Adaptive Parameters Adjustment for Group Reweighted Zero-Attracting LMS

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

*Northwestern Polytechnical University, China † Université Côte d'Azur, CNRS, OCA, France

Introduction and main contributions

- **Introduction:** Group zero-attracting LMS (GZA-LMS) and group reweighted zero-attracting LMS (GRZA-LMS) have been proposed for addressing system identification problems with structural group sparsity. Both algorithms however suffer from a trade-off between sparsity degree and estimation bias and, between convergence speed and steady-state performance. 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.
- **Contributions**
  - We derive closed-form expressions of the optimal step size and regularization parameter.
  - New algorithms achieve a lower mean-square deviation (MSD).

System model and group-sparse LMS

- Consider the time sequence $\{d_n, \boldsymbol{u}_n\}$ related via the linear model

$$d_n = \boldsymbol{u}_n^\top \boldsymbol{w}^\star + z_n$$

with $\boldsymbol{u}_n \in \mathbb{R}^L$ and $\boldsymbol{w}^\star \in \mathbb{R}^L$.

- To determine $\boldsymbol{w}^\star$, consider the MSE criterion with $\ell_{1,2}$-norm regularization:

$$\boldsymbol{w}^{o}_{\text{GZA}} = \arg\min_{\boldsymbol{w}} \frac{1}{2}\mathbb{E}\{[d_n - \boldsymbol{w}^\top\boldsymbol{u}_n]^2\} + \lambda\|\boldsymbol{w}\|_{1,2}, \qquad \|\boldsymbol{w}\|_{1,2} = \sum_{j=1}^{J}\|\boldsymbol{w}_{\mathcal{G}_j}\|_2$$

  - $\ell_{1,2}$-norm is used to promote the group-sparsity of the estimate.
  - GZA-LMS and GRZA-LMS algorithm:

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

where $e_n = d_n - \boldsymbol{w}_n^\top \boldsymbol{u}_n$, μ is the step size, $\rho = \mu\lambda$, $\boldsymbol{s}_n$ is vector form of $\boldsymbol{s}_{n,\mathcal{G}_j}$,

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

$\boldsymbol{\beta}_n$ is vector form of $\beta_{n,j}$, $\beta_{n,j} = 1/[\|\boldsymbol{w}_{n,\mathcal{G}_j}\|_2 + \varepsilon]$ corresponds to GRZA-LMS, $\beta_{n,j} = 1$ corresponds to GZA-LMS, symbol $\circ$ denotes the Hadamard product.

- Trade-off
  - μ controls the trade-off between convergence speed and steady-state performance.
  - ρ controls the trade-off between sparsity degree and estimation bias.

It is therefore crucial to adaptively adjust μ and ρ.

Transient behavior model of GRZA-LMS

- Define the weight error vector and its covariance matrix by

$$\tilde{\boldsymbol{w}}_n = \boldsymbol{w}_n - \boldsymbol{w}^\star \qquad \text{and} \qquad \boldsymbol{Q}_n = \mathbb{E}\{\tilde{\boldsymbol{w}}_n \tilde{\boldsymbol{w}}_n^\top\}$$

The recursion of $\tilde{\boldsymbol{w}}_n$ writes

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

- Assumptions:
  - **A1**: The weight error vector $\tilde{\boldsymbol{w}}_n$ is statistically independent of $\boldsymbol{u}_n$.
  - **A2**: The input regressor $\boldsymbol{u}_n$ is a zero-mean white signal with covariance matrix $\boldsymbol{R}_u = \sigma_u^2 \boldsymbol{I}$.
  - **A2'**: The input regressor $\boldsymbol{u}_n$ is Gaussian distributed.
- With the independence assumption **A1**, we have:

$$\text{MSE: } \mathbb{E}\{e_n^2\} = \sigma_z^2 + \text{trace}\{\boldsymbol{R}_u \boldsymbol{Q}_n\}$$

By utilizing white input assumption **A2**, we have:

$$\text{EMSE: } \zeta_n = \text{trace}\{\boldsymbol{R}_u \boldsymbol{Q}_n\} = \sigma_u^2\, \text{trace}\{\boldsymbol{Q}_n\} = \sigma_u^2 \xi_n \rightarrow \text{MSD}$$

