Adaptive Filter Theory and Application for the Identification of Sparse Systems

Cyro S. Hemsi

Abstract— **One of the most popular adaptive techniques available is the stochastic gradient algorithm, particularly a very simple implementation, the Least Mean Squares (LMS). In this paper, we focus on identifying sparse systems, as is often the case in telecommunications and acoustics applications. In this context, conventional adaptive filters, such as the LMS, are not able to exploit prior knowledge on the system sparsity, so sparse AFs have been shown to be more advantageous, as discussed in this paper. Initially, we offer a concise review on the AFs theory, including some sparse algorithms. Next, we propose a sparse AF derived from the LMS algorithm and the SparseStep approximation of the ℓ_0-pseudo-norm penalty. The proposed AF takes sparsity into account both to accelerate convergence and improve performance. Finally, the proposed filter is numerically validated by comparing it with well-known AFs.**

Keywords— **Adaptive Filters, System Identification, Sparsity, LMS Algorithm.**

I. INTRODUCTION

A useful approach for solving filter-optimization problems is to try to minimize the mean-squared error (MSE) between the filter output and the desired (or reference) signal. The resulting Wiener filter is optimum in the mean-squared sense under stationary conditions, but required *a priori* knowledge of the input signal statistics. In practice, a more effective solution resorts on the use of an adaptive filter (AF) that is able to self-tune its response through a recursive algorithm, thus also adjusting for non-stationary scenarios [1], [2], [3], [4].

One of the most popular AF techniques available is the stochastic gradient algorithm, particularly a very simple implementation, the Least Mean Squares (LMS) and its variant, the Normalized LMS (NLMS). Both filters are widely used in various applications in telecommunications and acoustics, such as system identification, channel equalization, echo and interference cancellation, multiple-input multiple-output (MIMO) channel estimation, and so forth. Note that, although other adaptive algorithms, such as the Recursive Least Squares (RLS), may have superior performance, in most practical applications, LMS-based algorithms are chosen due to their simplicity, low computational cost, good performance and robustness.

Nevertheless, many of the unknown systems to be identified in telecommunications and acoustics are sparse, i.e. few of the impulse response coefficients have large magnitude while most of them are close to zero, as in broadband wireless and digital TV channels subject to multipath and echo effects. In this context, conventional AFs, such as the LMS, although widely used in practice, are not able to exploit prior knowledge about the sparsity of the system. For example, [5] shows that the direct application of NLMS to a sparse system identification problem (long echo cancellation) resulted in unsatisfactory performance, due to adaptation noise of close-to-zero coefficients.

Several works in the literature discuss the use of sparse AFs, as in [6], [7], [8]. A common approach to achieve the sparsest set of the system's dominant coefficients is to directly penalize the number of non-zero elements in the solution, i.e. the ℓ_0-pseudo-norm of the coefficients vector. However, this approach requires high computational cost (exhaustive search in the space of solutions), thus being a non-polynomial (NP)-hard problem. Another approach is to penalize the sum of the magnitudes of the elements in the solution, that is, the ℓ_1-norm of the coefficients vector, by means of the least-absolute shrinkage and selection operator (LASSO) [11]. This paper proposes sparse AFs derived from a cost function that is penalized by an approximate ℓ_0-pseudo-norm, using the SparseStep approximation from [9], [10], which is a simple but precise continuous function that allows the problem to be computationally tractable. Comparing to the other approaches above, the proposed method is more attractive for adaptive estimation due to its lower complexity. This way, the proposed AFs take advantage of the sparsity knowledge, both to accelerate the convergence of the algorithm and improve the systems identification performance, with wide range of applications in telecommunications and acoustics.

The contributions of this paper are twofold: (a) We offer a concise review on the AF theory, including sparse techniques; (b) Furthermore, we propose an update equation for a new sparse AF based on an approximation of the ℓ_0-pseudo-norm. Finally, the performance of the proposed filter is numerically validated by comparison with well-known AFs.

This paper is organized as follows. Section 2 provides background on the theory of adaptive filters; in Section 3, the proposed update equation is derived. Section 4 provides numerical simulation and results, followed by Section 5, where final conclusions are drawn.

