



HAL
open science

A Modified Non-Negative LMS Algorithm and its Stochastic Behavior Analysis

Jie Chen, Cédric Richard, José C. M. Bermudez, Paul Honeine

► **To cite this version:**

Jie Chen, Cédric Richard, José C. M. Bermudez, Paul Honeine. A Modified Non-Negative LMS Algorithm and its Stochastic Behavior Analysis. Proc. 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2011, Pacific Grove (CA), USA, United States. pp.542-546, 10.1109/ACSSC.2011.6190060 . hal-01966036

HAL Id: hal-01966036

<https://hal.science/hal-01966036>

Submitted on 27 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A modified non-negative LMS algorithm and its stochastic behavior analysis

Jie Chen^(1,2), Cédric Richard⁽¹⁾, Jose Bermudez⁽³⁾, Paul Honeine⁽²⁾

⁽¹⁾Université de Nice Sophia-Antipolis, France

⁽²⁾Université de Technologie de Troyes, France

⁽³⁾Department of Electrical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil

I. INTRODUCTION

Optimizing an objective function subject to constraints is very common in statistical signal and image processing. Constraints represent prior knowledge that may lead to proper solutions, or may reduce the feasible space of solutions when the problem is ill-conditioned. Non-negativity is probably one of the most commonly used constraints. It is often imposed on the parameters to estimate in order to avoid physically absurd and uninterpretable results. Non-negative least-square problems (NNLS) have been addressed in applications ranging from image deblurring [1] to impulse response estimation [9]. Non-negative matrix factorization (NMF) [8], which is closely related to blind deconvolution problems, have found direct application in hyperspectral imaging [2].

A variety of methods have been proposed in the literature to tackle the NNLS problem. Active set algorithm of Lawson and Hanson [7] is a batch resolution technique for NNLS problems, which has become a standard among the most frequently used methods. Projected gradient methods [3], [10] are based on successive projections on the feasible region. Multiplicative algorithms are very popular to solve NMF problems [6]. All these algorithms are however based on batch processing, which is not suitable for online system identification problems. In [4], [5], online system identification subject to non-negativity constraints on the parameters to estimate was addressed. A LMS-type algorithm, called non-negative LMS algorithm (NN-LMS) was proposed to compute solutions adaptively. It relies on stochastic gradient descent combined with a fixed point iteration scheme to converge to a solution that satisfies Karush-Kuhn-Tucker conditions.

During convergence of NN-LMS algorithm, it has been observed that weights may have different convergence rates and accuracy, especially those in the active set since they become smaller and finally tend to zero as iterations proceed. In this paper we introduce a modified NN-LMS algorithm in order to alleviate these unbalanced convergence rates. We also proposed analytical models to characterize the stochastic behavior of this algorithm.

II. MODIFIED NN-LMS ALGORITHM

Consider an unknown system, only characterized by a set of real-valued discrete-time responses to known stationary inputs.

The problem is to derive a transversal filter model

$$y(n) = \boldsymbol{\alpha}^\top \mathbf{x}(n) + z(n), \quad (1)$$

with $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^\top$ the vector of model parameters, and $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-N+1)]^\top$ the observed data vector. The input signal $x(n)$ and the desired output signal $y(n)$ are assumed zero-mean stationary. Sequence $z(n)$ represents measurement noise and modeling errors. Due to the inherent physical characteristics of systems under investigation, the problem of identifying the optimum model can be formalized as follows

$$\begin{aligned} \boldsymbol{\alpha}^\circ &= \arg \min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) \\ &\text{subject to } \alpha_i \geq 0, \quad \forall i, \end{aligned} \quad (2)$$

with $J(\boldsymbol{\alpha}) = E\{[y(n) - \boldsymbol{\alpha}^\top \mathbf{x}(n)]^2\}$, and $\boldsymbol{\alpha}^\circ$ the solution of the constrained optimization problem. The KKT conditions at the optimum $\boldsymbol{\alpha}^\circ$ can be combined into the following expression [4], [5]

$$\alpha_i^\circ [-\nabla_{\alpha} J(\boldsymbol{\alpha}^\circ)]_i = 0, \quad (3)$$

where the extra minus sign is just used to make a gradient descent of $J(\boldsymbol{\alpha})$ apparent. Applying a fixed-point iteration strategy and using usual stochastic gradient approximations, we obtain the NN-LMS algorithm [4], [5]

