XXXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICACÕES E PROCESSAMENTO DE SINAIS - SBrT2019, 29 DE SETEMBRO A 02 DE OUTUBRO DE 2019, PETRÓPOLIS, RJ

The ℓ_0-Norm Constraint Coefficient Reusing Least Mean Square Algorithm

Leonardo C. Resende[1,2], Newton N. Siqueira[2,4] , Diego B. Haddad[3], Mariane R. Petraglia[2]

***Abstract*— An adaptive filtering algorithm should present fast convergence and good steady-state behaviour. Both metrics can be enhanced if one makes use of the sparsity (energy concentration in few coefficients) of the involved (unknown) transfer function. Unfortunately, such an approach can suffer from steady-state performance degradation when the signal-to-noise ratio is low, due to an increase of the adaptive weights variance. This work advances a new algorithm that combines the reusing coefficient strategy, which increases the robustness against the additive noise, with a sparsity-aware algorithm, which employs an approximation of the ℓ_0-norm in order to penalize non-sparse adaptive vector estimations.**

***Keywords*— Adaptive Filtering, Convergence Analysis, ℓ_0-Norm Constraint, Reusing Coefficients.**

I. Introduction

It is widely known that sparse signals and transfer functions are often present in applications such as channel equalization, system identification and acoustic echo cancellation [1]. This fact has motivated the insertion of such prior knowledge in the design of adaptive algorithms, in order to obtain improvement in the convergence rate and/or good steady-state performance. Traditional algorithms, such as the LMS (least mean square), NLMS (normalized LMS), AP (affine projection) and RLS (recursive least square), are intrinsically agnostic with respect to the energy concentration of the transfer function that the adaptive filter intends to emulate[1], so that sparsity-aware schemes may enhance their learning capabilities [2]. Most of these schemes can be classified into two categories: (i) proportionate algorithms, and (ii) regularization-based algorithms. The first of them comprises as examples the PNLMS [3], the PNLMS++ [4], the IPNLMS [5], the MPNLMS [6], and the IMPNLMS [7]. Among the regularization-based algorithms, one can list several proposals, such as the zero-attracting LMS (ZA-LMS) [1], the ℓ_0-LMS [8] and the non-uniform norm constraint LMS [9]. Since the sparsity-norm regularized filters are computationally less expensive and can achieve a better compromise between convergence and steady-state error than the proportionate approaches [10], this paper focuses on regularization-based algorithms, especially the ℓ_0-LMS, whose performance is analyzed in [11], [12].

Unfortunately, insertion of prior information on the sparsity of the ideal transfer function does not guarantee robustness against noise in a low signal-to-noise ratio (SNR) environment [13]. In this case, the high variance of the adaptive coefficients translates into poor steady-state performance [14]. This problem can be mitigated by the coefficient reuse (RC) strategy [15], which improves steady-state performance without significantly harming the convergence rate [16].

This paper proposes a deterministic optimization cost function, whose approximate solution introduces in the RC-LMS algorithm [17] a penalty for non-sparse coefficient vectors through an approximate ℓ_0-norm regularization. The resulting algorithm (named ℓ_0-RC-LMS) presents a low computational burden and combines the enhanced performance of sparsity aware based strategies with the robustness of reusing coefficient schemes in low SNR scenarios.

Section II succinctly describes the RC-NLMS algorithm and its derivation using the Lagrange multiplier technique. Section III presents the ℓ_0-LMS algorithm, which is usually derived using stochastic gradient. Section IV unifies the two paradigms, providing a generalized algorithm that incorporates both ℓ_0-norm regularization and the reuse of coefficients. Section V presents comparisons between the advanced approach and other algorithms. Finally, Section VI presents the concluding remarks of the paper.

II. RC-NLMS Algorithm

The algorithms that use the coefficient reuse technique perform well in steady-state operation, with a small reduction in the convergence rate which, in many scenarios, is negligible [15]. The RC strategy basically smoothes the dynamics of the adaptive weights, which is especially beneficial in environments with low SNR. The degree of coefficient reuse is directly influenced by an adjustable parameter $L \in \mathbb{N}$, with L greater than one (otherwise the RC-NLMS algorithm degenerates into the classic NLMS).