II. PROBLEM FORMULATION

In the optimal linear least-squares filter problem, we suppose that the desired signal $d(n)$ at a discrete time n is a realization of a scalar complex-valued random variable (RV) with zero mean and variance $\sigma_d^2 = \mathbb{E}[|d(n)|^2]$ (where $\mathbb{E}[\cdot]$ is the expectation operator). The desired signal has to be

Cyro S. Hemsi received his Ph.D. (2021) from the Department of Telecommunication and Control (PTC) of the University of São Paulo (USP), São Paulo-Brazil. E-mail: cyro.hemsi@alumni.usp.br.

estimated by the linear filter from the complex-valued input vector $\boldsymbol{u}(n)$, of size $M \times 1$, also a sequence of realizations of a RV with zero mean, whose positive-definite covariance matrix is $\boldsymbol{R}_u \triangleq \mathbb{E}[\boldsymbol{u}(n)\boldsymbol{u}^H(n)] > 0$ (where $(\cdot)^H$ is the complex-conjugate transpose) and cross-covariance vector is $\boldsymbol{R}_{du} \triangleq \mathbb{E}[\boldsymbol{u}(n)d^*(n)]$ (where $(\cdot)^*$ is the complex-conjugate). Assuming the signals are stationary, the problem of estimating d from $\boldsymbol{u}$ in the linear least mean squares sense is defined as:

$$\min_{\boldsymbol{w}\in\mathbb{C}^{M\times 1}} \quad J(\boldsymbol{w}) \triangleq \mathbb{E}[e(n)e^*(n)], \tag{1}$$

where $\boldsymbol{w}$ is the $M \times 1$ vector of the underlying finite-impulse response (FIR) filter tap-weights, $\boldsymbol{u}(n) = \begin{bmatrix} u(n) & u(n-1) & \dots & u(n-M+1) \end{bmatrix}^T$ is the tap-input vector (where $(\cdot)^T$ is the transpose) and $e(n) = d(n) - \boldsymbol{w}^H\boldsymbol{u}(n)$ is the estimation error. As per eq. (1), the cost function $J(\boldsymbol{w})$ associated with this estimation corresponds to a mean-squared error. The problem is to determine the operating conditions for which $J(\boldsymbol{w})$ is minimum; expanding the term on the right leads to:

$$J(\boldsymbol{w}) = \sigma_d^2 - \boldsymbol{R}_{du}^H\boldsymbol{w} - \boldsymbol{w}^H\boldsymbol{R}_{du} + \ \boldsymbol{w}^H\boldsymbol{R}_u\boldsymbol{w}, \tag{2}$$

in which $J(\boldsymbol{w})$ is a scalar-valued quadratic function of $\boldsymbol{w}$. Now, using Wirtinger calculus to differentiate with respect to $\boldsymbol{w}^H$, while formally treating $\boldsymbol{w}$ as a constant, and equating to zero, we obtain the optimal Wiener solution $\boldsymbol{w}^o$ for the tap-weight vector, as given by the well-known Wiener-Hopf (normal) equation:

$$\boldsymbol{w}^o = \boldsymbol{R}_u^{-1}\boldsymbol{R}_{du}. \tag{3}$$

The solution $\boldsymbol{w}^o$ is the global minimum of the cost function with minimum mean-squares error (MMSE) $J(\boldsymbol{w}^o) = \sigma_d^2 - \boldsymbol{R}_{du}^H\boldsymbol{R}_u^{-1}\boldsymbol{R}_{du}$.

Next, to avoid the computationally challenging inversion of the matrix $\boldsymbol{R}_u$ required in eq. (3), as well as be able to track time variations of the input signals' statistics, we resort on the steepest descent method. Firstly, the cost function at successive iterations is enforced to be monotonically decreasing, i.e.:

$$J(\boldsymbol{w}(n+1)) < J(\boldsymbol{w}(n)), \tag{4}$$

where $\boldsymbol{w}(n)$ is a guess of $\boldsymbol{w}^o$ at iteration time n and $\boldsymbol{w}(n+1)$ is its update at $n+1$. The procedure for updating the estimated values is linear (affine) of the form:

$$\boldsymbol{w}(n+1) \ = \boldsymbol{w}(n) + \mu\boldsymbol{p}(n), \quad i \geq 0, \tag{5}$$

