ADAPTIVE PARAMETER ADJUSTMENT FOR GROUP REWEIGHTED ZERO-ATTRACTING LMS

Danqi Jin ⋆, *Jie Chen* ⋆, *Cédric Richard* †, *Jingdong Chen* ⋆

⋆ Centre of Intelligent Acoustics and Immersive Communications
School of Marine Science and Technology, Northwestern Polytechinical University, China
† Laboratoire Lagrange, Université de Nice Sophia - Antipolis, CNRS, France
danqijin@mail.nwpu.edu.cn dr.jie.chen@ieee.org cedric.richard@unice.fr jingdongchen@ieee.org

ABSTRACT

Group zero-attracting LMS (GZA-LMS) and its reweighted variant (GRZA-LMS) have been proposed for system identification with structural group sparsity of the parameter vector. Similar to most adaptive filtering algorithms with regularized penalty, GZA-LMS/GRZA-LMS suffers from a trade-off between convergence rate and steady-state performance, meanwhile between the degree of sparsity and estimation bias. Therefore, it is pivotal to properly set the step-size and regularization parameter of the algorithms. Based on a transient behavior model of GZA-LMS/GRZA-LMS, a variable-parameter GRZA-LMS algorithm is proposed to address this issue. By minimizing the 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 in both white and colored input cases.

Index Terms— Sparse system identification, group sparsity, transient behavior model, variable parameter strategy, adaptive algorithms.

1. INTRODUCTION

Adaptive filtering algorithms serve as a very useful tool for online system identification [1, 2]. Within the myriad of adaptive algorithms, the least-mean-square (LMS) algorithm has been widely used due to its robustness, relatively good performance and low complexity. It is important to endow the standard LMS algorithm with other properties . One of the most useful properties is to promote sparsity of the estimate, frequently required in several applications such as online sparse channel identification. In such a scenario, though the impulse response can be long, only a few of the coefficients have significant values. Several algorithms based on the LMS were proposed to promote sparsity, such as proportionate normalized LMS (PNLMS) [3, 4], zero-attracting LMS (ZA-LMS) and reweighted zero-attracting LMS (RZA-LMS) [5]. These variants of LMS all claimed improved performance in sparse situations. Beyond element-wise sparsity, a further consideration is that some sparse systems can be group-sparse system [6, 7], utilizing such a structural priori information can achieve enhanced performance. Opposed to the general sparse system with impulse response having an arbitrary structure, a group-sparse system has impulse response composed of a few distinct clusters of nonzero coefficients, such as specular multipath acoustic and wireless channels [6–8]. By utilizing the mixed norm regularization, the $\ell_{1,\infty}$-regularized RLS algorithm [6], group ZA-LMS (GZA-LMS) and group RZA-LMS (GRZA-LMS) [7] algorithms were proposed to promote group-sparsity of estimates. To ensure a performance gain in such group-sparse scenarios [6, 7, 9], setting the algorithm parameters such as step size and regularization parameter remains a tricky task. On one hand, the step-size plays a crucial role to control 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 an inappropriate value of these parameters may even deteriorate the estimation performance.

Variable parameter strategies provide a simple but efficient way to achieve a reasonable trade-off of competing performance requirement [10]. Several variable step-size strategies have been proposed for LMS and ZA-LMS [10–14], and the step-size are adjusted according to the estimation error mostly. While for group-sparse LMS, there is little work addressing this issue. Motivated by [15], we propose in this paper to design a variable-parameter GZA-LMS (VP-GZA-LMS) and variable-parameter GRZA-LMS (VP-GRZA-LMS) algorithms. The proposed method is based on an optimization problem formulation stemming from the stochastic performance model of the algorithm. This makes the proposed strategy different from several other heuristic candidates. By minimizing the mean-square-deviation (MSD) at each iteration, we obtain closed-form expression of the optimal step-size and regularization parameter, leading to a faster convergence as well as a lower misadjustment.

Notation. Normal font x denotes scalars. Boldface lowercase letters $\boldsymbol{x}$ and uppercase letters $\boldsymbol{X}$ 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 operator $\max\{\cdot,\cdot\}$ and $\min\{\cdot,\cdot\}$ take the maximum or minimum of two arguments. $\bigcup$ and $\bigcap$ denote union and intersection of a collection of sets, respectively. $\emptyset$ denotes empty set.