As depicted in Fig. 1, consider the vector[2] $\boldsymbol{w}(k) \in \mathbb{R}^N$ that contains the adaptive coefficients $w_i(k)$, for $i \in \{0, ..., N-1\}$, at the k-th iteration. Given a reference (or desired) signal $d(k)$, often corrupted by additive noise $\nu(k)$, the error $e(k)$ is a stochastic assessment of the quality of the current filter estimate, which is computed through

$$e(k) \triangleq d(k) - y(k), \tag{1}$$

where $y(k) \triangleq \boldsymbol{w}^T(k)\boldsymbol{x}(k)$ is the adaptive filter output at the k-th iteration and $\boldsymbol{x}(k) \in \mathbb{R}^N$ is the input vector defined as

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k) & x(k-1) & \dots & x(k-N+1) \end{bmatrix}^T, \tag{2}$$

with $x(k)$ being the k-th sample of the input signal. Using these definitions, the RC-NLMS algorithm can be derived from

[1]Instituto Federal do Rio de Janeiro. [2]Programa de Engenharia Elétrica (PEE), COPPE/UFRJ, Rio de Janeiro-RJ, Brasil. [3] Coordenação de Engenharia de Computação, CEFET-RJ *campus* Petrópolis, Petrópolis-RJ, Brasil. [4] Coordenação de Telecomunicações, CEFET-RJ *campus* Nova Iguaçu, Nova Iguaçu, Brasil. E-mails: leonardo.resende@ifrj.edu.br, newton.siqueira@cefet-rj.br , diego.haddad@cefet-rj.br, mariane@pads.ufrj.br.

[1]This paper focuses on the identification system task.

[2]All vectors of this paper are of column-type.

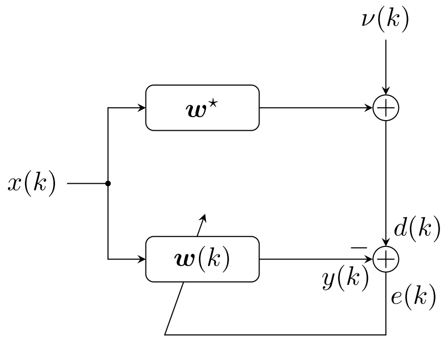

Fig. 1. Block diagram of an adaptive system identification task, where $\boldsymbol{w}^\star$ denotes the unknown system.

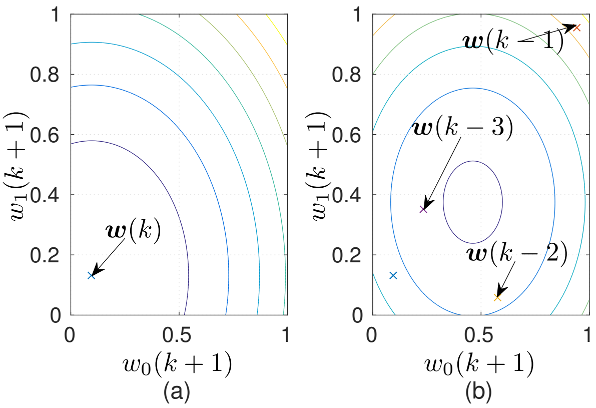

Fig. 2. Contour plots of the cost function $\mathcal{F}_{\text{RC}}[\boldsymbol{w}(k+1)]$ and coefficients vectors, assuming $\rho = 0.99$. (a) for $L = 1$ (i.e., the standard NLMS); (b) for $L = 4$.

the following optimization problem [15]:

$$\min_{\boldsymbol{w}(k+1)} \mathcal{F}_{\text{RC}}[\boldsymbol{w}(k+1)] = \sum_{l=0}^{L-1} \rho^l \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k-l)\|^2 \quad (3)$$
$$\text{s.t. } e_p(k) = (1-\beta)\overline{e}(k),$$

where β is the step size (or learning factor), $\overline{e}(k)$ is the filtered error, given by

$$\overline{e}(k) \triangleq d(k) - \theta(\rho) \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}^T(k-l)\boldsymbol{x}(k), \quad (4)$$

and

$$\theta(\rho) \triangleq \frac{\rho - 1}{\rho^L - 1}, \quad (5)$$

in which $\rho \in (0, 1]$ controls the weight given to past coefficient vectors [15]. Figure 2 illustrates the modifications engineered in the cost function by the coefficient reuse strategy for the bi-dimensional case (i.e., $N = 2$), where the corresponding contour plots and coefficient vectors are illustrated for $L = 1$ and $L = 4$.

