Signal Processing 128 (2016) 131–141

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# Reweighted nonnegative least-mean-square algorithm

Jie Chen [a,*], Cédric Richard [b], José Carlos M. Bermudez [c]

[a] Center of Intelligent Acoustics and Immersive Communications (CIAIC), School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China

[b] The Université de Nice Sophia-Antipolis, UMR CNRS 7293, Observatoire de la Côte d'azur, Laboratoire Lagrange, Parc Valrose, 06102 Nice, France

[c] The Department of Electrical Engineering, Federal University of Santa Catarina, 88040-900 Florianópolis, SC, Brazil

ARTICLE INFO

*Article history:*
Received 10 October 2015
Received in revised form
22 February 2016
Accepted 26 March 2016
Available online 1 April 2016

*Keywords:*
Online system identification
Nonnegativity constraints
Behavior analysis
Sparse system identification

ABSTRACT

Statistical inference subject to nonnegativity constraints is a frequently occurring problem in learning problems. The nonnegative least-mean-square (NNLMS) algorithm was derived to address such problems in an online way. This algorithm builds on a fixed-point iteration strategy driven by the Karush–Kuhn–Tucker conditions. It was shown to provide low variance estimates, but it however suffers from unbalanced convergence rates of these estimates. In this paper, we address this problem by introducing a variant of the NNLMS algorithm. We provide a theoretical analysis of its behavior in terms of transient learning curve, steady-state and tracking performance. We also introduce an extension of the algorithm for online sparse system identification. Monte-Carlo simulations are conducted to illustrate the performance of the algorithm and to validate the theoretical results.



## 1. Introduction

Online learning aims at determining a mapping from a dataset to the corresponding labels when the data are available in a sequential fashion. In particular, algorithms such as the Least-Mean-Square (LMS) and the Recursive Least-Square (RLS) algorithms minimize the mean square-error cost function in an online manner based on input/output measurement sequences [1,2]. In practice, rather than leaving the parameters to be estimated totally free and relying on data, it is often desirable to introduce some constraints on the parameter space. These constraints are usually introduced to impose some specific structures, or to incorporate prior knowledge, so as to improve the estimation accuracy and the interpretability of results in learning systems [3,4]. The nonnegativity constraint is one of the most frequently used constraints among several popular ones [5]. It can be imposed to avoid physically unreasonable solutions and to comply with physical characteristics. For example, quantities such as intensities [6,7], chemical concentrations [8], and material fractions of abundance [9] must naturally fulfill nonnegativity constraints. Nonnegativity constraints may also enhance the physical interpretability of some analyzed results. For instance, Nonnegative Matrix Factorization leads to more meaningful image decompositions than Principle Component Analysis (PCA) [10,11]. PCA and neural networks can also be conducted subject to nonnegativity constraints in order to enhance result interpretability [12,13]. Finally, there are important problems in signal processing that can be cast as optimization problems under nonnegativity constraints [14]. Other applications of learning systems related to nonnegativity constraints can be found in [15–17,5,18–20].

Nonnegativity constrained problems can be solved in a batch mode via active set methods [21,22], gradient projection methods [23,24], and multiplicative methods [25], to cite a few. Online system identification methods subject to nonnegativity constraints can also be of great interest in applications that require to adaptively identify a dynamic system. An LMS-type algorithm, called the non-negative least-mean-square (NNLMS) algorithm, was proposed in [26] to address the least-mean-square problem under nonnegativity constraints. It was derived based on a stochastic gradient descent approach combined with a fixed-point iteration strategy that ensures convergence toward a solution satisfying the Karush–Kuhn–Tucker (KKT) conditions. In [27], several variants of the NNLMS were derived to improve its convergence behavior in some sense. The steady-state performance of these algorithms was analyzed in [28]. It was observed that one limitation of the NNLMS algorithm is that the filter coefficients suffer from unbalanced convergence rates. In particular, convergence of small coefficients in the active set (the set of zero-valued optimum weights), that is, those tending to zero at steady-state, progressively slows down with time and almost stalls when approaching the steady-state (see [27] and also Fig. 3(a)). Another limitation of the NNLMS algorithm is its vulnerability to the occurrence of a large coefficient value spread. A large spread of coefficient values in NNLMS lead to a large spread of the weight updates, increasing the coefficient variances and complicating the choice of an adequate step-size.

