

ADAPTIVE PARAMETERS ADJUSTMENT FOR GROUP REWEIGHTED ZERO-ATTRACTING LMS

Danqi Jin^{*}, Jie Chen^{*}, Cédric Richard[†], Jingdong Chen^{*}

^{*} Center of Intelligent Acoustics and Immersive Communications

School of Marine Science and Technology, Northwestern Polytechnical University, China

[†] Laboratoire Lagrange, Université Côte d'Azur, CNRS, OCA, France

danqijin@mail.nwpu.edu.cn dr.jie.chen@ieee.org cedric.richard@unice.fr jingdongchen@ieee.org

ABSTRACT

Group zero-attracting LMS and its reweighted form have been proposed for addressing system identification problems with structural group sparsity in the parameters to estimate. Both algorithms however suffer from a trade-off between sparsity degree and estimation bias and, in addition, between convergence speed and steady-state performance like most adaptive filtering algorithms. It is therefore necessary to properly set their step size and regularization parameter. Based on a model of their transient behavior, we introduce a variable-parameter variant of both algorithms to address this issue. By minimizing their mean-square deviation at each time instant, we obtain closed-form expressions of the optimal step size and regularization parameter. Simulation results illustrate the effectiveness of the proposed algorithms.

Index Terms— Sparse system identification, group sparsity, zero-attracting LMS, adaptive parameter adjustment

1. INTRODUCTION

Adaptive filtering algorithms are powerful tools for online system identification [1, 2]. Among the set of existing solutions, the least-mean-squares (LMS) algorithm plays a central role due to its robustness, good performance and low complexity. A number of variants have been proposed in the literature to endow the LMS with useful additional characteristics such as the ability to promote sparse estimates. This property is required in applications such as channel identification where, though the impulse responses can be long, only a few coefficients may have significant values. Several algorithms were derived, such as the proportionate NLMS (PNLMS) [3, 4], the zero-attracting LMS algorithm (ZA-LMS) and its reweighted form (RZA-LMS) [5]. They offer improved performance in sparse scenarios provided that their parameters are appropriately set.

Beyond element-wise sparsity, a further consideration is that some sparse systems may be group-sparse [6, 7]. Compared to general sparse systems whose impulse response is not necessarily structured, group-sparse systems have an impulse response composed of a few distinct clusters of nonzero coefficients. Applications are abundant, e.g., specular multipath acoustic and wireless channels estimation [6–8]. Using such structural prior information should lead to enhanced performance. Based on mixed norm regularization, the $\ell_{1,\infty}$ -regularized RLS [6], the group ZA-LMS (GZA-LMS) and the

group reweighted ZA-LMS (GRZA-LMS) [7] were proposed to promote group-sparsity. Nevertheless, setting the algorithm parameters such as the step size and the regularization parameter to ensure a performance gain remains a tricky task [6, 7, 9]. On the one hand, the step size plays a critical role in the trade-off between the convergence speed and the steady-state performance. On the other hand, the regularization parameter controls the trade-off between the degree of sparsity and the estimation bias. It is worth noting that setting one of these parameters to an inappropriate value may deteriorate the estimation performance.

Variable parameter strategies usually provide an efficient way to achieve a reasonable trade-off between competing performance requirements [10]. Several variable step size strategies have been proposed for the LMS and ZA-LMS algorithms [10–14], mostly based on estimation error. There are however few works addressing this issue for group-sparse LMS. Motivated by our recent work [15], we propose in this paper to derive the so-called variable-parameter GZA-LMS (VP-GZA-LMS) and GRZA-LMS (VP-GRZA-LMS) algorithms. The step size and the regularization parameter are both adjusted in an online way, based on an optimization procedure that minimizes the mean-square deviation (MSD) at each iteration. Experiments illustrate the effectiveness of this strategy, which leads to a faster convergence rate as well as a lower misadjustment error.