The constrained and deterministic problem given in (3) can be solved by the Lagrange multiplier technique, which gives rise to the following *unconstrained* and equivalent[3] problem:

$$\min_{\boldsymbol{w}(k+1)} \mathcal{F}_{\text{RC}}[\boldsymbol{w}(k+1)] + \lambda \left[e_p(k) - (1-\beta)\overline{e}(k)\right], \quad (6)$$

whose solution generates the following filter updating equation:

$$\boldsymbol{w}(k+1) = \theta(\rho) \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}(k-l) + \frac{\beta \overline{e}(k)\boldsymbol{x}(k)}{||\boldsymbol{x}(k)||^2}. \quad (7)$$

It should be observed that the use of a small ρ value reduces the effect of previous $\boldsymbol{w}(k-l)$ vectors. Note that the solution $\boldsymbol{w}(k+1)$ minimizes the sum of weighted distances w.r.t. the last L vectors $\boldsymbol{w}(k-l)$, for $l \in \{0, 1, \ldots, L-1\}$, which smoothes adaptive coefficient oscillations. The computational effort involved in the coefficient reuse strategy may be reduced using the Set-Membership approach [18]. Furthermore, the trade-off between convergence rate and steady-state behavior can be relaxed through a time-variant coefficient reuse factor [16], [19].

III. ℓ_0-LMS ALGORITHM

Motivated by the Least Absolutely Shrinkage and Selection Operator (LASSO) [20] and by Compressive Sensing (CS) approaches [21], the ℓ_0-LMS algorithm [22] aims to accelerate the identification of a sparse system. This ability is obtained by minimizing the stochastic cost function

$$\mathcal{F}_{\ell_0-\text{LMS}}(k) \triangleq e^2(k) + \frac{\kappa}{\beta} F_\varrho\left[\boldsymbol{w}(k)\right], \quad (8)$$

in which the parameter $\kappa \in \mathbb{R}^+$ regularizes the amount of penalization of non-sparse solutions and $F_\varrho\left[\boldsymbol{w}(k)\right]$ is an almost everywhere differentiable function that approximates the ℓ_0-norm[4] and depends on an adjustable parameter $\varrho \in \mathbb{R}^+$ [11]. Such approximation is required due to the NP hardness of l_0-norm optimization [8]. One popular choice for $F_\varrho\left[\boldsymbol{w}(k)\right]$ is [11]

$$F_\varrho\left[\boldsymbol{w}(k)\right] \triangleq \sum_{n=0}^{N-1} \left(1 - e^{-\varrho|w_n(k)|}\right), \quad (9)$$

which is employed in this work. Alternative choices for this approximation function are given in [23]. Both ℓ_0-LMS and ℓ_0-NLMS updates[5] can be written as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \tilde{\beta}(k)\boldsymbol{x}(k)e(k) + \kappa \boldsymbol{f}_\varrho\left[\boldsymbol{w}(k)\right], \quad (10)$$

where $\boldsymbol{f}_\varrho\left[\cdot\right]$ is an approximation of the negative of the gradient of $F_\varrho\left[\boldsymbol{w}(k)\right]$ w.r.t. $\boldsymbol{w}(k)$ and $\tilde{\beta}(k)$ is the step-size, given by

$$\tilde{\beta}(k) = \begin{cases} \beta, & \text{for the } \ell_0\text{-LMS algorithm,} \\ \frac{\beta}{\boldsymbol{x}^T(k)\boldsymbol{x}(k)+\delta}, & \text{for the } \ell_0\text{-NLMS algorithm.} \end{cases} \quad (11)$$

This paper employs a low cost approximation of $\boldsymbol{f}_\varrho\left[\boldsymbol{w}(k)\right] = \nabla_{\boldsymbol{w}(k)} F_\varrho[\boldsymbol{w}(k)]$, so that the i-th coefficient of

[3]Equivalent in the sense that its solution is the same as that of the original constrained problem.
[4]Rigorously, the ℓ_0-norm is a pseudonorm.
[5]Other ℓ_p-norms could also be chosen [9].

