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, u_n\} related via the linear model

\[ d_n = w^* + u_n + z_n \]

with \( w^* \in \mathbb{R}^L \) and \( u_n \in \mathbb{R}^L \).

▸ To determine \( w^* \), consider the MSE criterion with \( \ell_2 \)-norm regularization:

\[ w_{GZA} = \arg \min_{w} \frac{1}{2} \mathbb{E}[\|d_n - w^T u_n\|^2] + \lambda \|w\|^2 \]  

where \( \lambda \) is a regularization parameter.

▸ GZA-LMS and GRZA-LMS algorithm:

\[ w_{GZA} = \arg \min_{w} \frac{1}{2} \mathbb{E}[\|d_n - w^T u_n\|^2] + \lambda \|w\|^2 \]

where \( d_n = w^* + u_n + \mu \beta_n u_n - \beta_n s_n \), \( \mu \) is the step size, and \( s_n \) is vector of form \( s_n \).

▸ Trade-off
  ▸ \( \mu \) controls the trade-off between convergence speed and steady-state performance.
  ▸ \( \rho \) controls the trade-off between sparsity degree and estimation bias. It is therefore crucial to adaptively adjust \( \mu \) and \( \rho \).

Transient behavior model of GRZA-LMS

▸ Define the weight error vector and its covariance matrix by

\[ \tilde{w}_n = w_n - w^* \]  

with

\[ Q_n = \mathbb{E}[\tilde{w}_n \tilde{w}_n^T] \]

The recursion of \( \tilde{w}_n \) writes

\[ \tilde{w}_{n+1} = \tilde{w}_n + \mu \tilde{u}_n z_n - \mu \tilde{u}_n^T u_n - \rho \tilde{\beta} \circ s_n \]

▸ Assumptions:
  ▸ A1: The weight error vector \( \tilde{w}_n \) is statistically independent of \( u_n \).
  ▸ A2: The input regressor \( u_n \) is a zero-mean white signal with covariance matrix \( R_u = \sigma_u^2 I \).
  ▸ A2': The input regressor \( u_n \) is Gaussian distributed.

▸ With the independence assumption A1, we have:

\[ \text{MSE: } \mathbb{E}[\tilde{w}_n^2] = \sigma_w^2 + \text{trace}(R_u Q_n) \]

By utilizing white input assumption A2, we have:

\[ \text{MSE: } \mathbb{E}[\tilde{w}_n^2] = \sigma_w^2 \text{trace}(R_u Q_n) = \sigma_w^2 \mathbb{E}[\tilde{w}_n^2] \rightarrow \text{MSE} \]

Under assumptions A1 and A2: \( \min \text{MSE} \iff \min \text{MSD} \)

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

\[ \text{trace}(Q_{n+1}) = \text{trace}(Q_n) + \mu^2 \gamma + \rho^2 h + 2 \mu \rho - 2 \mu_1 - 2 \rho_2 \]

with

\[ g = \sigma_w^2 \text{trace}(R_u) + \mathbb{E}[\tilde{w}_n^2] - \text{trace}(R_u) \]  

\[ h = \mathbb{E}\{ (\beta_n \circ s_n) (\beta_n \circ s_n) \}, \quad \ell = \mathbb{E}\{ \tilde{w}_n^2 (\beta_n \circ s_n) \} \]

\[ r_1 = \mathbb{E}\{ \tilde{w}_n^2 u_n^T u_n \}, \quad r_2 = \mathbb{E}\{ (\beta_n \circ s_n) \tilde{w}_n^T u_n \} \]

▸ How to derive an adaptive parameters adjustment strategy?

Parameter design using transient behavior model

▸ Given the MSD \( \zeta_n \), at time instant \( n \), we determine the parameters \( \{\mu_n, \rho_n\} \) that minimize the MSD \( \zeta_n \):

\[ \{\mu_n, \rho_n\} = \arg \min_{\mu, \rho} \zeta_{n+1} \]

Using the recursion of \( \text{trace}(Q_{n+1}) \), we have:

\[ \{\mu_n, \rho_n\} = \arg \min_{\mu, \rho} \text{trace}(Q_n) + \mu^2 \gamma + \rho^2 h + 2 \mu \rho - 2 \mu_1 - 2 \rho_2 \]

Equivalently, in matrix form:

\[ \zeta_{n+1} = [\mu \rho] H [\mu \rho]^T - 2 [r_1, r_2] [\mu \rho]^T + \zeta_n \]

where \( H = \begin{bmatrix} g & \ell \\ \ell & h \end{bmatrix} \), and \( H \) is a positive semidefinite matrix.

▸ Solution:

\[ \{\mu_n, \rho_n\}^T = H^{-1} [r_1, r_2]^T \]

▸ Take the adopted signal configurations:
  ▸ \( g, h, \ell, r_1, r_2 \)
  ▸ Impose nonnegative constraints as well as temporal smoothing for \( \mu_n \) and \( \rho_n \).

Simulation results

Consider non-stationary system identification scenarios:

▸ System parameter vectors:

\[ u^*_1 = [0.8, 0.5, 0.3, 0.2, 0.1, 0.05, -0.05, -0.1, -0.2, -0.3, -0.5, 0.0, 0.25, 0.5, 0.25, -0.25, -0.5]^T \]

\[ u^*_2 = [0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.1, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9, -0.9^T] \]

\[ u^*_3 = [1.2, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.2, 0.1, 0.1, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9, -1.2^T] \]

At time instant \( n = 1, 5000 \) and \( 10000 \), we set the system parameter vector to \( u^*_1, u^*_2, \) and \( u^*_3 \), respectively.

▸ Input signal:
  ▸ Experiment 1: Zero-mean white Gaussian with \( \sigma^2 = 1 \)  
  ▸ 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

\[ 0.5 N(0 \cdot \sigma_v, \sigma_v^2) + 0.5 N(-a \cdot \sigma_v, \sigma_v^2) \]

▸ Additive noise: \( z_n \) was zero-mean.i.i.d. Gaussian with \( \sigma_z^2 = 0.01 \).

▸ Parameters: \( L = 35 \), group size \( |g| = 5 \), \( \varepsilon = 0.1 \).

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).