Notation. Normal font x denotes scalars. Boldface lowercase letters \mathbf{x} and uppercase letters \mathbf{X} denote column vectors and matrices, respectively. The superscript $(\cdot)^\top$ and $(\cdot)^{-1}$ denote the transpose and inverse operators, respectively. $\mathbf{0}_N$ and $\mathbf{1}_N$ denote all-zero vector and all-one vector of length N . The operator $\text{tr}\{\cdot\}$ takes the trace of its matrix argument. The mathematical expectation is denoted by $\mathbb{E}\{\cdot\}$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. The operator $\max\{\cdot, \cdot\}$ and $\min\{\cdot, \cdot\}$ take the maximum or minimum of their arguments.

2. SYSTEM MODEL AND GROUP-SPARSE LMS

Consider an unknown system defined by the linear model:

$$d_n = \mathbf{u}_n^\top \mathbf{w}^* + z_n \quad (1)$$

at time instant n , where $\mathbf{w}^* \in \mathbb{R}^L$ is an unknown parameter vector, $\mathbf{u}_n \in \mathbb{R}^L$ is a zero-mean regression vector with positive definite covariance matrix, and d_n is the output signal assumed to be zero-mean. The error signal z_n is assumed to be stationary, independent and identically distributed (i.i.d.), with zero-mean and variance σ_z^2 , and independent of any other signal.

The work of Jie Chen was supported in part by NSFC grant 61671382. The work of Cédric Richard was supported in part by ANR and in part by DGA under Grant ANR-13-ASTR-0030 (ODISSEE Project). The work of Jingdong Chen was supported in part by NSFC grant 61425005.

Consider the mean-square error (MSE) cost $J(\mathbf{w})$, namely,

$$J(\mathbf{w}) = \frac{1}{2} \mathbb{E}\{[d_n - \mathbf{w}^\top \mathbf{u}_n]^2\} \quad (2)$$

It can be checked that \mathbf{w}^* is the minimizer of $J(\mathbf{w})$. In this paper we consider the problem of estimating the unknown parameter vector \mathbf{w}^* when it is group-sparse. This problem can be addressed by minimizing the following regularized MSE cost:

$$\begin{aligned} \mathbf{w}_{\text{GZA}}^\circ &= \arg \min_{\mathbf{w}} J_{\text{GZA}}(\mathbf{w}) \\ \text{with } J_{\text{GZA}}(\mathbf{w}) &= \frac{1}{2} \mathbb{E}\{[d_n - \mathbf{w}^\top \mathbf{u}_n]^2\} + \lambda \|\mathbf{w}\|_{1,2} \end{aligned} \quad (3)$$

where $\lambda \geq 0$ is the regularization parameter. The $\ell_{1,2}$ -norm of \mathbf{w} , defined below, allows to promote its group-sparsity:

$$\|\mathbf{w}\|_{1,2} = \sum_{j=1}^J \|\mathbf{w}_{\mathcal{G}_j}\|_2 \quad (4)$$

where $\{\mathcal{G}_j\}_{j=1}^J$ is a partition of the index set $\mathcal{G} = \{1, \dots, L\}$, and $\mathbf{w}_{\mathcal{G}_j}$ denotes the subvector of \mathbf{w} with entries indexed by \mathcal{G}_j . Calculating a subgradient of $J_{\text{GZA}}(\mathbf{w})$, then approximating second-order moments by instantaneous estimates, leads to the following subgradient update equation in subvector form:

$$\mathbf{w}_{n+1, \mathcal{G}_j} = \mathbf{w}_{n, \mathcal{G}_j} + \mu e_n \mathbf{u}_{n, \mathcal{G}_j} - \rho \mathbf{s}_{n, \mathcal{G}_j} \quad (5)$$

for $j \in \{1, \dots, J\}$, with:

$$\mathbf{s}_{n, \mathcal{G}_j} = \begin{cases} \frac{\mathbf{w}_{n, \mathcal{G}_j}}{\|\mathbf{w}_{n, \mathcal{G}_j}\|_2} & \text{for } \|\mathbf{w}_{n, \mathcal{G}_j}\|_2 \neq 0 \\ 0 & \text{for } \|\mathbf{w}_{n, \mathcal{G}_j}\|_2 = 0, \end{cases} \quad (6)$$

with $e_n = d_n - \mathbf{w}_n^\top \mathbf{u}_n$ the estimation error, $\mathbf{u}_{n, \mathcal{G}_j}$ the subvector of \mathbf{u}_n with entries indexed by \mathcal{G}_j , μ a positive step size, and $\rho = \mu\lambda$ the shrinkage parameter.

