adaptation capabilities become comparable to those of the APA.

To this end, note that the underlying cost function of the TRUE-LMS can be written as $J_{TRUE}(w) = \sum_k |e_k(w)|^2$, with $e_k(w) = d(i-k+1) - u_{i-k+1}\, w$. It is known from distributed optimization theory that cost functions that can be decomposed this way give rise to two gradient-based minimization procedures [21]. On one hand, *true gradient* methods evaluate all gradients $\nabla J_k(\cdot)$ at a global estimate w of the optimal solution. This is the approach taken in the TRUE-LMS adaptation (see (5)) and it is the reason behind its renaming from DR-LMS in this work. On the other hand, *incremental gradient* techniques evaluate each $\nabla J_k(\cdot)$ at a intermediate estimate w_{k-1} obtained from the previous partial cost function [22]. Observations advocate that incremental strategies have faster convergence then true gradient ones when far from the steady-state solution [21].

These arguments suggest an incremental counterpart of the TRUE-LMS as a candidate comparable to the APA. From the framework put forward in Section II-B, this transition can be seen as a topological change in a combination of LMS filters (see (7) and (8)). Another interpretation of this algorithm is as a particular case of the distributed algorithm from [22] (INC-LMS), where now all nodes statistics are the same. Indeed, the incremental data reuse LMS (iDR-LMS) recursion is

$$\begin{aligned} w_{0,i} &= w_{i-1} \\ w_{n,i} &= w_{n-1,i} + \mu_n u_k^*[d(k) - u_k w_{n-1,i}] \\ w_i &= w_{N,i}, \end{aligned} \tag{11}$$

where $n = 1, \ldots, N$ and $k = i - (n-1) \bmod K$. For the special case where $N = K$, the algorithm introduced in a different context in [4] is recovered. In this work, however, the conceptual comparison to the APA motivates the use of $N > K$ in (11), so as to better exploit the data and improve performance. This is accomplished by defining the index k, which goes over the data set again once n exceeds the set's boundary (K).

The recursion (11) is composed of N LMS filters, making its complexity $O(NM)$. Hence, it requires less operations than the classical APA ($O(K^2M)$ [16]) and even fast APA implementations (e.g., $O(3K^2) + O(KM)$ [15]) for a large range of N. As simulations show, the iDR-LMS matches or outperforms the APA for $N \ll K^2$, i.e., at a lower computational cost.

Besides its complexity, another important feature of this structure is that it inherits the well-known robustness and stability properties of the LMS filters it is based on [23]. When contrasted to the known computational problems of the APA in lower precision environments [24], the iDR-LMS is expected to show even better performance in embedded applications. Furthermore, note that in (11) the updates of iDR-LMS are made in blocks, i.e., after going through all components. Although this keeps the algorithm in the same time scale as the APA, a different time scale could be obtained by updating the overall coefficients w_i for every new $w_{n,i}$.

V. Simulations

This section starts by illustrating the relation derived in Section III between the APA and the TRUE-LMS. Both algorithms are used to identify an unknown system of length $M = 16$ modeled as $w^o = \mathrm{col}\{1\}/\sqrt{M}$. Figure 3 shows the result of changes in the regularization factor (ϵ) for $K = 4$, $\mu' = 0.05$, $\mu_{TRUE} = \mu'$, $\mu_{APA} = \mu'\epsilon$, and $u(i)$ and $v(i)$ are realizations of a zero-mean Gaussian i.i.d. RVs with $\sigma_u^2 = 1$ and $\sigma_v^2 = 10^{-3}$. Likewise, Figure 4 displays the effect of vanishing inputs ($\sigma_u^2 \to 0$) for $K = 4$, $\mu_{TRUE} = \mu_{APA} = 0.3/\sigma_u^2$, $\epsilon = 0.1$, and $\sigma_v^2 = 10^{-6}$. The learning curves ($\mathrm{MSD}(i) = \mathrm{E}\,|w^o - w_{i-1}|^2$) were averaged over 100 independent runs. Note that μ_{APA} and σ_v^2 were chosen for clarity's sake, so as to emphasize the relation between APA and TRUE-LMS. Also, in Fig. 4 the step sizes were increased as the variance decreased to keep the all curves in the same time frame. Nonetheless, it is clear that as $\sigma_u^2/\epsilon \to 0$, the APA's behavior approaches that of the TRUE-LMS.