2. SYSTEM MODEL AND GROUP-SPARSE LMS

Consider an unknown system with output d_n characterized by the linear model

$$d_n = \boldsymbol{u}_n^\top \boldsymbol{w}^\star + v_n, \tag{1}$$

where $\boldsymbol{w}^\star \in \mathbb{R}^L$ is an unknown parameter vector, and $\boldsymbol{u}_n \in \mathbb{R}^L$ is a zero-mean regression vector with positive definite covariance ma-

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.

trix. The reference signal d_n is assumed to be zero mean. The error signal v_n is assumed to be stationary, independent and identically distributed (i.i.d.), with zero mean and variance σ_v^2, and independent of any other signal.

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

$$J(\boldsymbol{w}) = \frac{1}{2}\,\mathbb{E}\big\{[d_n - \boldsymbol{w}^\top \boldsymbol{u}_n]^2\big\}. \tag{2}$$

with $\boldsymbol{w}^\star$ being the minimizer of $J(\boldsymbol{w})$. In this paper we consider the problem of estimating the unknown parameter vector $\boldsymbol{w}^\star$ with a group-sparse structure. This can be addressed by minimizing the following regularized MSE cost:

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

where the $\ell_{1,2}$-norm is used to promote the group-sparsity of the estimate, and $\lambda \geq 0$ is the regularization parameter. The $\ell_{1,2}$-norm of a vector $\boldsymbol{w}$ is defined as $\|\boldsymbol{w}\|_{1,2} = \sum_{j=1}^{J} \|\boldsymbol{w}_{\mathcal{G}_j}\|_2$, where $\{\mathcal{G}_j\}_{j=1}^{J}$ is a group partition of the whole index set $\mathcal{G} = \{0, 1, \ldots, L-1\}$, satisfying: (1) $\bigcup_{j=1}^{J} \mathcal{G}_j = \mathcal{G}$, (2) $\mathcal{G}_j \bigcap \mathcal{G}_l = \emptyset$ when $j \neq l$. And $\boldsymbol{w}_{\mathcal{G}_j}$ denotes a sub-vector of $\boldsymbol{w}$ indexed by $\mathcal{G}_j$. Using a sub-gradient update, the iteration of $\boldsymbol{w}$ in GZA-LMS is given in sub-vector form:

$$\boldsymbol{w}_{n+1,\mathcal{G}_j} = \boldsymbol{w}_{n,\mathcal{G}_j} + \mu\, e_n \boldsymbol{u}_{n,\mathcal{G}_j} - \rho\; \boldsymbol{s}_{n,\mathcal{G}_j} \tag{4}$$

for $j = 1, \ldots, J$, where

$$\boldsymbol{s}_{n,\mathcal{G}_j} = \begin{cases} \frac{\boldsymbol{w}_{n,\mathcal{G}_j}}{\|\boldsymbol{w}_{n,\mathcal{G}_j}\|_2} & \text{for } \|\boldsymbol{w}_{n,\mathcal{G}_j}\|_2 \neq 0 \\ 0 & \text{for } \|\boldsymbol{w}_{n,\mathcal{G}_j}\|_2 = 0, \end{cases} \tag{5}$$

where $e_n = d_n - \boldsymbol{w}_n^\top \boldsymbol{u}_n$ is the estimation error, μ is a positive step-size, $\boldsymbol{u}_{n,\mathcal{G}_j}$ is a sub-vector of $\boldsymbol{u}_n$ corresponding to $\boldsymbol{w}_{n,\mathcal{G}_j}$, and the shrinkage parameter $\rho = \mu\lambda$.