The GRZA-LMS was proposed to reinforce group-sparsity and then get enhanced performance in group-sparse system identification. Consider the optimization problem:

$$\begin{aligned} \mathbf{w}_{\text{GRZA}}^\circ &= \arg \min_{\mathbf{w}} J_{\text{GRZA}}(\mathbf{w}) \\ \text{with } J_{\text{GRZA}}(\mathbf{w}) &= \frac{1}{2} \mathbb{E}\{[d_n - \mathbf{w}^\top \mathbf{u}_n]^2\} \\ &\quad + \lambda \sum_{j=1}^J \log \left[1 + \frac{\|\mathbf{w}_{\mathcal{G}_j}\|_2}{\varepsilon} \right] \end{aligned} \quad (7)$$

where the log-sum penalty term is used to make group-sparsity attractor takes effort only for groups at the same level as ε [5]. Similarly, using a stochastic subgradient update yields the GRZA-LMS:

$$\mathbf{w}_{n+1, \mathcal{G}_j} = \mathbf{w}_{n, \mathcal{G}_j} + \mu e_n \mathbf{u}_{n, \mathcal{G}_j} - \rho \beta_{n, j} \mathbf{s}_{n, \mathcal{G}_j}, \quad (8)$$

where $\beta_{n, j} = 1/[\|\mathbf{w}_{n, \mathcal{G}_j}\|_2 + \varepsilon]$ is a weighting coefficient. Equivalently, equation (8) in vector form is given by:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu e_n \mathbf{u}_n - \rho \beta_n \circ \mathbf{s}_n, \quad (9)$$

where β_n and \mathbf{s}_n are vector forms of $\beta_{n, j}$ and $\mathbf{s}_{n, \mathcal{G}_j}$, respectively, with dimension $L \times 1$. Symbol \circ denotes the Hadamard product.

Observe from equations (5) and (8) that GRZA-LMS reduces to GZA-LMS by replacing the parameters $\beta_{n, j}$ with 1, that is,

$$\beta_{n, j} = \begin{cases} \frac{1}{\|\mathbf{w}_{n, \mathcal{G}_j}\|_2 + \varepsilon} & \text{GRZA-LMS} \\ 1 & \text{GZA-LMS.} \end{cases} \quad (10)$$

We shall derive a variable parameter strategy for both GZA-LMS and GRZA-LMS algorithms based on the general form (9), while specific algorithms can be obtained by setting $\beta_{n, j}$ according to (10).

3. MODEL-BASED PARAMETER DESIGN OF GRZA-LMS

3.1. Transient Behavior Model of GRZA-LMS

Define the weight error vector $\tilde{\mathbf{w}}_n$ as the difference between the estimated weight vector \mathbf{w}_n and \mathbf{w}^* :

$$\tilde{\mathbf{w}}_n = \mathbf{w}_n - \mathbf{w}^* \quad (11)$$

To derive our variable parameter strategy, we analyze the transient behavior of the mean-square deviation (MSD) of \mathbf{w}_n over time, defined as: $\xi_n = \mathbb{E}\{\|\tilde{\mathbf{w}}_n\|^2\}$. To keep the calculations mathematically tractable, we introduce the independence assumption [1]:

A1: The weight-error vector $\tilde{\mathbf{w}}_n$ is statistically independent of the input vector \mathbf{u}_n .

This assumption is commonly used in the adaptive filtering literature since it helps simplify the analysis, and the performance results obtained under this assumption match well the actual performance of filters for sufficiently small step sizes [1].