* Corresponding author.
*E-mail addresses:* dr.jie.chen@ieee.org (J. Chen), cedric.richard@unice.fr (C. Richard), j.bermudez@ieee.org (J.C.M. Bermudez).

http://dx.doi.org/10.1016/j.sigpro.2016.03.017
0165-1684/

The Exponential NNLMS algorithm was proposed in [27] to alleviate the first limitation. This algorithm applies a Gamma scaling function to each component of the NNLMS update. Although this NNLMS variant leads to more balanced coefficient convergence rates, it does not completely solve large coefficient update spread problem, as the scaling function is still unbounded on the coefficient values. Moreover, the exponential scaling operation tends to be computationally expensive for real-time implementations requiring a large number of coefficients. Thus, it is of interest to investigate alternative algorithms that may simultaneously address these two NNLMS limitations.

In this paper, we propose a variant of the NNLMS algorithm that balances more efficiently the convergence rate of the different filter coefficients. The entries of the gradient correction term are reweighted by a bounded differentiable sign-like function at each time instant. This gives the filter coefficients balanced convergence rates and largely reduces the sensitivity of the algorithm to large coefficient spreads. The stochastic behavior of the algorithm is then studied in detail. A statistical analysis of its transient and steady-state behavior leads to analytical models that are able to accurately predict the algorithm performance. In particular, contrary to a previous analysis [27] the algorithm tracking behavior is studied using a nonstationarity model that allows for a bounded covariance matrix of the optimal solution, a more practical scenario. The accuracy of the derived analytical models is verified through Monte Carlo simulations. Finally, the applicability of the proposed algorithm to problems whose definition does not specify nonnegativity constraints on the coefficients is illustrated through an example of identification of sparse system responses.

The rest of this paper is organized as follows. Section 2 reviews the problem of system identification under nonnegativity constraints and briefly introduces the NNLMS algorithm. Section 3 motivates and introduces the new variant of the NNLMS algorithm. In Section 4, the behavior in the mean and mean-square-error sense, and the tracking performance of this algorithm are studied. Section 5 provides simulation results to illustrate the properties of the algorithm and the accuracy of the theoretical analysis. Section 6 concludes the paper.

In this paper normal font letters ($x$) denote scalars, boldface small letters ($\boldsymbol{x}$) denote vectors, boldface capital letters ($\mathbf{X}$) denote matrices with $\boldsymbol{I}$ being the identity matrix. All vectors are column vectors. The superscript $(\cdot)^\top$ represents the transpose of a matrix or a vector, $\mathrm{tr}\{\cdot\}$ denotes trace of a matrix, and $E\{\cdot\}$ denotes statistical expectation. Either $\boldsymbol{D}_{\boldsymbol{x}}$ or $\boldsymbol{D}\{x_1, \ldots, x_N\}$ denote a diagonal matrix whose main diagonal is the vector $\boldsymbol{x} = [x_1, \ldots, x_N]^\top$. The operator $\mathrm{diag}\{\cdot\}$ forms a column vector with the main diagonal entries of its matrix argument.

## 2. Online system identification subject to nonnegativity constraints

Consider an unknown system with input-output relation characterized by the linear model:

$$y(n) = \boldsymbol{\alpha}^{\star\top}\boldsymbol{x}(n) + z(n) \tag{1}$$

with $\boldsymbol{\alpha}^\star = [\alpha_1^\star, \alpha_2^\star, \ldots, \alpha_N^\star]^\top$ an unknown parameter vector, and $\boldsymbol{x}(n) = [x(n), x(n-1), \ldots, x(n-N+1)]^\top$ the vector of regressors with positive definite correlation matrix $\boldsymbol{R}_{\boldsymbol{x}} > 0$. The input signal $x(n)$ and the desired output signal $y(n)$ are assumed zero-mean stationary. The modeling error $z(n)$ is assumed zero-mean stationary, independent and identically distributed (i.i.d.) with variance $\sigma_z^2$, and independent of any other signal. We seek to identify this system by minimizing the following constrained mean-square error criterion:

$$\boldsymbol{\alpha}^o = \arg\min_{\alpha} J(\boldsymbol{\alpha}) \quad \text{subject to } \alpha_i \geq 0, \quad \forall\, i \tag{2}$$