where the positive scalar μ is the step-size parameter that controls convergence and $\boldsymbol{p}(n)$ is an update direction vector. Now, from eq. (2), the gradient vector $\nabla_{\boldsymbol{w}^H}J(\boldsymbol{w}) = -\boldsymbol{R}_{du} + \boldsymbol{R}_u\boldsymbol{w}$ and, reinforcing eq. (4), it follows that the update direction $\boldsymbol{p}(n)$ must satisfy:

$$\boldsymbol{p}(n) = -\boldsymbol{B}\nabla_{\boldsymbol{w}^H}J(\boldsymbol{w}(n)), \tag{6}$$

where $\boldsymbol{B}$ is a positive-definite Hermitian matrix. This way, at each iteration, $\boldsymbol{p}(n)$ points to the steepest descent direction. For $\boldsymbol{B}$ equals to the identity matrix, eq. (5) results in:

$$\boldsymbol{w}(n+1) \ = \boldsymbol{w}(n) + \mu(\boldsymbol{R}_{du} - \boldsymbol{R}_u\boldsymbol{w}(n)), \quad n \geq 0, \tag{7}$$

with an initial guess $\boldsymbol{w}_0$. Note that, if, instead, $\boldsymbol{B}$ is chosen as the inverse of the Hessian matrix, the resulting update equation $\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \mu\boldsymbol{R}_u^{-1}(\boldsymbol{R}_{du} - \boldsymbol{R}_u\boldsymbol{w}(n))$ corresponds to the Newton's method.

So far, the steepest descent formulation relies explicitly on the knowledge of second-order moments of the inputs, $\boldsymbol{R}_{du}$ and $\boldsymbol{R}_u$, not available in many practical cases. Next, adaptive filter algorithms replace the input signal statistics by stochastic gradients, i.e., the following instantaneous approximations: $\hat{\boldsymbol{R}}_u = \boldsymbol{u}(n)\boldsymbol{u}^H(n)$ and $\hat{\boldsymbol{R}}_{du} = \boldsymbol{u}(n)d^*(n)$. This way, from eq. (7), the complex-LMS update recursion (Widrow-Hoff algorithm), results:

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \mu\hat{e}^*(n)\boldsymbol{u}(n), \tag{8}$$

where $\hat{\boldsymbol{w}}(n)$ is an estimate of the filter weights at iteration n, $\hat{e}(n) = d(n) - \hat{\boldsymbol{w}}^H(n)\boldsymbol{u}(n)$ and μ is a positive step-size. Note that the transversal tap-weights $\hat{\boldsymbol{w}}(n)$ in eq. (8) denotes an approximation of the $\boldsymbol{w}(n)$ in eq. (7), where deterministic gradients are used. In practice, the LMS filter suffers from sensitivity to time-varying scaling of the input signal, thus leading to the NLMS update equation:

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \frac{\mu}{\epsilon + \|\boldsymbol{u}(n)\|_2^2}\hat{e}^*(n)\boldsymbol{u}(n), \tag{9}$$

where ϵ is a regularization factor that prevents division by zero and $\|\cdot\|_2$ is the Euclidean norm. In summary, as in Figure 1, an adaptive filter applied to the task of unknown system identification comprises of two iterative procedures: (1) a filtering step, which involves computing at each instant n an estimated output $y(n) = \hat{d}(n)$ using the tap input $\boldsymbol{u}(n)$, as well as computing the corresponding estimation error $\hat{e}(n)$ between the estimated AF output and the reference signal (typically contaminated by measurement noise $\boldsymbol{v}(n)$) and (2) an update step, which involves the adaptive adjustment of the filter's tap-weights $\hat{\boldsymbol{w}}$, using a weight update equation.

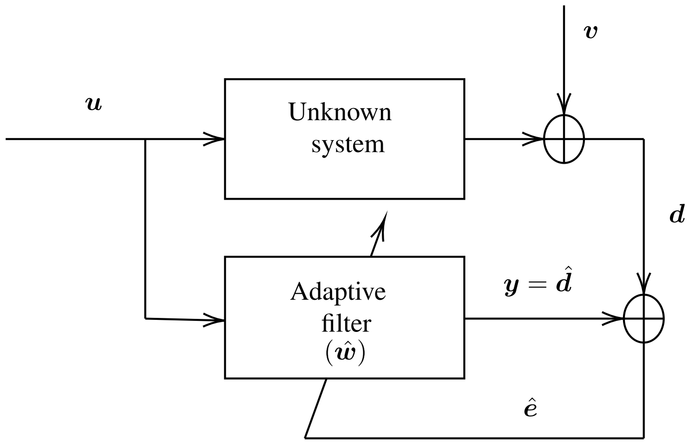

Fig. 1: System identification block diagram.