Subtracting \mathbf{w}^* from both sides of (9), using $e_n = z_n - \tilde{\mathbf{w}}_n^\top \mathbf{u}_n$, leads to the update equation of $\tilde{\mathbf{w}}_n$:

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n + \mu \mathbf{u}_n z_n - \mu \mathbf{u}_n \mathbf{u}_n^\top \tilde{\mathbf{w}}_n - \rho \beta_n \circ \mathbf{s}_n. \quad (12)$$

Using A1 and $e_n = z_n - \tilde{\mathbf{w}}_n^\top \mathbf{u}_n$, the MSE of the GRZA-LMS is given by:

$$\mathbb{E}\{e_n^2\} = \sigma_z^2 + \text{tr}\{\mathbf{R}_u \mathbf{Q}_n\} \quad (13)$$

with $\mathbf{Q}_n = \mathbb{E}\{\tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^\top\}$. The quantity $\text{tr}\{\mathbf{R}_u \mathbf{Q}_n\}$ is the excess mean-square error (EMSE) at time instant n , denoted by ζ_n . Note that $\xi_n = \text{tr}\{\mathbf{Q}_n\}$. Again, to keep the calculations mathematically tractable, we introduce the following assumptions [1]:

A2: The input regressor \mathbf{u}_n is a zero-mean white signal with covariance matrix $\mathbf{R}_u = \sigma_u^2 \mathbf{I}$.

A2': The input regressor \mathbf{u}_n is Gaussian distributed.

Though introducing A2 and A2' simplify the derivation, as illustrated with simulation results, it turns out that the resulting algorithms work well in non-Gaussian correlated input scenarios where these assumptions do not hold. As shown in the sequel, A2' is only used with (16) to make the calculation of g tractable. Under A2, we can relate the MSD to the EMSE via a scaling factor:

$$\zeta_n = \sigma_u^2 \text{tr}\{\mathbf{Q}_n\} = \sigma_u^2 \xi_n. \quad (14)$$

We shall now determine a recursion for $\text{tr}\{\mathbf{Q}_n\}$ in order to relate the MSD at two consecutive time instants n and $n+1$. Post-multiplying (12) by its transpose, taking the expectation and matrix trace, using A1 and A2, we get:

$$\text{tr}\{\mathbf{Q}_{n+1}\} = \text{tr}\{\mathbf{Q}_n\} + \mu^2 g + \rho^2 h + 2\mu\rho\ell - 2\mu r_1 - 2\rho r_2 \quad (15)$$

with

$$g = \sigma_z^2 \text{tr}\{\mathbf{R}_u\} + \mathbb{E}\{\mathbf{u}_n^\top \tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^\top \mathbf{u}_n \mathbf{u}_n^\top \mathbf{u}_n\} \quad (16)$$

$$h = \mathbb{E}\{(\beta_n \circ \mathbf{s}_n)^\top (\beta_n \circ \mathbf{s}_n)\} \quad (17)$$

$$\ell = \mathbb{E}\{\tilde{\mathbf{w}}_n^\top \mathbf{u}_n \mathbf{u}_n^\top (\beta_n \circ \mathbf{s}_n)\} \quad (18)$$

$$r_1 = \mathbb{E}\{\tilde{\mathbf{w}}_n^\top \mathbf{u}_n \mathbf{u}_n^\top \tilde{\mathbf{w}}_n\} \quad (19)$$

$$r_2 = \mathbb{E}\{(\beta_n \circ \mathbf{s}_n)^\top \tilde{\mathbf{w}}_n\}. \quad (20)$$

We have dropped the time index n in the left-hand side of (16)–(20) for compactness.

3.2. Parameter Design Using Transient Behavior Model

We shall now derive our parameter design strategy for GRZA-LMS using model (15). Given the MSD ξ_n at time instant n , we need to determine the parameters $\{\mu_n, \rho_n\}$ that minimize the MSD ξ_{n+1} :

$$\{\mu_n^*, \rho_n^*\} = \arg \min_{\mu, \rho} \xi_{n+1} | \xi_n. \quad (21)$$

Using recursion (15), the above optimization problem becomes:

$$\begin{aligned} \{\mu_n^*, \rho_n^*\} &= \arg \min_{\mu, \rho} \text{tr}\{\mathbf{Q}_{n+1}\} \\ &= \arg \min_{\mu, \rho} \text{tr}\{\mathbf{Q}_n\} + \mu^2 g + \rho^2 h + 2\mu\rho\ell - 2\mu r_1 - 2\rho r_2. \end{aligned} \quad (22)$$

Equivalently, equation (22) can be written in matrix form as:

$$\xi_{n+1} = [\mu \ \rho] \mathbf{H} [\mu \ \rho]^\top - 2[r_1 \ r_2] [\mu \ \rho]^\top + \xi_n, \quad (23)$$

which is a quadratic function of $[\mu \ \rho]$, with $\mathbf{H} = \begin{bmatrix} g & \ell \\ \ell & h \end{bmatrix}$.