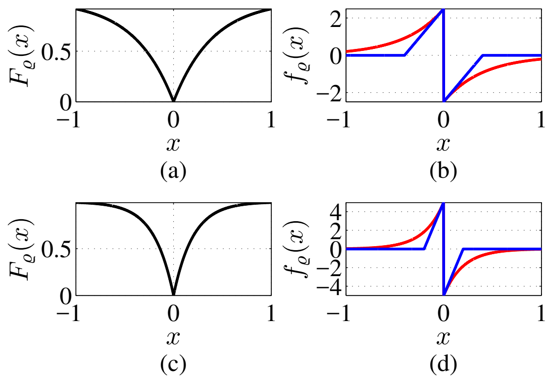

Fig. 3. Curves of $F_\varrho(x)$ and the low-cost approximation $f_\varrho(x)$ (in blue) of the exact function $-\frac{\partial F_\varrho(x)}{\partial x}$ (in red). (a) $F_\varrho(x)$ with $\varrho = 2.5$; (b) $f_\varrho(x)$ and its exact version with $\varrho = 2.5$; (c) $F_\varrho(x)$ with $\varrho = 5$; (d) $f_\varrho(x)$ and its exact version with $\varrho = 5$.

$\boldsymbol{f}_\varrho[\boldsymbol{w}(k)]$, denoted by $f_\varrho[w_i(k)]$, can be written as

$$f_\varrho[w_i(k)] \approx -\frac{\partial F_\varrho[\boldsymbol{w}(k)]}{\partial w_i(k)}, \tag{12}$$

which is related to a zero-attracting term [24]. The quantity $f_\varrho[w_i(k)]$ can be expressed as [22]

$$f_\varrho[w_i(k)] = \begin{cases} \varrho^2 w_i(k) + \varrho, & -\frac{1}{\varrho} \leq w_i(k) < 0 \\ \varrho^2 w_i(k) - \varrho, & 0 < w_i(k) \leq \frac{1}{\varrho} \\ 0, & \text{elsewhere} \end{cases}. \tag{13}$$

Figure 3 presents some examples of the univariable functions $F_\varrho(x)$ and $f_\varrho(x)$. Note that $f_\varrho(x)$ implements a *zero-point attraction* function, which in some contexts might produce large steady-state misalignment [12]. The ℓ_0-LMS can be modified in order to reduce its computational complexity and increase its robustness against impulsive noise (e.g., see [25]).

IV. Proposed Algorithm

The algorithm proposed in this work minimizes the cost function

$$\min_{\boldsymbol{w}(k+1)} \sum_{l=0}^{L-1} \rho^l \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k-l)\|^2 + \alpha F_\varrho[\boldsymbol{w}(k+1)], \tag{14}$$

where the term $\alpha F_\varrho[\boldsymbol{w}(k+1)]$ penalizes non-sparse solutions, subject to the linear constraint [17]

$$e_p(k) = (1 - \beta||\boldsymbol{x}(k)||^2)\bar{e}(k). \tag{15}$$

The solution of (14)-(15) provides a new non-normalized algorithm that is sparsity-aware and employs coefficient reuse. In the following, an approximated solution is derived.

Using the Lagrange multiplier technique for solving (14)-(15), the following unconstrained problem is obtained:

$$\min_{\boldsymbol{w}(k+1)} \sum_{l=0}^{L-1} \rho^l \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k-l)\|^2 + F_\varrho[\boldsymbol{w}(k+1)]$$
$$+\lambda[e_p(k) - (1 - \beta||\boldsymbol{x}(k)||^2)\bar{e}(k)] \triangleq G[\boldsymbol{w}(k+1)], \tag{16}$$

where λ is the Lagrange multiplier.

Zeroing the gradient $\nabla_{\boldsymbol{w}(k+1)} G[\boldsymbol{w}(k+1)]$ w.r.t. $\boldsymbol{w}(k+1)$ results in:

$$\nabla_{\boldsymbol{w}(k+1)} G[\boldsymbol{w}(k+1)] = 2\sum_{l=0}^{L-1} \rho^l \boldsymbol{w}(k+1) - 2\sum_{l=0}^{L-1} \rho^l \boldsymbol{w}(k-l)$$
$$+\frac{\alpha}{2}\boldsymbol{f}_\varrho[\boldsymbol{w}(k+1)] - \frac{\lambda}{2}\boldsymbol{x}(k) = \boldsymbol{0}, \tag{17}$$

where $\boldsymbol{f}_\varrho[\boldsymbol{w}(k+1)] \triangleq -\nabla F_\varrho[\boldsymbol{w}(k+1)]$ and $\mathcal{F}_\varrho[\boldsymbol{w}(k+1)] \approx \mathcal{F}_\varrho[\boldsymbol{w}(k)]$ [2].