where the nonnegativity of the estimated coefficients is imposed by inherent physical characteristics of the system, and $J(\boldsymbol{\alpha})$ is the mean-square error criterion

$$J(\boldsymbol{\alpha}) = E\{[y(n) - \boldsymbol{\alpha}^\top\boldsymbol{x}(n)]^2\} \tag{3}$$

and $\boldsymbol{\alpha}^o$ is the solution of the constrained optimization problem (2). The Lagrange function associated with this problem is given by $L(\boldsymbol{\alpha}, \lambda) = J(\boldsymbol{\alpha}) - \lambda^\top\boldsymbol{\alpha}$, with $\lambda$ being the vector of nonnegative Lagrange multipliers. The KKT conditions for (2) at the optimum $\boldsymbol{\alpha}^o$ can be combined into the following expression [29,26]

$$\alpha_i^o[-\nabla_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}^o)]_i = 0 \tag{4}$$

where $\nabla_{\boldsymbol{\alpha}}$ stands for the gradient operator with respect to $\boldsymbol{\alpha}$. Implementing a fixed-point strategy with (4) leads to the component-wise gradient descent updating rule [26]

$$\alpha_i(n+1) = \alpha_i(n) + \eta(n) f_i(\boldsymbol{\alpha}(n))\alpha_i(n)[-\nabla_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}(n))]_i \tag{5}$$

where $\eta(n)$ is the positive step size that controls the convergence rate, $f_i(\boldsymbol{\alpha}(n))$ is a nonnegative scalar function of vector $\boldsymbol{\alpha}$ that weights its $i$th component $\alpha_i$ in the update term. Selecting different functions $f_i(\boldsymbol{\alpha}(n))$ leads to different adaptive algorithms. Particularly, making $f_i(\boldsymbol{\alpha}(n)) = 1$, using stochastic gradient approximations as done in deriving the LMS algorithm, and rewriting the update equation in vectorial form, we obtain the NNLMS algorithm [26]:

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \eta(n)\boldsymbol{D}_{\boldsymbol{\alpha}}(n)\boldsymbol{x}(n)e(n) \tag{6}$$

where $\boldsymbol{D}_{\boldsymbol{\alpha}}(n)$ is the diagonal matrix in which the elements of $\boldsymbol{\alpha}(n)$ compose the main diagonal, and $e(n)$ is the estimation error at time instant $n$:

$$e(n) = y(n) - \boldsymbol{\alpha}^\top(n)\boldsymbol{x}(n). \tag{7}$$

This iteration is similar in some sense to the expectation maximization (EM) algorithm [30]. The algorithm requires to be initialized with positive values. Suppose that $\boldsymbol{\alpha}(n)$ is nonnegative at time $n$. If the step size satisfies

$$0 < \eta(n) \leq \min_i \frac{1}{-e(n)x_i(n)}, \tag{8}$$

for all $i \in \{j: e(n)x_j(n) < 0\}$, the nonnegativity constraint is satisfied at time $n+1$ with (6). If $e(n)x_i(n) \geq 0$, there is no restriction on the step size to guarantee the nonnegativity constraint. Convergence of the NNLMS algorithm was analyzed in [26]. Its steady-state excess mean-square error (EMSE) was studied in [28].

## 3. Motivating facts and the algorithm

### 3.1. Motivation

The weight update in (6) corresponds to the classical stochastic gradient LMS update with its $i$th component scaled by $\alpha_i(n)$. The mean value of the update vector $\boldsymbol{D}_{\boldsymbol{\alpha}}(n)\boldsymbol{x}(n)e(n)$ is thus no longer in the direction of the gradient of $J(\boldsymbol{\alpha})$, as is the case for LMS. On the other hand, it is exactly this scaling by $\alpha_i(n)$ that enables the corrections $\alpha_i(n)x_i(n)e(n)$ to reduce gradually to zero for coefficients $\alpha_i(n)$ tending to zero, which leads to low-variance estimates for these coefficients.[1] If a coefficient $\alpha_k(n)$ that approaches zero turns

[1] The update for these weights will be necessarily small in amplitude as $\alpha_i(n)$ tends to zero, thus leading to a small variance of the adaptive coefficient.