By decomposing g with respect to the two additive terms in the right-hand side of (16), one can show that the Hessian matrix \mathbf{H} can be written as the sum of a covariance matrix and a positive semidefinite matrix. Matrix \mathbf{H} is thus positive semidefinite. In practice, since a covariance matrix is almost always positive definite [16], we shall assume that \mathbf{H} is positive definite, which allows us to obtain the optimal parameters via:

$$[\mu_n^* \ \rho_n^*]^\top = \mathbf{H}^{-1} [r_1 \ r_2]^\top. \quad (24)$$

Some elementary algebra leads to:

$$\mu_n^* = \frac{hr_1 - \ell r_2}{gh - \ell^2} \quad (25)$$

$$\rho_n^* = \frac{gr_2 - \ell r_1}{gh - \ell^2}. \quad (26)$$

This result cannot be used in practice since it requires statistics that are not available in online learning scenarios. We shall now introduce an approximation for these quantities. Time index n is added in variables g_n, h_n, ℓ_n, r_{1n} and r_{2n} for clearance.

Consider first the quantity g_n . With A1 and A2-A2', we obtain:

$$\begin{aligned} g_n &= \sigma_z^2 \text{tr}\{\sigma_u^2 \mathbf{I}\} + \text{tr}\{2\mathbf{R}_u \mathbf{Q}_n \mathbf{R}_u + \text{tr}\{\mathbf{R}_u \mathbf{Q}_n\} \mathbf{R}_u\} \\ &= \sigma_z^2 \sigma_u^2 L + (2 + L) \sigma_u^2 \zeta_n. \end{aligned} \quad (27)$$

Next, using A1 with $r_{1,n}$ yields:

$$r_{1,n} = \zeta_n. \quad (28)$$

Then, approximating the expectations in (17), (18) and (20) by their instantaneous argument yields:

$$h_n \approx (\beta_n \circ \mathbf{s}_n)^\top (\beta_n \circ \mathbf{s}_n) \quad (29)$$

$$\ell_n \approx \tilde{\mathbf{w}}_n^\top \mathbf{u}_n \mathbf{u}_n^\top (\beta_n \circ \mathbf{s}_n) \quad (30)$$

$$r_{2,n} \approx (\beta_n \circ \mathbf{s}_n)^\top \tilde{\mathbf{w}}_n. \quad (31)$$

Now we construct an approximation for \mathbf{w}^* at time instant n in order to evaluating the weight error vector $\tilde{\mathbf{w}}_n$. As proposed in [17], one strategy is to use a one-step approximation of the form:

$$\hat{\mathbf{w}}_n^* = \mathbf{w}_n - \eta_n \nabla J(\mathbf{w}_n) \quad (32)$$

with η_n a positive step size to be determined. Given ξ_n , we seek η_n that minimizes ξ_{n+1} . Following the same reasoning as (21)–(26) leads to $\eta_n = r_{1,n}/g_n$. Then, we approximate the gradient $\nabla J(\mathbf{w}_n)$ with the instantaneous value $-e_n \mathbf{u}_n$. Finally, we obtain the one-step approximation $\hat{\mathbf{w}}_n^* = \mathbf{w}_n - \mathbf{p}_n$ with $\mathbf{p}_n = -\frac{r_{1,n}}{g_n} e_n \mathbf{u}_n$.