A. Sparse Adaptive Filters

Often, the unknown systems to be identified in telecommunications and acoustics are sparse, as in broadband wireless and digital TV applications, where the impulse responses to be estimated are typically subject to multipath and echo effects. In this case, the conventional AFs discussed so far are not able to exploit the prior knowledge on the sparsity of the system, that is, about the channel dispersion. By adding a sparsity-inducing penalty to the AF cost function, several algorithms for sparse channel estimation have been proposed in the literature, e.g., [6], [7], [8]. As discussed in this section, LMS-based AFs for sparse channels can be based on different sparsity-inducing penalties [12], among the most prominent are:

- The ℓ_1-norm regularization. Since the ℓ_0-pseudo-norm minimization is a well-known NP-hard problem, the first approach is the convex relaxation, replacing it by the ℓ_1-norm [11], given by $\|\boldsymbol{w}(n)\|_1 = \sum_{i=0}^{M-1} |w_i(n)|$. This penalty is incorporated into the LMS cost function, resulting in the zero-attracting LMS (ZA-LMS) method [13], as follows:

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \mu \hat{e}^*(n)\boldsymbol{u}(n) - \lambda sgn(\hat{\boldsymbol{w}}(n)), \quad (10)$$

  where λ is a regularization parameter and $sgn(\cdot)$ is the component-wise complex signum function. Note that $sgn(w_i(n)) = w_i(n)/|w_i(n)|$ when $w_i(n) \neq 0$ and $sgn(w_i(n)) = 0$ when $w_i(n) = 0$, for $i = 0, \ldots, M-1$, where M is the filter length. In eq. (10), the term $-\lambda sgn(\hat{\boldsymbol{w}}(n))$ attracts the tap coefficients towards zero.
- The ℓ_0-pseudo-norm approximation. The ℓ_0-pseudo-norm or counting norm (i.e., the number of non-zero elements in $\boldsymbol{w}(n)$) can be approximated in several ways, as proposed in the literature. For example, the following approximation function [14] can be used:

$$\|\boldsymbol{w}(n)\|_0 \approx \sum_{i=0}^{M-1} (1 - e^{-\beta |w_i(n)|}), \quad (11)$$

  where β is a positive parameter and M is the filter length, which leads to the update equation:

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \mu \hat{e}^*(n)\boldsymbol{u}(n) - \lambda\beta sgn(\hat{\boldsymbol{w}}(n)) e^{-\beta|\hat{\boldsymbol{w}}|}. \quad (12)$$

  Besides, other ℓ_0 approximations can be found in [15], [16].
- Heuristic approaches. We consider the re-weighted ZA-LMS (RZA-LMS) method [13] that attracts to zero only the less relevant tap-weights, as follows:

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \mu \hat{e}^*(n)\boldsymbol{u}(n) - \lambda \frac{sgn(\hat{\boldsymbol{w}}(n))}{1 + \delta |\hat{\boldsymbol{w}}(n)|}, \quad (13)$$

  where δ limits the coefficient magnitudes that are attracted to zero. Note that starting from the NLMS algorithm, one can obtain the ZA-NLMS and RZA-NLMS AFs, which are counterparts of the above algorithms. Another heuristic approach, the proportionate NLMS (P-NLMS) [17], [18] allows the adaptation step-size for each tap-weight to be calculated proportionally to the previous value of the tap-weight, thus allowing active coefficients to be adjusted faster than non-active ones.