Under assumptions **A1** and **A2**: min MSE $\Longleftrightarrow$ min MSD

- Determine a recursion to relate the MSD at two consecutive time instants:

$$\text{trace}\{\boldsymbol{Q}_{n+1}\} = \text{trace}\{\boldsymbol{Q}_n\} + \mu^2 g + \rho^2 h + 2\mu\rho\ell - 2\mu r_1 - 2\rho r_2$$

with

$$g = \sigma_z^2\, \text{trace}\{\boldsymbol{R}_u\} + \mathbb{E}\{\boldsymbol{u}_n^\top \tilde{\boldsymbol{w}}_n \tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n \boldsymbol{u}_n^\top \boldsymbol{u}_n\} \leftarrow \textbf{A2'}$$

$$h = \mathbb{E}\{(\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)^\top (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)\}, \quad \ell = \mathbb{E}\{\tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n \boldsymbol{u}_n^\top (\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)\}$$

$$r_1 = \mathbb{E}\{\tilde{\boldsymbol{w}}_n^\top \boldsymbol{u}_n \boldsymbol{u}_n^\top \tilde{\boldsymbol{w}}_n\}, \quad r_2 = \mathbb{E}\{(\boldsymbol{\beta}_n \circ \boldsymbol{s}_n)^\top \tilde{\boldsymbol{w}}_n\}.$$

- How to derive an adaptive parameters adjustment strategy?

Parameter design using transient behavior model

- Given the MSD ξ_n at time instant n, we determine the parameters $\{\mu_n, \rho_n\}$ that minimize the MSD ξ_{n+1}:

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

Using the recursion of trace$\{\boldsymbol{Q}_{n+1}\}$, we have:

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

Equivalently, in matrix form:

$$\xi_{n+1} = [\mu\ \rho]\, \boldsymbol{H}\, [\mu\ \rho]^\top - 2\,[r_1\ r_2]\,[\mu\ \rho]^\top + \xi_n, \quad \text{with } \boldsymbol{H} = \begin{bmatrix} g & \ell \\ \ell & h \end{bmatrix}$$

which is a quadratic function of $[\mu\ \rho]$, and $\boldsymbol{H}$ is a positive semidefinite matrix.

- Solution:

$$[\mu_n^\star\ \rho_n^\star]^\top = \boldsymbol{H}^{-1}[r_1\ r_2]^\top,$$

i.e.,

$$\mu_n^\star = \frac{h r_1 - \ell r_2}{gh - \ell^2}, \qquad \rho_n^\star = \frac{g r_2 - \ell r_1}{gh - \ell^2}.$$

  - Adopt approximations for quantities: g, h, ℓ, r_1, r_2.
  - Impose nonnegative constraints as well as temporal smoothing for $\mu_n^\star$ and $\rho_n^\star$.

Simulation results

Consider non-stationary system identification scenarios:

- System parameter vectors:

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

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

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

At time instant $n = 1$, 8000 and 16000, we set the system parameter vector to $\boldsymbol{w}_1^\star$, $\boldsymbol{w}_2^\star$ and $\boldsymbol{w}_3^\star$, respectively.

- Input signal:
  - Experiment 1: Zero-mean white Gaussian with $\sigma_u^2 = 1$. ← **A1, A2, A2'**
  - Experiment 2: Generated from a first-order AR process

$$u_n = 0.5\, u_{n-1} + v_n$$

with zero-mean random variable v_n generated from Gaussian mixture distribution

$$0.5\,\mathcal{N}(a \cdot \sigma_v, \sigma_v^2) + 0.5\,\mathcal{N}(-a \cdot \sigma_v, \sigma_v^2).$$

- Additive noise: z_n was zero-mean i.i.d. Gaussian with $\sigma_z^2 = 0.01$.
- Parameters: $L = 35$, group size $|\mathcal{G}_j| = 5$, $\varepsilon = 0.1$. We set the parameters of all the algorithms so that the initial convergence rate of their MSD was almost the same.
- MSD learning curves (average of 100 MC runs)