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \eta e(n) \mathbf{D}_\alpha(n) \mathbf{x}(n) \quad (4)$$

where η is a positive step size that must be tuned to construct a contraction scheme and to control the convergence rate. $\mathbf{D}_\alpha(n)$ stands for the diagonal matrix with diagonal entries given by $\alpha(n)$, and $e(n) = y(n) - \boldsymbol{\alpha}^\top(n) \mathbf{x}(n)$. It is interesting to note that the $\alpha_i(n)$'s in the second term of the r.h.s of expression (4) are factors that modify the step size along each axis. This results in different convergence rates and accuracy for each component of $\boldsymbol{\alpha}(n)$, especially for the weights corresponding to the active set as they become smaller through iterations and finally tend to zero. In order to compensate this unbalance of convergence rates, we now introduce the modified NN-LMS algorithm. Considering KKT condition (3), we can equivalently write

$$[\alpha_i^\circ]^\gamma [-\nabla_{\alpha} J(\boldsymbol{\alpha}^\circ)]_i = 0, \quad (5)$$

with γ a positive real number. Implementing the fixed-point iteration strategy with equation (5) and using instantaneous

estimates for the gradient leads us to the following algorithm

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \eta e(n) \mathbf{D}_x(n) \boldsymbol{\alpha}^\gamma(n) \quad (6)$$

Without ambiguity $\boldsymbol{\alpha}^\gamma(n)$ denotes the exponential value γ applied to each component of $\boldsymbol{\alpha}(n)$. We recommend to specify parameter γ in the form $\gamma = \frac{p}{q}$ where p and q are both odd numbers, and $0 < p \leq q$. The oddness of p and q preserves the sign of $\alpha_i(n)$ at each iteration. Although it can be shown that weights are always non-negative thanks to an appropriate step size, the fixed-point 0 can be reached by the left side 0^- with less strict step size selection. See [4] for more details with NN-LMS. Just as the gamma correction in image processing, an exponential value $0 < \gamma < 1$ reduces the dynamic range of parameter vector $\boldsymbol{\alpha}(n)$. Each component $\alpha_i(n)$ will be closer to 1 no matter it is larger or smaller than 1.

III. ALGORITHM BEHAVIOR MODELING

Let us denote by $\boldsymbol{\alpha}^*$ the solution of the unconstrained least-mean-square problem $\min_{\boldsymbol{\alpha}} E\{[y(n) - \boldsymbol{\alpha}^\top \mathbf{x}(n)]^2\}$. Defining the weight-error vector by $\mathbf{v}(n) = \boldsymbol{\alpha}(n) - \boldsymbol{\alpha}^*$, the update equation (4) can be written as

$$\mathbf{v}(n+1) = \mathbf{v}(n) + \eta e(n) \mathbf{D}_x(n) (\mathbf{v}(n) + \boldsymbol{\alpha}^*)^\gamma. \quad (7)$$

It is assumed that $z(n)$ is stationary, zero-mean and statistically independent of any other signal, which implies that $E\{z(n) \mathbf{D}_x(n)\} = 0$. After some calculation and approximation, the expected value of the weight vector error can be expressed recursively by

$$E\{\mathbf{v}(n+1)\} = E\{\mathbf{v}(n)\} - \eta \mathbf{D}_r(n) \mathbf{R}_x E\{\mathbf{v}(n)\} - \eta \mathbf{D}_s(n) \text{diag}\{\mathbf{R}_x E\{\mathbf{v}(n)\} E\{\mathbf{v}^\top(n)\}\}. \quad (8)$$

where $\mathbf{D}_r(n)$ and $\mathbf{D}_s(n)$ are diagonal matrices with diagonal entries defined by

$$r_{ii}(n) = (E\{v_i(n)\} + \alpha_i^*)^\gamma - \gamma (E\{v_i(n)\} + \alpha_i^*)^{\gamma-1} E\{v_i(n)\} \\ s_{ii}(n) = \gamma (E\{v_i(n)\} + \alpha_i^*)^{\gamma-1}$$

We now briefly present the model for the stochastic behavior of the second-order moments of the adaptive weights. As the mean-square estimation error is expressed

$$E\{e^2(n)\} = \sigma_z^2 + \text{trace}\{\mathbf{R}_x \mathbf{K}(n)\}, \quad (9)$$

we thus derive a recursion for $\mathbf{K}(n) = E\{\mathbf{v}(n) \mathbf{v}^\top(n)\}$ starting from the weight error update equation (7). Premultiplying equation (7) by its transpose, taking the expected value, and using the statistical properties of $z(n)$ leads us to a recursion of the form