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

$$\begin{aligned} &\boldsymbol{w}^o_{\text{GRZA}} = \arg\min_{\boldsymbol{w}} J_{\text{GRZA}}(\boldsymbol{w}) \quad \text{with} \\ &J_{\text{GRZA}}(\boldsymbol{w}) = \frac{1}{2}\,\mathbb{E}\big\{[d_n - \boldsymbol{w}^\top \boldsymbol{u}_n]^2\big\} + \lambda \sum_{j=1}^{J} \log\Big(1 + \frac{\|\boldsymbol{w}_{\mathcal{G}_j}\|_2}{\varepsilon}\Big), \end{aligned} \tag{6}$$

where the log-sum penalty has been introduced to make group-sparsity attractor take effort only on groups at the same level of ε [5]. Similarly, using a sub-gradient update yields the GRZA-LMS:

$$\boldsymbol{w}_{n+1,\mathcal{G}_j} = \boldsymbol{w}_{n,\mathcal{G}_j} + \mu\, e_n \boldsymbol{u}_{n,\mathcal{G}_j} - \rho\, \beta_{n,j} \boldsymbol{s}_{n,\mathcal{G}_j}, \tag{7}$$

where $\beta_{n,j} = \frac{1}{\|\boldsymbol{w}_{n,\mathcal{G}_j}\|_2 + \varepsilon}$ is a weighting coefficient. Equivalently, (7) in vector form is given by:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\, e_n \boldsymbol{u}_n - \rho\, \boldsymbol{\beta}_n \circ \boldsymbol{s}_n, \tag{8}$$

where $\boldsymbol{\beta}_n$ and $\boldsymbol{s}_n$ are vector form of $\beta_{n,j}$ and $\boldsymbol{s}_{n,\mathcal{G}_j}$ with dimension $L \times 1$, and $\circ$ denotes Hadamard product.

Furthermore, we observe from (4) and (7) that the GRZA-LMS reduces to GZA-LMS by simply replacing the parameters $\beta_{n,j}$ with 1 for $j = 1, \ldots, J$, that is,

$$\beta_{n,j} = \begin{cases} \frac{1}{\|\boldsymbol{w}_{n,\mathcal{G}_j}\|_2 + \varepsilon} & \text{GRZA-LMS} \\ 1 & \text{GZA-LMS.} \end{cases} \tag{9}$$

We will thus derive a variable parameter strategy for GZA-LMS and GRZA-LMS on the unified form (8), while specific algorithms can be obtained by setting $\beta_{n,j}$ according to (9).

3. MODEL-BASED PARAMETER DESIGN OF GRZA-LMS

3.1. Transient Behavior Model of GRZA-LMS

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

$$\tilde{\boldsymbol{w}}_n = \boldsymbol{w}_n - \boldsymbol{w}^\star. \tag{10}$$

To derive a variable parameter strategy, we first study the stochastic transient behavior of the second-order moments of $\tilde{\boldsymbol{w}}_n$ over time. Besides, to keep the calculations mathematically tractable, we introduce the commonly used independence assumption [1]:

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

Subtracting $\boldsymbol{w}^\star$ from both sides of (8), and using $e_n = v_n - \tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n$, yields the update of $\tilde{\boldsymbol{w}}_n$:

$$\tilde{\boldsymbol{w}}_{n+1} = \tilde{\boldsymbol{w}}_n + \mu \boldsymbol{u}_n v_n - \mu \boldsymbol{u}_n \boldsymbol{u}_n^\top \tilde{\boldsymbol{w}}_n - \rho \boldsymbol{\beta}_n \circ \boldsymbol{s}_n. \tag{11}$$

Using the independence Assumption A1 and relation $e_n = v_n - \tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n$, the MSE of the GRZA-LMS is given by

$$\mathbb{E}\{e_n^2\} = \sigma_v^2 + \text{tr}\{\boldsymbol{R}_u \boldsymbol{Q}_n\} \tag{12}$$

with $\boldsymbol{Q}_n = \mathbb{E}\{\tilde{\boldsymbol{w}}_n \tilde{\boldsymbol{w}}_n^\top\}$. The quantity $\text{tr}\{\boldsymbol{R}_u \boldsymbol{Q}_n\}$ is the excess-mean-square-error (EMSE) at time instant n, denoted by ζ_n. The trace of $\boldsymbol{Q}_n$ is the MSD, denoted by $\xi_n = \text{tr}\{\boldsymbol{Q}_n\}$. Besides, we introduce the whiteness assumption A2 [1] to simplify the derivation:

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

Though introducing Assumption A2 to simplify the derivation, it turns out that the resulting algorithms work well in moderately correlated input scenarios where assumption A2 does not hold. Under Assumption A2, we relate MSD to EMSE via a scaling factor:

$$\zeta_n = \sigma_u^2\, \text{tr}\{\boldsymbol{Q}_n\} = \sigma_u^2\, \xi_n. \tag{13}$$

Therefore, we need to determine a recursion for $\text{tr}\{\boldsymbol{Q}_n\}$ in order to relate the MSD at two consecutive time instants n and $n+1$. Post-multiplying (11) by its transpose, taking the expectation and matrix trace, using Assumptions A1 and A2, we get:

$$\text{tr}\{\boldsymbol{Q}_{n+1}\} = \text{tr}\{\boldsymbol{Q}_n\} + \mu^2 g + \rho^2 h + 2\mu\rho l - 2\mu r_1 - 2\rho r_2 \tag{14}$$

with

$$g = \sigma_v^2\, \text{tr}\{\boldsymbol{R}_u\} + \mathbb{E}\{\boldsymbol{u}_n^\top \tilde{\boldsymbol{w}}_n \tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n \boldsymbol{u}_n^\top \boldsymbol{u}_n\} \tag{15}$$

$$h = \mathbb{E}\big\{(\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)^\top (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)\big\} \tag{16}$$

$$l = \mathbb{E}\big\{\tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n \boldsymbol{u}_n^\top (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)\big\} \tag{17}$$

$$r_1 = \mathbb{E}\big\{\tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n \boldsymbol{u}_n^\top \tilde{\boldsymbol{w}}_n\big\} \tag{18}$$

$$r_2 = \mathbb{E}\big\{(\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)^\top \tilde{\boldsymbol{w}}_n\big\}. \tag{19}$$

We have dropped the time index n in left hand side of (15) – (19) for compactness.

3.2. Parameter Design Using Transient Behavior Model

Now we derive a parameter design strategy for GRZA-LMS using model (14). Given the MSD ξ_n at time instant n, we solve for the parameters that minimize the MSD ξ_{n+1}:

$$\{\mu_n^\star, \rho_n^\star\} = \arg\min_{\mu,\rho} \, \xi_{n+1} \,|\, \xi_n. \tag{20}$$

Using the recursion (14), the above optimization problem becomes:

$$\begin{aligned}\{\mu_n^\star, \rho_n^\star\} &= \arg\min_{\mu,\rho} \, \mathrm{tr}\{\boldsymbol{Q}_{n+1}\} \\ &= \arg\min_{\mu,\rho} \, \mathrm{tr}\{\boldsymbol{Q}_n\} + \mu^2 g + \rho^2 h + 2\mu\rho l - 2\mu r_1 - 2\rho r_2.\end{aligned} \tag{21}$$

Equivalently, equation (21) can be written in matrix form as:

$$\xi_{n+1} = [\mu \; \rho] \, \boldsymbol{H} \, [\mu \; \rho]^\top - 2\,[r_1 \; r_2]\,[\mu \; \rho]^\top + \xi_n, \tag{22}$$

which is a quadratic function of $[\mu \; \rho]^\top$, with $\boldsymbol{H} = \begin{bmatrix} g & l \\ l & h \end{bmatrix}$.

By decomposing g into two additive terms, it can be proved that Hessian matrix $\boldsymbol{H}$ could be written as the sum of a covariance matrix and a positive semidefinite matrix, yielding $\boldsymbol{H}$ is positive semidefinite. Since in practice a covariance matrix is always almost positive definite [16], we assume further that $\boldsymbol{H}$ is positive definite, which allows us to obtain the optimal parameters via:

$$[\mu_n^\star \; \rho_n^\star]^\top = \boldsymbol{H}^{-1} [r_1 \; r_2]^\top. \tag{23}$$