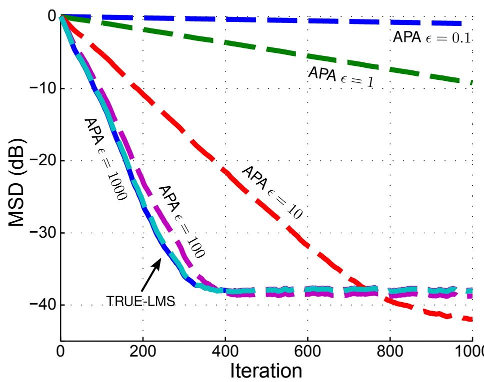

Figure 3. TRUE-LMS and APA for different ϵ

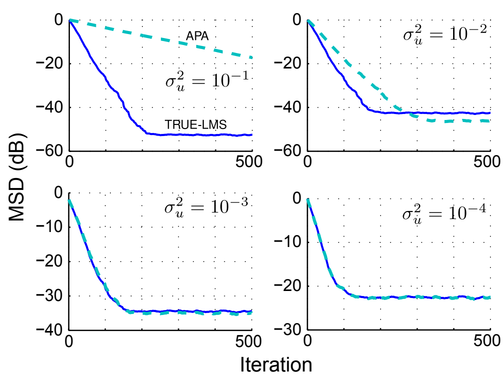

Figure 4. TRUE-LMS and APA for different σ_u^2

Next, the performance of the novel iDR-LMS is compared to that of the APA under different system identification scenarios. Figures 5, 6, and 7 present stationary environments where $M = 100$, $w^o = \text{col}\{1\}/\sqrt{M}$, $\mu_0 = 0.05$, $\mu_{APA} = \mu_0$, $\mu_n = \mu_0/M\sigma_u^2$ and x and v, zero-mean Gaussian i.i.d. sequences with variances $\sigma_x^2 = 1$ and $\sigma_v^2 = 10^{-3}$. For the white input experiments, $u(i) = x(i)$, and for the correlated input ones, $u(i) = \alpha\, u(i-1) + \sqrt{1-\alpha^2}\, x(i)$, with $\alpha = 0.95$. Figure 8, in addition to the colored input signal, simulates a non-stationary system of length $M = 60$ modeled as $w_i^o = w_{i-1}^o + q(i)$, where $q(i)$ is a zero-mean i.i.d. Gaussian RV with variance $\sigma_q^2 = 10^{-6}$. In this case, $\mu_{APA} = 0.1$ and $\mu_n = 0.003$. All learning curves are averages of 100 independent trials.

In every simulated scenarios, the iDR-LMS is able to match or outperform the APA for some N relatively close to K, keeping its complexity lower than classical and fast APAs [15], [16]. Finally, Fig. 5 and 6 reveal a trade off between steady state error and convergence speed imposed by the number of components (N) that bears significant similarity to the known compromise for step sizes in single AFs [16].

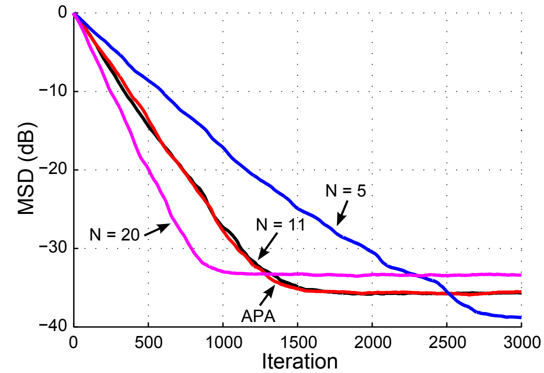

Figure 5. APA and iDR-LMS for white signals and $K = 10$

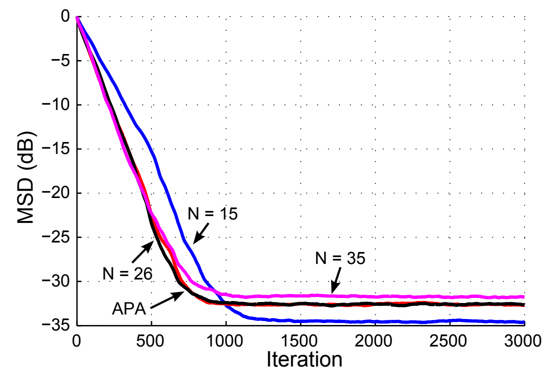

Figure 6. APA and iDR-LMS for white signals and $K = 20$

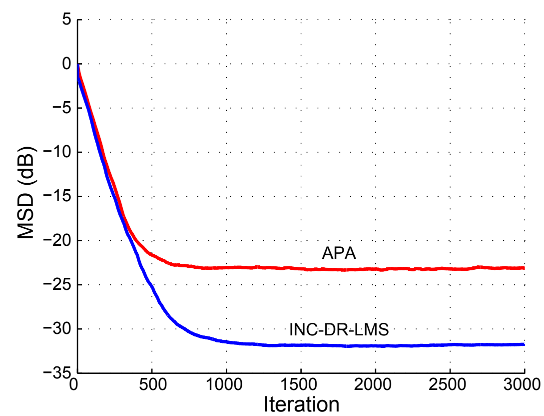

Figure 7. APA and iDR-LMS for correlated inputs, $K = 10$, and $N = 20$

VI. Conclusion

A new DR algorithm based on the incremental combination of LMS filters was introduced by putting forward new definitions of combinations that include DR strategies. After illustrating the relation between combinations of AFs and DR algorithms, the TRUE-LMS was proven to be a limiting case of the APA. The iDR-LMS was then introduced as an incremental counterpart of its recursion. Simulations showed that the new algorithm can match and even outperform the APA with lower complexity. Future works include performance analysis, inspired by adaptive networks [22], and solutions to reduce the steady state/convergence rate trade off.