III. Proposed Sparse AF with ℓ_0 Approximation

In this paper, we investigate a new sparse AF algorithm, derived from a smooth and continuous approximation of the ℓ_0-pseudo-norm through a weighted version of the ℓ_2-norm. Comparing to other ℓ_0-pseudo-norm approximations from the literature, the one we adopted inherits a leakage term from the leaky-LMS algorithm, also discussed in this section, that is known to improve the stability and robustness of the LMS algorithm. The approximation adopted in this paper is proposed in [9] as the SparseStep:

$$\|\boldsymbol{w}(n)\|_0 \approx \sum_{i=0}^{M-1} \frac{w_i^2(n)}{w_i^2(n) + \gamma}, \quad (14)$$

where $0 < \gamma \ll 1$. Note that for decreasing values of γ, the approximation of the counting norm becomes increasingly more accurate. Thus, the regularized cost function with the weighted ℓ_2-norm term becomes:

$$J(\boldsymbol{w}) = \frac{1}{2}|e(n)|^2 + \lambda \sum_{i=0}^{M-1} \frac{w_i^2(n)}{w_i^2(n) + \gamma}, \quad (15)$$

which allows efficient convex optimization techniques. In fact, this is a weighted version of the leaky-LMS cost function:

$$J(\boldsymbol{w}) = \frac{1}{2}|e(n)|^2 + \lambda \|\boldsymbol{w}(n)\|_2^2, \quad (16)$$

where the ℓ_2-norm term that is added prevents any excessive growth of the weight coefficients and λ is the corresponding positive leakage factor. From eq. (15), the proposed SparseStep LMS (SS-LMS) AF update equation is given by:

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \mu(\hat{e}^*(n)\boldsymbol{u}(n) - \lambda \boldsymbol{\Theta}(n)\hat{\boldsymbol{w}}(n)), \quad (17)$$

where $\boldsymbol{\Theta}(n)$ is a diagonal matrix, whose elements are the weights:

$$\theta_{i,i}(n) = \frac{\gamma^2}{(w_i^2(n) + \gamma^2)^2}. \quad (18)$$

Comparing to the leaky-LMS, the identity matrix is replaced by $\boldsymbol{\Theta}(n)$.
Additionally, we also apply a variable step-size (VSS) technique [19] to the proposed algorithm, expected to speed up convergence at the first iterations and improve the steady-state performance, leading to the SS-VSS-LMS AF, in which μ from eq. (17) is replaced by a variable step-size $\hat{\mu}(n) \in (0, \mu)$, to be computed at each iteration, as follows:

$$\hat{\mu}(n) = \mu \frac{\hat{\boldsymbol{p}}^H(n)\hat{\boldsymbol{p}}(n)}{\hat{\boldsymbol{p}}^H(n)\hat{\boldsymbol{p}}(n) + k}, \quad (19)$$

where k is a positive threshold parameter that relates to the estimated SNR by $k \propto 1/snr$ (where snr is in linear scale) and μ is the maximum step-size value. The vector $\hat{\boldsymbol{p}}(n)$ at iteration n is computed as:

$$\hat{\boldsymbol{p}}(n) = \beta \hat{\boldsymbol{p}}(n-1) + (1-\beta)\frac{\hat{e}^*(n)\boldsymbol{u}(n)}{\|\boldsymbol{u}(n)\|_2^2}, \quad (20)$$

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

Equivalently, combining eqs. (9) and (18), the SS-NLMS filter update equation is given by:

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \frac{\mu}{\epsilon + \|\boldsymbol{u}(n)\|_2^2}(\hat{e}^*(n)\boldsymbol{u}(n) - \lambda \boldsymbol{\Theta}(n)\hat{\boldsymbol{w}}(n)), \quad (21)$$

while its VSS version can be obtained by replacing μ with $\hat{\mu}(n)$ from eq. (19).

IV. Numerical Results

The objective of this section is to validate the proposed estimator through numerical simulations of channel estimation under several operating conditions, given by signal-to-noise ratios (SNR) and sparsity levels (Sp). In this study, Sp is the percentage of sparsity in the channel, as defined by:

$$\text{Sp} \triangleq (1 - K/L)100\%, \quad (22)$$

where K is the number of dominant coefficients, L is the channel length and the channel is said to be K-sparse.