Quantities g_n and $r_{1,n}$ depend on the EMSE ζ_n , which is not available. Given (13), we suggest to use the following estimator:

$$\hat{\zeta}_n = \max\{\hat{e}_n^2 - \sigma_v^2, 0\}, \quad (33)$$

$$\text{where: } \hat{e}_n = (1 - \gamma)e_n + \gamma\hat{e}_{n-1} \quad (34)$$

which provides an instantaneous approximation of the EMSE, with γ a temporal smoothing factor in the interval $[0, 1)$. To further improve the estimation accuracy of ζ_n , we use $\zeta_{n,\min} = \sigma_u^2 \text{tr}\{\mathbf{Q}_n\}$ calculated with (22) as a lower bound for ζ_n . Indeed, since we minimized $\text{tr}\{\mathbf{Q}_n\}$ with respect to $\{\mu, \rho\}$ at iteration $n-1$, $\hat{\zeta}_n$ cannot be less than $\zeta_{n,\min}$ due to the approximation introduced in the derivation and the inherent variability of signal realizations. This, rather than (33), we suggest to use:

$$\hat{\zeta}_n = \max\{\hat{e}_n^2 - \sigma_v^2, \zeta_{n,\min}\}. \quad (35)$$

Note that non-negativity of μ and ρ is also required. We did not consider this constraint in (22) in order to get closed-form solutions as given by (25) and (26). We further impose a temporal smoothing over parameters μ_n^* and ρ_n^* , as well as a possible upper bound μ_{\max} for the step size μ to ensure the stability of the algorithm:

$$\mu_n = \min\{\gamma' \mu_{n-1} + (1 - \gamma') \mu_n^*, \mu_{\max}\} \quad (36)$$

$$\rho_n = \gamma' \rho_{n-1} + (1 - \gamma') \rho_n^* \quad (37)$$

with γ' a temporal smoothing factor in $[0, 1)$.

4. SIMULATION RESULTS

We shall now present simulation results to illustrate the effectiveness of our algorithms in non-stationary system identification scenarios. In all the experiments, the initial weight vector \mathbf{w}_0 was set to the all-zero vector $\mathbf{0}_L$. The MSD learning curves were obtained by averaging results over 100 Monte-Carlo runs. The VP-GZA-LMS and VP-GRZA-LMS were compared with the standard LMS, GZA-LMS, GRZA-LMS, ZA-VSSLMS [11], and WZA-VSSLMS [11] algorithms. Note that the last two algorithms are variable step size algorithms for general sparse system identification. For group-sparse algorithms, the group size $|\mathcal{G}_j|$ was set to 5 for all j , and ε in (10) was set to 0.1. We set the parameters of all the algorithms so that the initial convergence rate of their MSD was almost the same. Two experiments were performed to compare their tracking behavior and steady-state performance with uncorrelated and correlated inputs.

In the first experiment, we considered a zero-mean white Gaussian input signal u_n to be consistent with A2-A2'. The variance of u_n was set to $\sigma_u^2 = 1$. The additive noise z_n was an i.i.d. zero-mean white Gaussian noise with variance $\sigma_z^2 = 0.01$. The order of the unknown time-varying system was set to $L = 35$. At time instant $n = 1, 8000$ and 16000 , we set the system parameter vector to \mathbf{w}_1^* , \mathbf{w}_2^* and \mathbf{w}_3^* , respectively. Parameter vector \mathbf{w}_2^* was a non-sparse vector, while \mathbf{w}_1^* and \mathbf{w}_3^* were group-sparse. They were defined as:

$$\begin{aligned} \mathbf{w}_1^* &= [0.8, 0.5, 0.3, 0.2, 0.1, \mathbf{0}_{15}, -0.05, -0.1, -0.2, -0.3, \\ &\quad -0.5, \mathbf{0}_5, 0.5, 0.25, 0.5, -0.25, -0.5]^\top; \end{aligned}$$

$$\mathbf{w}_2^* = [0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, \mathbf{1}_{17}, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9]^T;$$

$$\mathbf{w}_3^* = [1.2, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.2, 0.5, 0.4, \mathbf{0}_{15}, -0.4, -0.5, -0.2, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9, -1.2]^T.$$

The results are provided in Fig. 1. Observe that all the algorithms outperformed the LMS in stages \mathbf{w}_1^* and \mathbf{w}_3^* , demonstrating their effectiveness for group-sparse system identification. VP-GZA-LMS and VP-GRZA-LMS algorithms converged as fast as the other algorithms when estimating \mathbf{w}_1^* but reached a lower misadjustment error, especially for VP-GRZA-LMS. The estimation of the non-sparse \mathbf{w}_2^* caused a moderate performance degradation, mainly in their convergence speed. Indeed, their convergence speeds slowed down compared to the other algorithms but they however reached a smaller MSD. The estimation of \mathbf{w}_3^* confirms improved performance and tracking ability of VP-GZA-LMS and VP-GRZA-LMS.