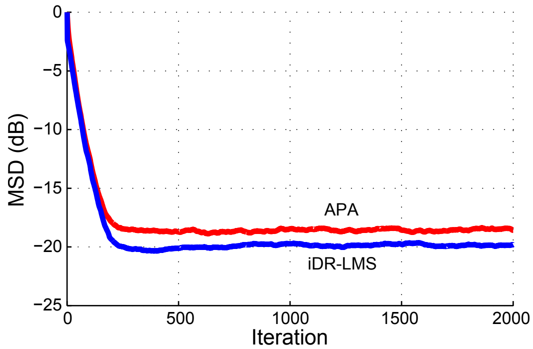

Figure 8. APA and iDR-LMS for correlated inputs, non-stationary system, $K = 10$, and $N = 20$

REFERENCES

[1] J. Benesty and T. Gänsler, "On data-reuse adaptive algorithms," in *International Workshop on Acoustic Echo and Noise Control*, 2003, pp. 31–34.

[2] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Elec. Comm. Japan*, vol. J67-A, pp. 126–132, 1984.

[3] H.-C. Shin, W.-J. Song, and A.H. Sayed, "Mean-square performance of data-reusing adaptive algorithms," *IEEE Signal Process. Lett.*, vol. 12[12], pp. 851–854, 2005.

[4] B.A. Schnaufer and W.K. Jenkins, "New data-reusing LMS algorithms for improved convergence," in *Asilomar Conf. on Signals, Syst. and Comput.*, 1993, pp. 1584–1588.

[5] J. Apolinário Jr., M.L.R. Campos, and P.S.R. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," *IEEE Trans. Signal Process.*, vol. 48[11], pp. 3235–3242, 2000.

[6] R.A. Soni, K.A. Gallivan, and W.K. Jenkins, "Low-complexity data reusing methods in adaptive filtering," *IEEE Trans. Signal Process.*, vol. 52[2], pp. 394–405, 2004.

[7] Y.A. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*, Springer, 2006.

[8] L.F.O. Chamon and C.G. Lopes, "Combination of adaptive filters for relative navigation," in *European Signal Processing Conference*. EURASIP, 2011, pp. 1771–1775.

[9] L.F.O. Chamon, W.B. Lopes, and C.G. Lopes, "Combination of adaptive filters with coefficients feedback," in *International Conference in Acoustic, Speech, and Signal Processing*. IEEE, 2012, pp. 3785–3788.

[10] J. Arenas-García, A.R. Figueiras-Vidal, and A.H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54[3], pp. 1078–1090, 2006.

[11] 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[5], pp. 1853—-1864, 2008.

[12] W.B. Lopes and C.G. Lopes, "Incremental-cooperative strategies in combination of adaptive filters," in *International Conference in Acoustic, Speech, and Signal Processing*. IEEE, 2011, pp. 4132–4135.

[13] C.G. Lopes, E.H. Satorius, P. Estabrook, and A.H. Sayed, "Adaptive carrier tracking for mars to earth communications during entry, descent, and landing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46[4], pp. 1865–1879, 2010.

[14] M.T.M. Silva and V.H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. Signal Process.*, vol. 56[7], pp. 3137—-3149, 2008.

[15] M.C. Tsakiris and P.A. Naylor, "Fast exact affine projection algorithm using displacement structure theory," in *International Conference on Digital Signal Processing*, 2009.

[16] A.H. Sayed, *Adaptive filters*, Wiley-IEEE Press, 2008.

[17] S. Roy and J.J. Shynk, "Analysis of the data-reusing LMS algorithm," in *Midwest Symposium on Circuits and Systems*, 1989, pp. 1127–1130.

[18] S. Shaffer and C.S. Williams, "Comparison of LMS, alpha LMS, and data reusing LMS algorithms," in *Asilomar Conference on Circuits, Systems, and Computers*, 1983, pp. 260–264.

[19] A. Leon-Garcia, *Probability, Statistics, and Random Processes For Electrical Engineering*, Prentice Hall, 2008.

[20] H.-C. Shin and A.H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 52[1], pp. 90–102, 2004.

[21] D.P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM J. Optim.*, vol. 7[4], pp. 913–926, 1997.

[22] C.G. Lopes and A.H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Process.*, vol. 55[8], pp. 4064–4077, 2007.

[23] B. Hassibi, A.H. Sayed, and T. Kailath, "H^∞ optimality of the LMS algorithm," vol. 44[2], pp. 267–280, 1996.

[24] M.C. Tsakiris and C.G. Lopes, "A robust affine projection algorithm with feedback compensation of the condition number," in *International Conference in Acoustic, Speech, and Signal Processing*, 2009, pp. 2057–2060.