$$\mathbf{K}(n+1) = \mathbf{K}(n) - \eta \mathbf{P}_1(n) \mathbf{K}(n) - \eta \mathbf{K}(n) \mathbf{P}_1^\top(n) + \eta^2 \sigma_z^2 \mathbf{P}_2(n) + \eta^2 \sigma_z^2 (\mathbf{P}_3(n) + \mathbf{P}_3^\top(n)) + \eta^2 \sigma_z^2 \mathbf{P}_4(n) - \eta (\mathbf{P}_5(n) + \mathbf{P}_5^\top(n)) + \eta^2 \mathbf{P}_6(n) + \eta^2 (\mathbf{P}_7(n) + \mathbf{P}_7^\top(n)) + \eta^2 \mathbf{P}_8(n) \quad (10)$$

where \mathbf{P}_1 to \mathbf{P}_8 are functions of $\mathbf{K}(n)$, $E\{\mathbf{v}(n)\}$ and \mathbf{R}_x . Due to space constraints in the abstract, their expressions will be given in the final paper.

We now present simulation examples to illustrate the accuracy of the first and second-order models derived in this paper. We consider an example where the correlated inputs $x(n)$ are generated by the AR(1) model $x(n) = \frac{1}{2} x(n-1) + w(n)$. The noise $w(n)$ is i.i.d. and drawn from a zero-mean Gaussian distribution so that $\sigma_x^2 = 1$. Noise $z(n)$ is drawn from a Gaussian distribution with zero mean and $\sigma_z^2 = 0.1$. The impulse response $\boldsymbol{\alpha}^*$ is given by

$$\boldsymbol{\alpha}^* = [0.8 \ 0.6 \ 0.5 \ 0.4 \ 0.3 \ 0.2 \ 0.1 \ -0.1 \ -0.3 \ -0.6]^\top$$

Note that it contains negative entries in order to better test the behavior of the algorithm. The mean value $E\{\alpha_i(n)\}$ of each coefficient is shown in Figure 1 and the excess mean square error $J_{emse} = \text{trace}\{\mathbf{R}_x \mathbf{K}(n)\}$ is shown in Figure 2. The simulation curves (blue line) were obtained from Monte Carlo simulation averaged over 100 realizations. The theoretical curves (red line) were obtained from the models. One can notice that all the curves are perfectly superimposed. We then compare NN-LMS and modified NN-LMS with adjusted parameters so that they reach the same steady state error. The results are shown in the Figure 3. Unlike NN-LMS, whose small weights converge more slowly, our modified NN-LMS has more balanced first-order convergence rate. It also has better second-order convergence performance than NN-LMS.

REFERENCES

- [1] F. Benvenuto, R. Zanella, L. Zanni, and M. Bertero. Nonnegative least-squares image deblurring: improved gradient projection approaches. *Inverse Problems*, 26(1), 2010.
- [2] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics and Data Analysis*, 52(1):155–173, 2007.
- [3] P. H. Calamai and J. J. Moré. Projected gradient methods for linearly constrained problems. *Mathematical Programming*, 39(1):93–116, 1987.
- [4] J. Chen, C. Richard, J.-C. M. Bermudez, and P. Honeine. Non-negative least-mean-square algorithm. *IEEE Transactions on Signal Processing*, 2010 (submitted).
- [5] J. Chen, C. Richard, P. Honeine, H. Lantéri, and C. Theys. System identification under non-negativity constraints. In *Proc. EUSIPCO'10*, Aalborg, Denmark, 2010.
- [6] H. Lantéri, M. Roche, O. Cuevas, and C. Aime. A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints. *Signal Processing*, 81(5):945–974, 2001.
- [7] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, 1995.
- [8] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [9] Y. Lin and D. D. Lee. Bayesian regularization and nonnegative deconvolution for room impulse response estimation. *IEEE Transactions on Signal Processing*, 54(3):839–847, Mar. 2006.
- [10] S. Theodoridis, K. Slavakis, and I. Yamada. Adaptive learning in a world of projections. *Signal Processing Magazine, IEEE*, 28(1):97–123, 2011.