Using matrix calculation leads to:

$$\mu_n^\star = \frac{h r_1 - l r_2}{g h - l^2} \tag{24}$$

$$\rho_n^\star = \frac{g r_2 - l r_1}{g h - l^2}. \tag{25}$$

However, the above result cannot be used since it requires statistics that are not available in online learning scenarios. We now adopt an approximation for these quantities. The subscript n as time index in variables g_n, h_n, l_n, r_{1n} and r_{2n} is now added for clearance . With Assumption A2 the quantity g_n can be computed as:

$$\begin{aligned} g_n &= \sigma_v^2 \, \mathrm{tr}\{\sigma_u^2 \boldsymbol{I}\} + \mathrm{tr}\Big\{2\boldsymbol{R}_u \boldsymbol{Q}_n \boldsymbol{R}_u + \mathrm{tr}\{\boldsymbol{R}_u \boldsymbol{Q}_n\} \boldsymbol{R}_u\Big\} \\ &= \sigma_v^2 \sigma_u^2 \, L + (2 + L) \sigma_u^2 \, \zeta_n. \end{aligned} \tag{26}$$

Using the independence Assumption A1 yields:

$$r_{1_n} = \zeta_n. \tag{27}$$

Then, approximating the expectation in (16), (17) and (19) by their instantaneous argument yields:

$$h_n \approx (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)^\top (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n) \tag{28}$$

$$l_n \approx \tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n \boldsymbol{u}_n^\top (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n) \tag{29}$$

$$r_{2n} \approx (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)^\top \tilde{\boldsymbol{w}}_n. \tag{30}$$

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

$$\hat{\boldsymbol{w}}_n^\star = \boldsymbol{w}_n - \eta_n \nabla J(\boldsymbol{w}_n) \tag{31}$$

where η_n is a positive step-size to be determined. Given the MSD ξ_n, we seek η_n that minimizes ξ_{n+1}. Following (20)–(25) leads to $\eta_n = r_{1n}/g_n$. Further, we approximate the true gradient $\nabla J(\boldsymbol{w}_n)$ with instantaneous value $-e_n \boldsymbol{u}_n$. Finally, we have $\hat{\boldsymbol{w}}_n^\star = \boldsymbol{w}_n - \boldsymbol{p}_n$ with $\boldsymbol{p}_n = -\frac{r_{1n}}{g_n} e_n \boldsymbol{u}_n$.

We then adopt the estimator $\hat{\zeta}_n$ for unknown EMSE ζ_n:

$$\hat{e}_n = (1 - \lambda) e_n + \lambda \hat{e}_{n-1} \tag{32}$$

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

where $\lambda \in [0, 1)$ is a smoothing factor. To further improve the estimation accuracy, we use $\zeta_{n_{\min}} = \sigma_u^2 \, \mathrm{tr}\{\boldsymbol{Q}_n\}$ calculated via (21) as a lower bound of ζ_n, since we have minimized $\mathrm{tr}\{\boldsymbol{Q}_n\}$ at iteration $n - 1$. Due to the approximation introduced in the derivation and the inherent properties of signal realization, ζ_n is no less than $\sigma_u^2 \, \mathrm{tr}\{\boldsymbol{Q}_n\}$. Then instead of (33), we use

$$\hat{\zeta}_n = \max\big\{\hat{e}_n^2 - \sigma_v^2, \, \zeta_{n_{\min}}\big\}. \tag{34}$$

Non-negative constraint of μ and ρ is needed. We did not considered it in (21) in order to get closed-form solutions (24) and (25). Now we need to constrain μ and ρ with the following operators:

$$\mu_n^\star = \max\{\mu_n^\star, \, 0\} \tag{35}$$

$$\rho_n^\star = \max\{\rho_n^\star, \, 0\}. \tag{36}$$

We further impose a temporal smoothing over parameters $\mu_n^\star$ and $\rho_n^\star$, meanwhile a predefined upper bound $\mu_{\max}$ of step-size to ensure the stability of the algorithm:

$$\mu_n = \min\{\beta \mu_{n-1} + (1 - \beta) \mu_n^\star, \, \mu_{\max}\} \tag{37}$$