Rewriting (17), we obtain

$$\boldsymbol{w}(k+1) = \theta(\rho) \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}(k-l) - \frac{\theta(\rho)}{2}\alpha \boldsymbol{f}_\varrho[\boldsymbol{w}(k)]$$
$$-\frac{\lambda}{2}\theta(\rho)\boldsymbol{x}(k) \tag{18}$$

Replacing (18) in (15), we get

$$\underbrace{d(k) - \theta(\rho)\sum_{l=0}^{L-1} \rho^l \boldsymbol{w}^T(k-l)\boldsymbol{x}(k)}_{=\bar{e}(k)} + \frac{\theta(\rho)}{2}\alpha \boldsymbol{f}_\varrho^T[\boldsymbol{w}(k)]\boldsymbol{x}(k)$$
$$+\frac{\lambda}{2}\theta(\rho)\|\boldsymbol{x}(k)\|^2 = \bar{e}(k) - \beta\|\boldsymbol{x}(k)\|^2\bar{e}(k), \tag{19}$$

and assuming that $\frac{\theta(\rho)}{2}\alpha \boldsymbol{f}_\varrho^T[\boldsymbol{w}(k)]\boldsymbol{x}(k)$ can be approximated by zero (see [26]), we conclude that

$$\Rightarrow \lambda = \frac{-2\beta\bar{e}(k)}{\theta(\rho)}. \tag{20}$$

Finally, by replacing (20) in (18), we obtain the filter update equation for the proposed ℓ_0-RC-LMS algorithm:

$$\boldsymbol{w}(k+1) = \theta(\rho) \sum_{l=0}^{L-1} \rho^l \boldsymbol{w}(k-l) + \frac{\theta(\rho)}{2}\alpha \boldsymbol{f}_\varrho[\boldsymbol{w}(k)]$$
$$+\beta\boldsymbol{x}(k)\bar{e}(k). \tag{21}$$

Note that the update equation (21) can be efficiently computed by defining the intermediate vector $\boldsymbol{\phi}(k)$ as

$$\boldsymbol{\phi}(k) \triangleq \theta(\rho)\boldsymbol{w}(k) + \theta(\rho)\rho\boldsymbol{w}(k-1) + \ldots + \theta(\rho)\rho^{L-1}\boldsymbol{w}(k-L+1), \tag{22}$$

so that (21) can be rewritten as

$$\boldsymbol{w}(k+1) \;=\; \boldsymbol{\phi}(k) + \beta\left[d(k) - \boldsymbol{\phi}^T(k)\boldsymbol{x}(k)\right]\boldsymbol{x}(k)$$
$$+\frac{\theta(\rho)}{2}\alpha \boldsymbol{f}_\varrho[\boldsymbol{w}(k)]. \tag{23}$$

Using the recursion

$$\boldsymbol{\phi}(k+1) = \rho\boldsymbol{\phi}(k) + \theta(\rho)\boldsymbol{w}(k+1) - \theta(\rho)\rho^L\boldsymbol{w}(k-L+1), \tag{24}$$

the updating of the intermediate vector $\boldsymbol{\phi}(k)$ requires $3N$ multiplications and $2N$ sums (assuming that both terms $\theta(\rho)$ and $\theta(\rho)\rho^L$ have been previously stored in memory). It should be noted that, employing (24), the complexity of the ℓ_0-RC-LMS algorithm does not increase with L. If D denotes the fraction of the N coefficients that are zero, each iteration of the ℓ_0-RC-LMS algorithm at steady-state requires approximately $(4+D)N$ sums and $(4+D)N+1$ multiplications, compared to the $2N$ sums and $2N+1$ multiplications involved in the original LMS algorithm. The pseudocode of the proposed ℓ_0-RC-LMS algorithm is presented in Algorithm 1, where N_{it} is the total number of iterations performed.