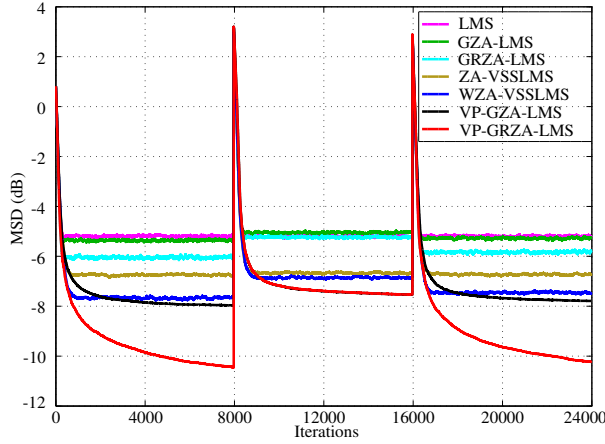


Fig. 1. MSD learning curves for experiment 1 (white input).

In the second experiment, we used the same setting except that the input signal was generated with a first-order AR process given by $u_n = \alpha u_{n-1} + v_n$, where v_n was an i.i.d. zero-mean random variable distributed according to the following Gaussian mixture distribution: $0.5 \mathcal{N}(a \cdot \sigma_v, \sigma_v^2) + 0.5 \mathcal{N}(-a \cdot \sigma_v, \sigma_v^2)$. Its parameters were set to $a = 3/2$ and $\sigma_v^2 = 4/13$, so that $\sigma_u^2 = 1$. The correlation coefficient of u_n was set to $\alpha = 1/2$. In this way, assumptions A2–A2' were both relaxed in order to test the robustness of our approach. The learning curves of all the algorithms are provided in Fig. 2. The evolution of the step size and the regularization parameter over time of VP-GZA-LMS and VP-GRZA-LMS are provided in Fig. 3. Though there was some performance degradation of VP-GZA-LMS and VP-GRZA-LMS algorithms compared with the first experiment, the VP-GRZA-LMS algorithm still led to the lowest steady-state MSD along with the fastest convergence speed among all the competing algorithms for \mathbf{w}_1^* and \mathbf{w}_3^* . The performance of VP-GZA-LMS was almost at the same level as the best of the competing algorithms. Despite the loss of assumptions A2–A2', VP-GZA-LMS and VP-GRZA-LMS algorithms still worked well with non-Gaussian correlated inputs. The results in Fig. 3 show that VP-GZA-LMS and VP-GRZA-LMS set the step size and regularization parameter to large values at the beginning of each estimation stage in order to ensure tracking ability and promote sparsity. Then they gradually reduced them to ensure small MSD.

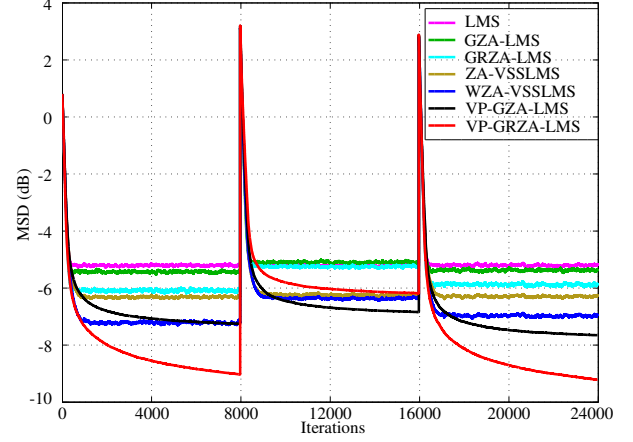


Fig. 2. MSD learning curves for experiment 2 (non-Gaussian colored input).