The simulation experiments are run for all combinations of SNR in the range $\{5, 10, 15, 20, 25, 30\}$ (in dB) and Sp in the range $\{0, 20, 40, 60, 80\}$ (in %). We report the performance of the estimators using the mean square error (MSE) between the desired and the estimated outputs, the mean square deviation (MSD) between the optimal and the estimated channel coefficients, defined as $MSD = E\{\|\boldsymbol{w} - \hat{\boldsymbol{w}}(n)\|_2^2\}$, and the number of iterations until filter convergence. For each SNR and Sp combination, a grid of values of the filters hyper-parameters μ and λ is tested and the best values are selected, as follows. First, we run the conventional LMS and NLMS filters for each value in their μ ranges and find the best μ values, taking into account the achieved MSEs and required numbers of iterations until convergence. These μ values are then used for all the LMS- and NLMS-based filters under analysis. Next, we run the LMS- and NLMS-based sparse filters for each value in their λ ranges and determine for each filter the best λ value based on the MSE. Each simulation experiment consists of 100 independent Monte Carlo runs, each of which consisting of 1,000 iterations of the AFs, and the results are averaged. The simulations are presented for a channel length $L = 20$. For each simulation run, the positions of the K dominant taps in the sparse channel are randomly selected and the channel coefficients follow a real-valued Gaussian distribution with variance $\sigma_d^2 = 1/\sqrt{K}$. The additive noise is also Gaussian with $\sigma_n^2 = 10^{-SNR/10}$. In each case, the performance metrics MSD, MSE and numbers of iterations are reported for each AF, as follows.

Figure 2 shows the heat-maps of MSD performance obtained with the LMS, RZA-LMS, SS-LMS and SS-VSS-LMS filters for each combination of SNR and Sp. The simulations are run with LMS hyper-parameters $\mu = \{1\text{e}{-2}, 2\text{e}{-2}, 3\text{e}{-2}, 4\text{e}{-2}, 5\text{e}{-2}\}$ and $\lambda = \{1\text{e}{-4}, 2\text{e}{-4}, 3\text{e}{-4}, 5\text{e}{-4}, 1\text{e}{-3}, 2\text{e}{-3}, 3\text{e}{-3}, 5\text{e}{-3}\}$, the best values being selected as described above. Additional parameters are $\delta = 20$ for RZA-LMS, $\gamma = 0.015$ for SS-LMS, $\beta = 0.95$ and k ranging from $1\text{e}{-5}$ to $1\text{e}{-4}$ for the SS-VSS-LMS. Figure 3 shows the heat-maps of MSD performance obtained with the NLMS, RZA-NLMS, SS-NLMS and SS-VSS-NLMS filters for each combination of SNR and Sp values, run with the NLMS hyper-parameters $\mu = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and same λ range as for the LMS. Additional parameters are $\epsilon = 1\text{e}{-4}$ and $\delta = 20$ for RZA-NLMS and the other parameters as for the LMS. Figures 4 and 5 show the heat-maps of corresponding MSE performance, respectively, for the LMS- and NLMS-based filters for each combination of the SNR and Sp values, using the same parameter values above. Next, Figure 6 depicts the MSD ensemble learning curves of the LMS algorithms for a grid point at SNR=15dB and Sp=80%. Finally, Table I presents the numbers of iterations required until convergence at a few grid points.

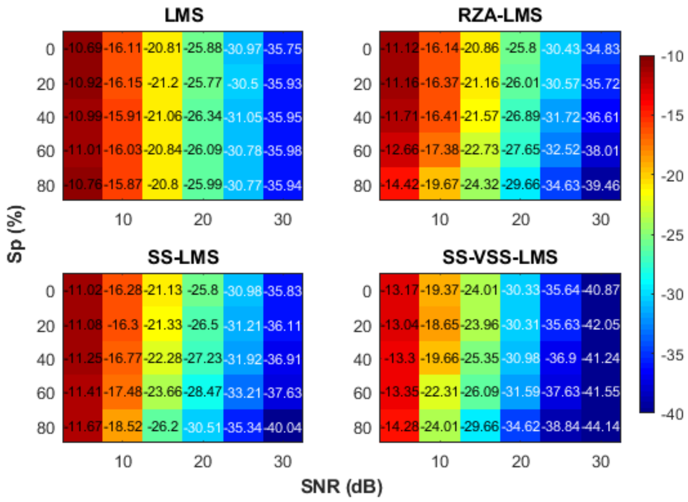

Fig. 2: Heat-maps of MSD (dB) for LMS-based sparse AFs.