Algorithm 1 ℓ_0-RC-LMS Algorithm

procedure ℓ_0-RC-LMS(N,L,ρ,$x(k)$,N_{it},$d(k)$,ϱ,α)
  if $\rho = 1$ **then** ▷ Initialization
    $\theta(\rho) \leftarrow \frac{1}{L}$
  else
    $\theta(\rho) \leftarrow \frac{\rho-1}{\rho^L-1}$
  end if
  $\boldsymbol{w}(0) \leftarrow \mathbf{0}_{N\times1}, \ldots, \boldsymbol{w}(-L+1) \leftarrow \mathbf{0}_{N\times1}$
  $k \leftarrow 0$
  $\boldsymbol{\phi}(k) \leftarrow \mathbf{0}_{N\times1}$
  while $k \leq N_{\text{it}}$ **do** ▷ Main loop
    Evaluate $\overline{e}(k)$ using (4)
    $\boldsymbol{w}(k+1) \leftarrow \boldsymbol{\phi}(k) + \beta\left[d(k) - \boldsymbol{\phi}^T(k)\boldsymbol{x}(k)\right]\boldsymbol{x}(k)$
    $\boldsymbol{w}(k+1) \leftarrow \boldsymbol{w}(k+1) + \frac{\theta(\rho)}{2}\alpha\boldsymbol{f}_{\varrho}[\boldsymbol{w}(k)]$
    $k \leftarrow k+1$
    Evaluate $\boldsymbol{\phi}(k+1)$ using (24)
  end while
end procedure

The fast identification of sparse transfer functions, observed in the ℓ_0-LMS algorithm, and the robustness in low SNR environments of the reuse of coefficients are combined in the ℓ_0-RC-LMS algorithm. This is confirmed by the results of the simulations presented in the next section.

V. Results

The algorithms used for purposes of comparison with the proposed algorithm are LMS, ℓ_0-LMS and RC-LMS. The following parameters were employed: $L = 2$, $\rho = 0.9$, $\varrho = 2\cdot10^{-3}$, $\beta_{\text{LMS}} = 2\cdot10^{-2}$, $\beta_{\ell_0\text{LMS}} = 1.2\cdot10^{-2}$, $\beta_{\text{RCLMS}} = 1.5\cdot10^{-2}$,$\beta_{\ell_0\text{RCLMS}} = 1.5\cdot10^{-2}$, $\sigma_\nu^2 = 2\cdot10^{-2}$ and $\kappa = 10^{-3}$. A white Gaussian noise (WGN) with variance σ_ν^2 was added to the reference signal. All averaged results come from 400 independent Monte Carlo trials. The performance of the adaptive filter algorithm is analyzed through the mean square deviation (MSD), defined by

$$\text{MSD}(k) \triangleq \mathbb{E}\left\{\|\boldsymbol{w}^\star - \boldsymbol{w}(k)\|^2\right\}, \tag{25}$$

where $\mathbb{E}\{\cdot\}$ denotes the statistical expectation operator. The ideal transfer function coefficients are $w_n^\star = 1$, for $n \in \{0,1,2\}$, and $w_n^\star = 0$, for $n \notin \{0,1,2\}$ [27].

Figure 4 displays the MSD as a function of the number of iterations. Note that the proposed algorithm presents better steady-state performance than the other algorithms.

Figure 5 presents the steady-state MSD as a function of β. The parameters used are the same as those used in the simulations of Fig. 4. Notice that the proposed scheme has the best steady-state performance, independent of the β value.

In order to evaluate the tracking behavior of the proposed algorithm, Fig. 6 shows the evolution of the MSD in a case where there is an abrupt change of the unknown plant after 5000 iterations. The initial and last ideal transfer functions are given by

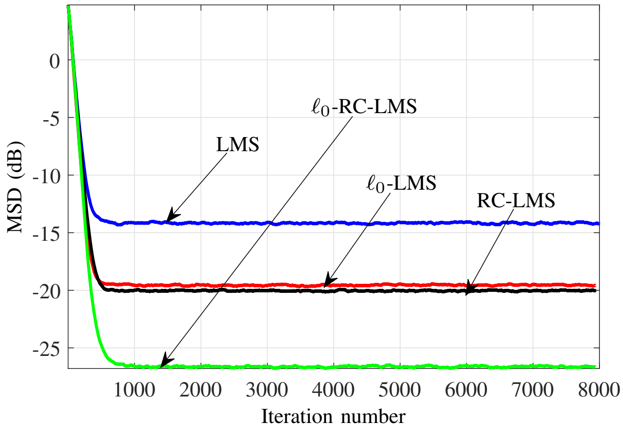

Fig. 4. MSD evolution (in dB) for the LMS, ℓ_0-LMS, RC-LMS and ℓ_0-RC-LMS algorithms.