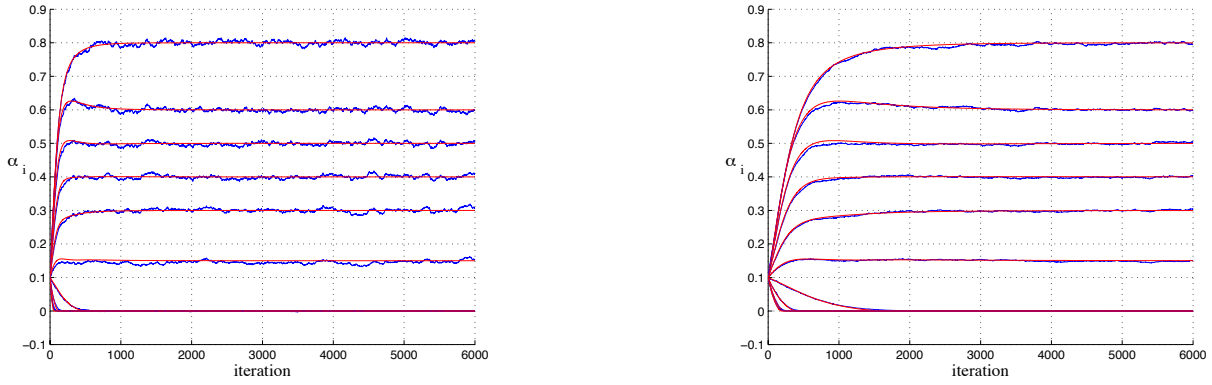


Fig. 1. Convergence of the coefficients $\alpha_i(n)$ for the modified NN-LMS, in the case of correlated input $x(n)$. Two different step sizes are considered: $\eta = 10^{-2}$ on the left, and $\eta = 3 \cdot 10^{-3}$ on the right. The theoretical curves (red line) obtained from the model (8) and simulation curves (blue line) are perfectly superimposed.

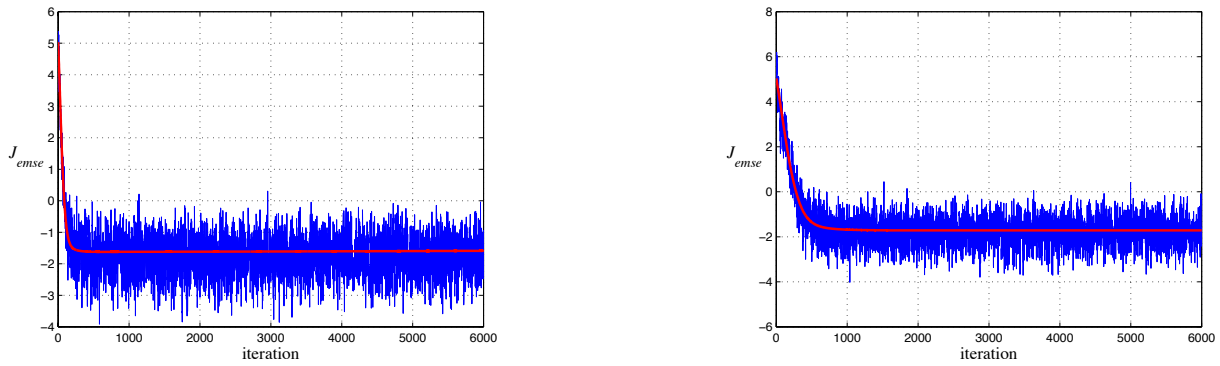


Fig. 2. Corresponding convergence of $J_{emse}(n)$ of the modified NN-LMS in these experiments. The theoretical curves (red line) were obtained with the model (10).

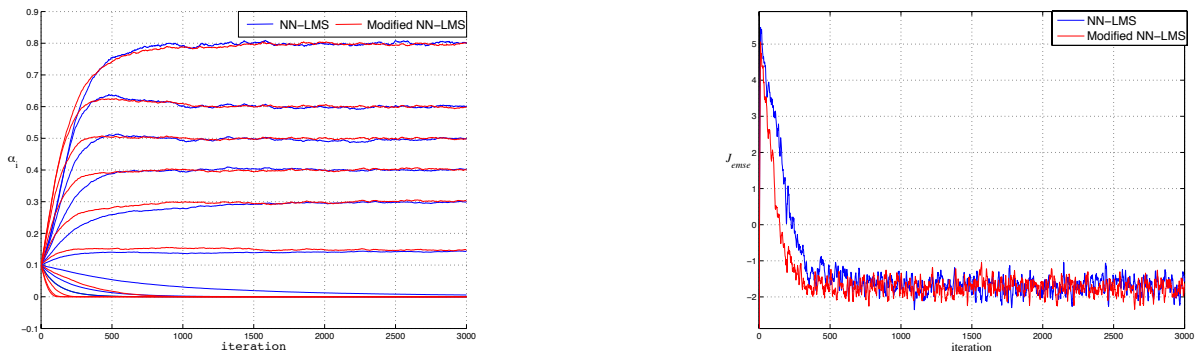


Fig. 3. Comparison of NN-LMS and modified NN-LMS. Left: first-order performance. The modified algorithm has a more balanced weight convergence rate. Right: Excess mean square error.