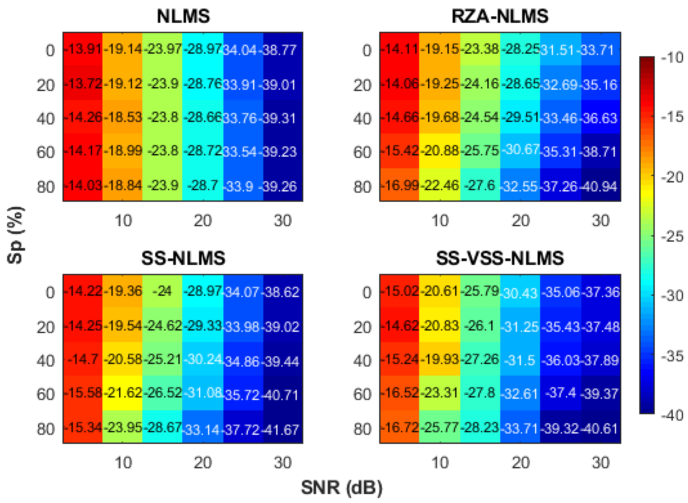

Fig. 3: Heat-maps of MSD (dB) for NLMS-based sparse AFs.

The following summary of observations can be drawn from the numerical simulations:

- All AFs simulated were highly sensitive to SNR.
- NLMS-based AFs performed better than the corresponding LMS ones.
- Sparse LMS and NLMS outperformed conventional LMS and NLMS as Sp increased, as observed in Figures 2 and 3 (MSD) and Figures 4 and 5 (MSE). Moreover, the proposed SS-LMS and SS-NLMS outperformed their ZA (not shown) and RZA counterparts, as well as the P-NLMS (not shown).
- The proposed SS-VSS-LMS and SS-VSS-NLMS outperformed the step-invariant AFs, as shown in the figures above and detailed by the ensemble learning curves in Figure 6.
- The numbers of iterations until convergence were equivalent for the step-invariant AFs and slightly larger for the VSS, as illustrated in Table I.

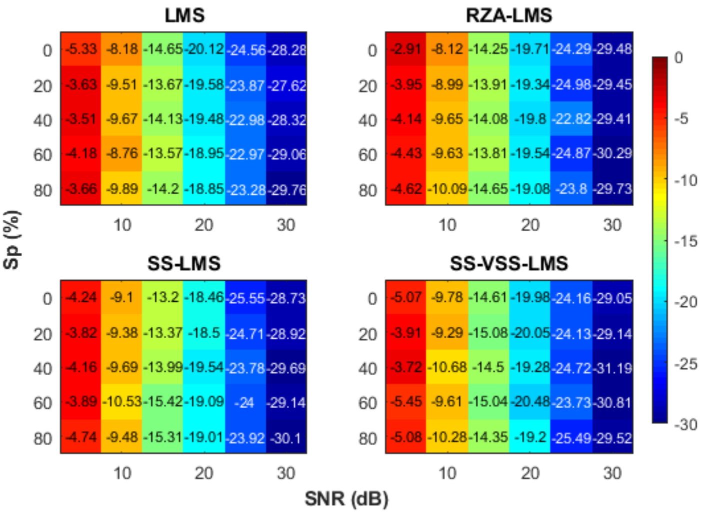

Fig. 4: Heat-maps of MSE (dB) for LMS-based sparse AFs.

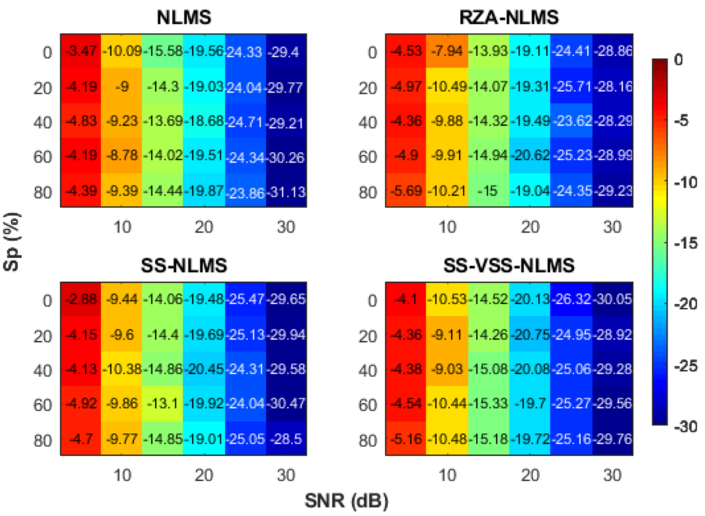

Fig. 5: Heat-maps of MSE (dB) for NLMS-based sparse AFs.