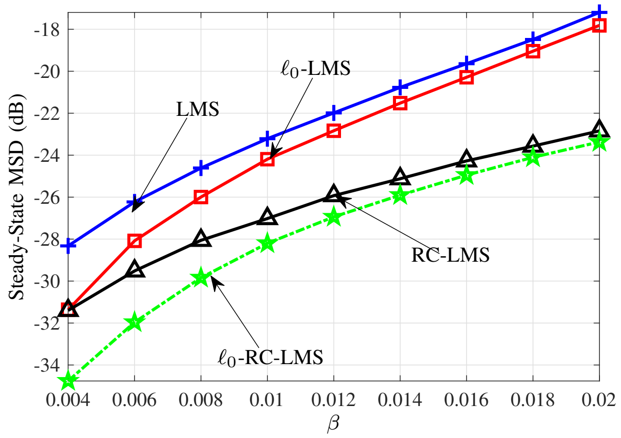

Fig. 5. Steady-state MSD (in dB) for the LMS, ℓ_0-LMS, RC-LMS and ℓ_0-RC-LMS algorithms.

$$\text{Initial} \;: w_n^\star = \begin{cases} 1, & \text{for } n \in \{0,1,2\} \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

$$\text{Last} \;: w_n^\star = \begin{cases} 1, & \text{for } n = 0 \\ -0.8, & \text{for } n = 1 \\ 0.3, & \text{for } n = 2 \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

From this figure, it can be observed that the proposed algorithm has good tracking ability.

Figure 7 presents the MSD behavior of the ℓ_0-RC-LMS algorithm as a function of the β parameter for different values of L. The other parameter values were $\rho = 0.9$ and $\sigma_\nu^2 = 10^{-4}$. The results were obtained from 100 independent Monte Carlo trials, with WGN input signal and ideal transfer function given in (26). Note that increasing the number of previous coefficient vectors L in the ℓ_0-RC-LMS produces better steady state performance.

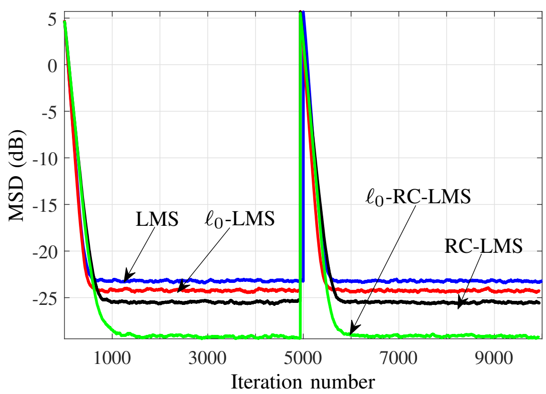

Fig. 6. MSD evolutions for the time-variant plant scenario for LMS, ℓ_0-LMS, RC-LMS and ℓ_0-RC-LMS algorithms.

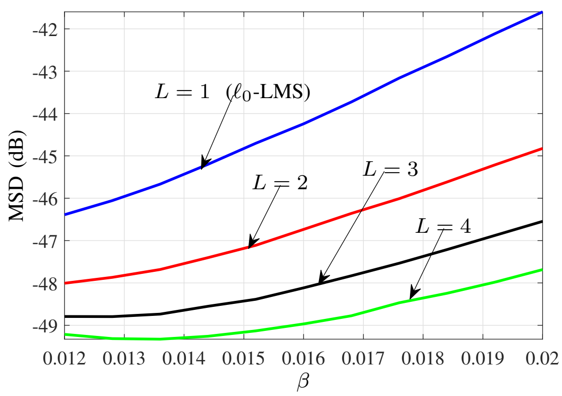

Fig. 7. Steady-state MSD of the proposed ℓ_0-RC-LMS algorithm w.r.t. β for $L \in \{1, 2, 3, 4\}$.

VI. Conclusions

In this paper, a new deterministic optimization problem is formulated, whose solution provides an adaptive algorithm that presents two desirable properties, namely: fast convergence when the unknown transfer function is sparse and robust performance to high energy additive noise. The proposed algorithm presents better steady-state behavior than traditional approaches, resulting in MSD improvements of more than 3 dB in the performed simulations without a noticeable loss in the convergence rate. In addition, the proposed algorithm presented experimentally good tracking capability.

References

[1] Yilun 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, April 2009.

[2] M. V. Lima, T. N. Ferreira, W. A. Martins, and P. S. Diniz, "Sparsity-aware data-selective adaptive filters," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4557–4572, 2014.

[3] D. L. 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.

[4] S. L. Gay, "An efficient fast converging adaptive filter for network echo cancellation," in *Signals, Systems & Computers, 1998. Conference Record of the Thirty-Second Asilomar Conference on*, vol. 1, pp. 394–398, IEEE, 1998.