$$\rho_n = \beta \rho_{n-1} + (1 - \beta) \rho_n^\star. \tag{38}$$

4. SIMULATION RESULTS

Now we present simulation results to illustrate the effectiveness of our algorithms in non-stationary system identification applications. The input signal was generated via a first-order AR process defined as $u_n = \alpha \, u_{n-1} + z_n$, where z_n is an i.i.d. zero-mean Gaussian variable with variance $\sigma_z^2 = 1 - \alpha^2$ (so that $\sigma_u^2 = 1$), and α is the correlation coefficient of u_n. We obtained processes u_n with different levels of correlation through varying the value of α. v_n was an i.i.d. zero-mean white Gaussian noise with variance $\sigma_v^2 = 0.01$. In all the experiments, the initial weight vector $\boldsymbol{w}_0$ was set to the all-zero vector $\boldsymbol{0}_L$. The MSD learning curves were obtained by averaging results over 100 Monte-Carlo runs. Besides, the VP-GZA-LMS and VP-GRZA-LMS were compared with the standard LMS, GZA-LMS, GRZA-LMS, ZA-VSSLMS [11], WZA-VSSLMS [11] algorithms. The last two were variable step-size algorithms for general sparse system. For group-sparse algorithms, the group size was set to 5, and ε of (9) was 0.1. We set parameters of all algorithms so that their initial convergence speed were almost the same[1]. Then by comparing the steady-state MSD, we evaluate their performance. Two experiments were designed to illustrate the tracking and steady-state behaviors of the algorithms with uncorrelated and correlated input signals.

In the first experiment, we compared all the algorithms using white input signals to make it coincident with design Assumption A2. The order of the unknown time-varying system was set to $L = 35$. At time instant $n = 1$, $n = 8000$ and $n = 16000$, we set the

[1] In non-stationary system identification problems, we set the parameters only to guarantee almost the same convergence rate during the first group of system parameters $\boldsymbol{w}_1^\star$, not for all $\boldsymbol{w}^\star$.

system parameter vector to $\boldsymbol{w}_1^\star$, $\boldsymbol{w}_2^\star$ and $\boldsymbol{w}_3^\star$ respectively. $\boldsymbol{w}_2^\star$ was a non-sparse one, while the rest had a group-sparse structure. The parameter vectors $\boldsymbol{w}_1^\star$, $\boldsymbol{w}_2^\star$ and $\boldsymbol{w}_3^\star$ were defined as:

$$\boldsymbol{w}_1^\star = [0.8,\, 0.5,\, 0.3,\, 0.2,\, 0.1,\, \mathbf{0}_{15},\, -0.05,\, -0.1,\, -0.2,\, -0.3,\, -0.5,\, \mathbf{0}_5,\, 0.5,\, 0.25,\, 0.5,\, -0.25,\, -0.5]^\top;$$

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

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

The result is illustrated in Fig. 1. It shows that all other algorithms outperform LMS in stages $\boldsymbol{w}_1^\star$ and $\boldsymbol{w}_3^\star$, demonstrating their effectiveness for group-sparse systems. Further, for VP-GZA-LMS and VP-GRZA-LMS algorithms, they converged as fast as other competing algorithms when estimating the group-sparse $\boldsymbol{w}_1^\star$, meanwhile maintaining a lower misadjustement, especially for VP-GRZA-LMS. The estimation of the non-sparse $\boldsymbol{w}_2^\star$ caused a moderate performance degradation, mainly in their convergence speed. In this stage, their convergence speeds slowed down compared to the other algorithms but they reached a smaller MSD. The estimation of $\boldsymbol{w}_3^\star$ confirms the superior performance and tracking capability of VP-GZA-LMS and VP-GRZA-LMS algorithms.

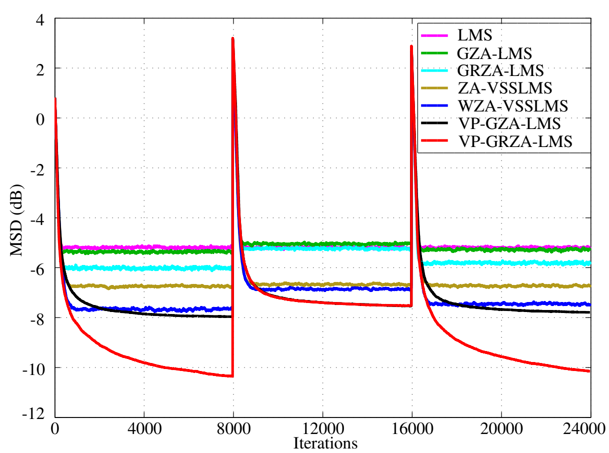

Fig. 1. Transient behaviors of the compared algorithms in a time-varying system with a white input.