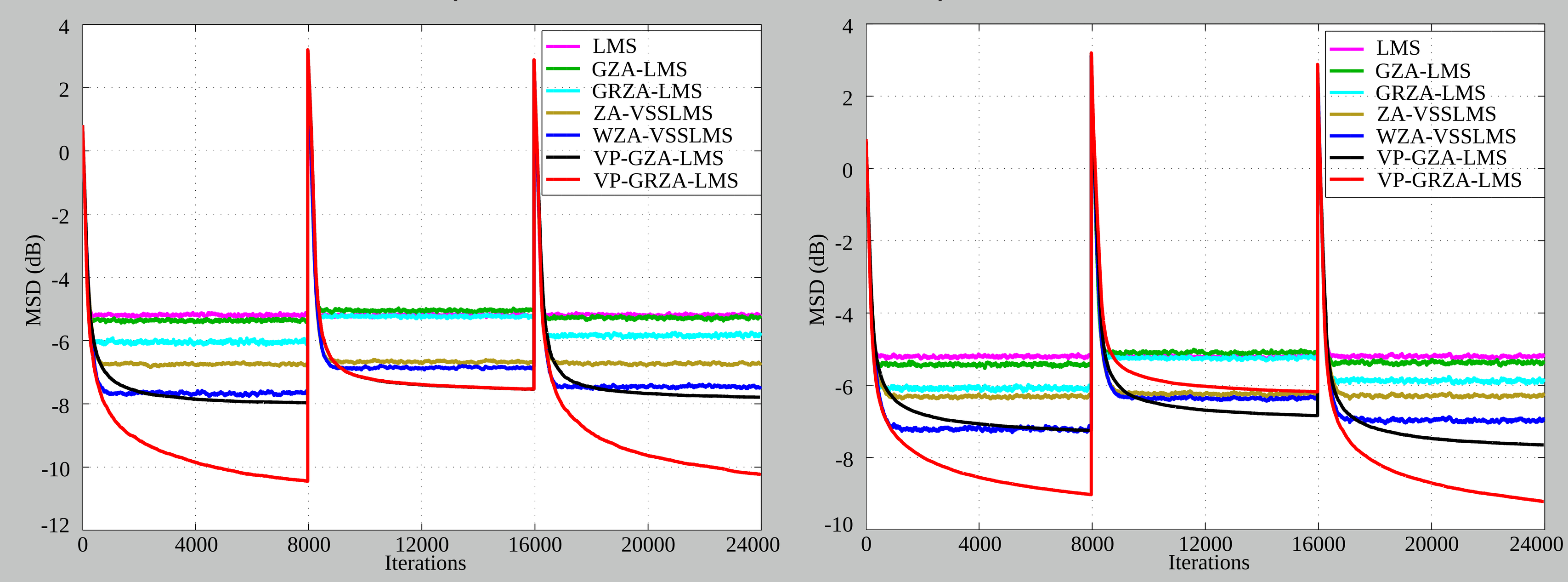

Figure: MSD learning curves (left: white input; right: non-Gaussian colored input).

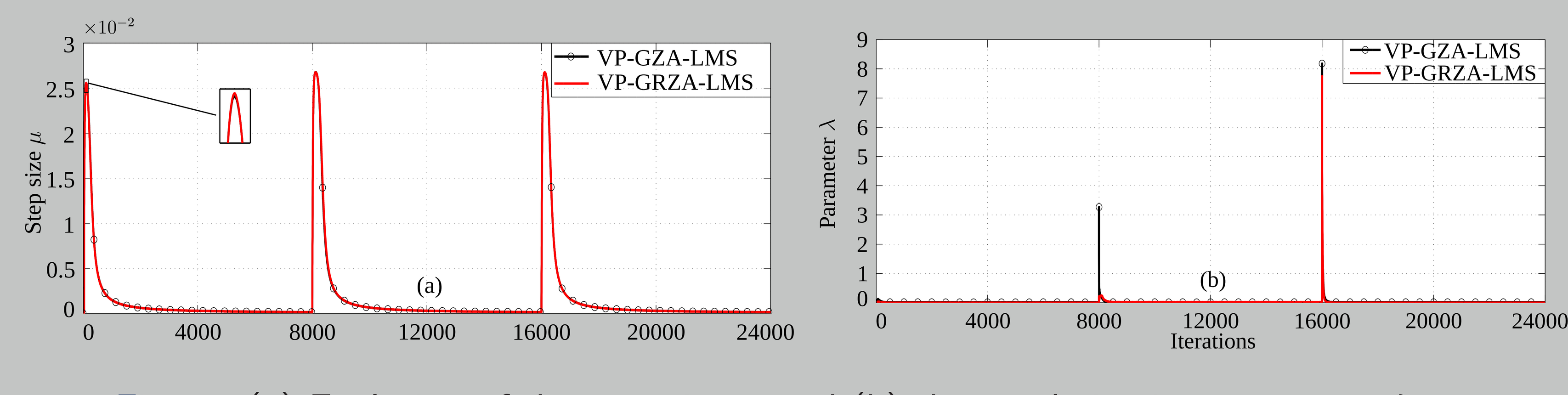

Figure: (a) Evolution of the step size μ and (b) the regularization parameter λ.

 Contact: danqijin@mail.nwpu.edu.cn; dr.jie.chen@ieee.org