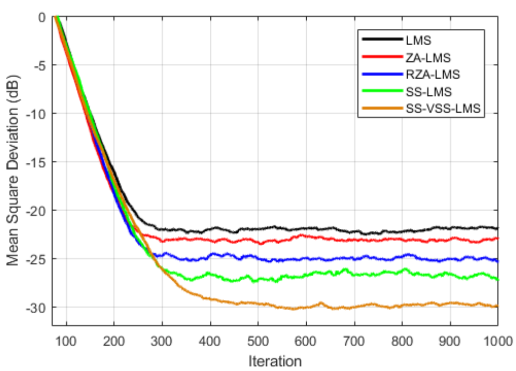

Fig. 6: MSD ensemble learning curves for LMS-based AFs (SNR=15dB, Sp=80% and $\mu = 2\text{e}{-2}$).

V. Conclusions

In this study, we proposed the SS-LMS and SS-NLMS AFs, derived from the SparseStep approximation for the ℓ_0-pseudo-norm penalty. We also extended these filters to accommodate variable step-size by defining the SS-VSS-LMS and SS-VSS-NLMS AFs. Finally, the proposed filters were evaluated through numerical simulations for various SNRs and sparsity levels, outperforming some well-known sparse AFs. A complexity analysis will be addressed in future work.

Table I: Numbers of iterations of LMS-based AFs.

SNR (dB)	Sp (%)	LMS	RZA-LMS	SS-LMS	SS-VSS-LMS
5	80	283	286	280	304
10	40	307	335	347	364
15	80	353	331	388	408
20	60	391	367	366	425

Acknowledgment

The author would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Finance Code 001, for partially funding this research.

References

[1] S. Haykin, *Adaptive Filter Theory*. Pearson, 2014.
[2] A. Sayed, *Adaptive Filters*. IEEE Press, Wiley, 2008.
[3] V. Nascimento and M. Silva, "Adaptive filters," in *Academic Press Library in Signal Processing, vol. 1*, ch. 12, pp. 619–761, Academic Press, 2014.
[4] P. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Jan. 2008.
[5] O. Tanrıkulu and K. Doğançay, "Selective-partial-update proportionate normalized least-mean-squares algorithm for network echo cancellation," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. II–1889–II–1892, 2002.
[6] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.
[7] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 255–271, 2010.
[8] G. Gui, Q. Wan, W. Peng, and F. Adachi, "Sparse multipath channel estimation using compressive sampling matching pursuit algorithm," May 2010.
[9] G. J. van den Burg, P. J. Groenen, and A. Alfons, "SparseStep: Approximating the counting norm for sparse regularization," *Econ. Inst. Res. Pap.*, no. 1, pp. 1–15, 2017.
[10] J. de Rooi and P. Eilers, "Deconvolution of pulse trains with the l_0 penalty," *Analytica Chimica Acta*, vol. 705, no. 1, pp. 218–226, 2011.
[11] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
[12] G. Gui and F. Adachi, "Improved adaptive sparse channel estimation using least mean square algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, pp. 1–18, Aug. 2013.
[13] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3125–3128, 2009.
[14] Y. Gu, J. Jin, and S. Mei, "ℓ_0-norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, 2009.
[15] Y. Chen, Y. Gu, and A. Hero, "Regularized Least-Mean-Square algorithms," *ArXiv e-prints*, Dec. 2010.
[16] G. Gui and F. Adachi, "Improved least mean square algorithm with application to adaptive sparse channel estimation," *EURASIP Journal on Wireless Communications and Networking*, Aug. 2013.
[17] D. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.
[18] F. Souza, O. Tobias, and R. Seara, "Considerações sobre o algoritmo PNLMS com fatores de ativação individuais," Jan. 2009.
[19] G. Gui, L. Xu, L. Shan, and F. Adachi, "Adaptive MIMO channel estimation using sparse variable step-size NLMS algorithms," in *2014 IEEE International Conference on Communication Systems*, pp. 605–609, 2014.