In the second experiment, we used the same setting except that the correlation coefficient α was changed to 0.5. The learning curves of all algorithms are provided in Fig. 2. Besides, for VP-GZA-LMS and VP-GRZA-LMS, the evolution of step-size and regularization parameter over time are provided in Fig. 3. Though there was some performance degradation of VP-GZA-LMS and VP-GRZA-LMS algorithm compared with the first experiment, the VP-GRZA-LMS algorithm still yielded the lowest steady-state MSD along with the fastest convergence speed among all the competing algorithms for group-sparse systems $\boldsymbol{w}_1^\star$ and $\boldsymbol{w}_3^\star$. While for VP-GZA-LMS, its performance is almost on the same level as the best of the competing algorithms. Despite the loss of the whiteness Assumption, the VP-GZA-LMS and VP-GRZA-LMS algorithms still work well with correlated inputs for group-sparse system. Additionally, results in Fig. 3 shows that VP-GZA-LMS and VP-GRZA-LMS set the step-size and the regularization parameter to large values in order to ensure tracking ability and promote sparsity at the beginning of each estimation phase. Then they gradually reduced these values to ensure small MSD.

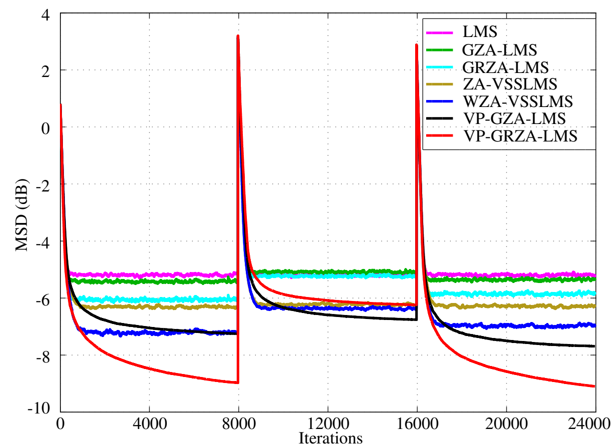

Fig. 2. Transient behaviors of compared algorithms in a time-varying system with a colored input.

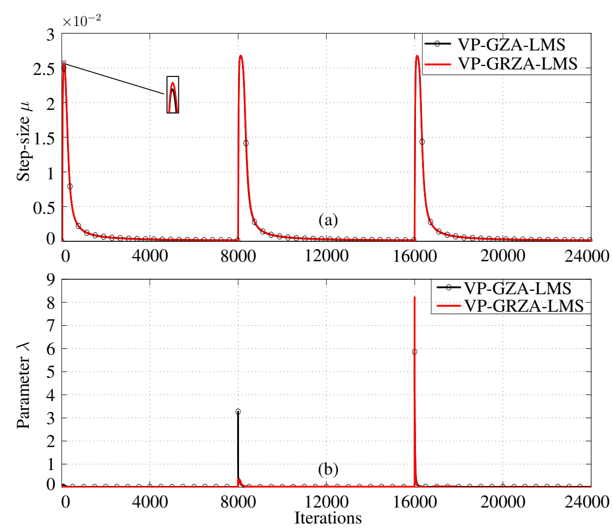

Fig. 3. (a) Step-sizes μ and (b) Regularization parameter λ of VP-GZA-LMS and VP-GRZA-LMS algorithms in the case with the colored input.

5. CONCLUSIONS

In this paper, we introduced VP-GZA-LMS and VP-GRZA-LMS algorithms to address online group-sparse system identification problems. Based on the transient behavior model of the GRZA-LMS, 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 algorithms. In addition, VP-GZA-LMS and VP-GRZA-LMS depend on a few number of hyperparameters that do not drastically affect the performance.

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.