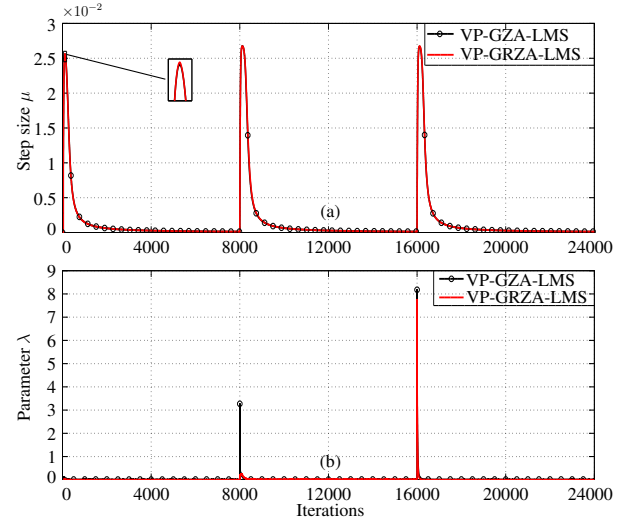


Fig. 3. (a) Evolution of the step size μ and (b) the regularization parameter λ of VP-GZA-LMS and VP-GRZA-LMS algorithms for experiment 2.

5. CONCLUSION

In this paper, we introduced VP-GZA-LMS and VP-GRZA-LMS algorithms to address online group-sparse system identification problems. Based on a model of the transient behavior of the GRZA-LMS algorithm, we proposed to minimize the MSD with respect to the step size and regularization parameter, simultaneously, at each iteration. This led to a convex optimization problem with a closed-form solution. Simulation results demonstrated the effectiveness of VP-GZA-LMS and VP-GRZA-LMS algorithms over other existing variable step size methods, even for non-Gaussian correlated inputs.

6. REFERENCES

- [1] A. H. Sayed, *Adaptive Filters*, John Wiley & Sons, Inc., 2008.
- [2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1985.
- [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, Sep. 2000.
- [4] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proceedings of the IEEE ICASSP*, May 2002, vol. 2, pp. II–1881–II–1884.
- [5] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proceedings of the IEEE ICASSP*, Taipei, China, Apr. 2009, pp. 3125–3128.
- [6] Y. Chen and A. O. Hero, "Recursive $\ell_{1,\infty}$ group lasso," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 3978–3987, Aug. 2012.
- [7] Y. Chen, Y. Gu, and A. O. Hero, "Regularized least-mean-square algorithms," 2010, Arxiv preprint arXiv:1012.5066.
- [8] W. F. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958–981, Jun. 1995.
- [9] J. Huang and T. Zhang, "The benefit of group sparsity," *The Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, Aug. 2010.
- [10] C. Paleologu, J. Benesty, and S. Ciochină, "A variable step-size proportionate NLMS algorithm for echo cancellation," *Revue Roumaine des Sciences Techniques – Serie Electrotechnique et Energetique*, vol. 53, pp. 309–317, 2008.
- [11] M. S. Salman, M. N. S. Jahromi, A. Hocanin, and O. Kukrer, "A zero-attracting variable step-size LMS algorithm for sparse system identification," in *IX International Symposium on Telecommunications (BIHTEL)*, Oct. 2012, pp. 1–4.
- [12] T. Fan and Y. Lin, "A variable step-size strategy based on error function for sparse system identification," *Circuits, Systems, and Signal Processing*, vol. 36, no. 3, pp. 1301–1310, Mar. 2017.
- [13] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Processing Letters*, vol. 13, no. 10, pp. 581–584, Oct 2006.
- [14] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.
- [15] J. Chen, C. Richard, Y. Song, and D. Brie, "Transient performance analysis of zero-attracting LMS," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1786–1790, Dec. 2016.
- [16] S. Haykin, *Adaptive Filter Theory*, Pearson Education India, 4th edition, 2005.
- [17] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2733–2748, 2015.