[5] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," *Orlando, FL, USA*, pp. 1881–1884, 2002.

[6] H. Deng and M. Doroslovacki, "Proportionate adaptive algorithms for network echo cancellation," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1794–1803, 2006.

[7] L. Liu, M. Fukumoto, and S. Saiki, "An improved mu-law proportionate NLMS algorithm.," in *ICASSP*, pp. 3797–3800, Citeseer, 2008.

[8] Y. Gu, J. Jin, and S. Mei, "ℓ_0-norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, pp. 774–777, Sep. 2009.

[9] F. Y. Wu and F. Tong, "Non-uniform norm constraint LMS algorithm for sparse system identification," *IEEE Communications Letters*, vol. 17, pp. 385–388, February 2013.

[10] B. K. Das, L. A. Azpicueta-Ruiz, M. Chakraborty, and J. Arenas-García, "A comparative study of two popular families of sparsity-aware adaptive filters," in *2014 4th International Workshop on Cognitive Information Processing (CIP)*, pp. 1–6, May 2014.

[11] K. da S. Olinto, D. B. Haddad, and M. R. Petraglia, "Transient analysis of ℓ_0-LMS and ℓ_0-NLMS algorithms," *Signal Processing*, vol. 127, pp. 217–226, 2016.

[12] G. Su, J. Jin, Y. Gu, and J. Wang, "Performance analysis of ℓ_0-norm constraint least mean square algorithm," *IEEE Transactions on Signal Processing*, vol. 60, pp. 2223–2235, May 2012.

[13] R. M. S. Pimenta, L. C. Resende, N. N. Siqueira, D. B. Haddad, and M. R. Petraglia, "A new proportionate adaptive filtering algorithm with coefficient reuse and robustness against impulsive noise," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 465–469, Sep. 2018.

[14] K. Nishikawa, "Low variance adaptation method for the LMS-type adaptive filters based on the gaussian model," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, May 2019.

[15] H. Cho, C. W. Lee, and S. W. Kim, "Derivation of a new normalized least mean squares algorithm with modified minimization criterion," *Signal Processing*, vol. 89, no. 4, pp. 692 – 695, 2009.

[16] S.-E. Kim, J.-W. Lee, and W.-J. Song, "Steady-state analysis of the NLMS algorithm with reusing coefficient vector and a method for improving its performance," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 4120–4123, IEEE, 2011.

[17] L. C. Resende, N. N. Siqueira, R. M. S. Pimenta, D. B. Haddad, and M. R. Petraglia, "LMS algorithm with reuse of coefficients and robustness against impulsive noise," in *XXII Congresso Brasileiro de Automática*, pp. 1–5, Sept. 2018.

[18] R. M. S. Pimenta, L. C. Resende, D. B. Haddad, M. R. Petraglia, and N. N. Siqueira, "Novo algoritmo proporcional com reúso de coeficientes e seleção de dados," in *Anais do XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, pp. 1–5, Sept. 2017.

[19] L. C. Resende, D. B. Haddad, and M. R. Petraglia, "A variable step-size NLMS algorithm with adaptive coefficient vector reusing," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 0181–0186, May 2018.

[20] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, SERIES B*, vol. 58, pp. 267–288, 1994.

[21] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 24, pp. 118–121, July 2007.

[22] 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.

[23] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Transactions on Medical Imaging*, vol. 28, pp. 106–121, Jan 2009.

[24] Y. Liu, C. Li, and Z. Zhang, "Diffusion sparse least-mean squares over networks," *IEEE Transactions on Signal Processing*, vol. 60, pp. 4480–4485, Aug 2012.

[25] L. O. Oliveira, D. B. Haddad, and M. R. Petraglia, "Análise de Transiente do Algoritmo ℓ_0-sign-LMS," in *Anais do XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, pp. 1–5, Sept. 2017.

[26] D. B. Haddad, M. R. Petraglia, and A. Petraglia, "A unified approach for sparsity-aware and maximum correntropy adaptive filters," in *2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 170–174, Aug 2016.

[27] L. C. Resende, D. B. Haddad, G. d. R. Ferreira, P. H. Campelo, and M. R. Petraglia, "Sparsity-aware reuse of coefficients normalised least mean squares," *Electronics Letters*, vol. 55, no. 9, pp. 561–563, 2019.