Signal Processing 128 (2016) 340–349

Contents lists available at ScienceDirect

Signal Processing

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

A class of improved least sum of exponentials algorithms

Shiyuan Wang [a,*], Yunfei Zheng [a], Shukai Duan [a], Lidan Wang [a], Chi K. Tse [b]

[a] The School of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
[b] The Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 6 December 2015
Received in revised form
28 March 2016
Accepted 4 May 2016
Available online 6 May 2016

Keywords:
Least sum of exponentials
Energy conservation relation
Variable scaling factor
Kernel method

ABSTRACT

A class of *improved least sum of exponentials* (ILSE) algorithms is proposed by incorporating a scaling factor into the cost function of LSE in this paper. The even-order moment information regarding error is influenced by the scaling factor. However, the ILSE algorithm based on a fixed scaling factor can only provide a tradeoff between the convergence rate and steady-state excess-mean-square error (EMSE). Therefore, a *variable scaling factor ILSE* (VS-ILSE) algorithm is also proposed to improve the convergence rate and steady-state EMSE, simultaneously. To facilitate analysis, the energy conservation relation of ILSE is established, providing a sufficient condition for mean square convergence and a theoretical value of the steady-state EMSE. In addition, the kernel extensions of ILSE and VS-ILSE are further developed for performance improvement. Simulation results illustrate the theoretical analysis and the excellent performance of the proposed methods.

1. Introduction

Adaptive filters (AF) have been widely applied in system identification, channel equalization, echo cancelation and many other signal processing applications [1–3]. According to the number of used data, AFs can be divided into three categories, i.e., the least mean-square algorithms (LMS), the affine projection algorithms (APA), and the recursive least squares algorithms (RLS) [4,5]. Among these algorithms, LMS is the most popular filtering algorithm due to its robustness, good tracking capability and simplicity [1,4]. Moreover, most other filters are developed on the basis of LMS.

In the LMS family, different p-powers of the error, i.e., $|e|^p$, are used as the cost function to realize the desirable performance. The sign algorithm (SA) [6] utilizes the error with $p=1$ to combat the impulsive interferences. For Gaussian interferences, $p=2$ is used to generate the traditional LMS. The least mean absolute third (LMAT) [7] with $p=3$ and least mean fourth (LMF) [8] with $p=4$ employ a high order power of the error to achieve a better tradeoff between the transient and steady-state performance. In addition, a linear combination of $p=2$ and $p=4$ is proposed in the least-mean mixed-norm (LMMN) algorithm [9] to further improve the performance of LMS and LMF. Another combination of $p=1$ and $p=2$ is used to construct the robust mixed-norm (RMN) algorithm [10], leading to improvement of the robustness to impulsive interferences. However, the LMMN and RMN only consider a linear combination of two kinds of errors with higher order errors neglected. Therefore, higher order powers of the error should be introduced to make full use of the information hidden in the error. To this end, the cost function based on the even moments of the error [11], namely the weighted even moment (WEM) algorithm, is used to achieve better performance, especially in the case of Gaussian noise. Moreover, WEM necessitates appropriate setting of the combination parameters, which may be infeasible in practice [12]. Therefore, to achieve a weighted sum of the even-order moments of the error as the cost function, the smoothed least mean p-power error (SLMP) algorithm [13] introduces a smoothing factor in the least mean p-power error (MPE) criterion [14] in order to take into account the influence of different moments of the error on the performance. The smoothing factor is very crucial for the performance of SLMP. Thus, an optimization method for determining the parameters is required in practice.

In addition, the logarithmic and exponential cost functions employ the high order power of the error using Taylor series expansion. The logarithmic cost function [15–17] intrinsically combines the higher and the lower order measures of the error into a single continuous update progress. The representative logarithmic algorithms include the least logarithmic absolute difference (LLAD) and the least mean logarithmic square (LMLS) algorithms [16]. Unfortunately, only when the error is small, the logarithmic algorithms can be regarded as a linear combination of different error powers. The exponential cost function [12,18,19] has been

* Corresponding author.
E-mail address: wangsy.swu@gmail.com (S. Wang).

http://dx.doi.org/10.1016/j.sigpro.2016.05.005
0165-1684/

proven to be a real linear combination of infinite number of the error powers. Nevertheless, the exponential algorithms only accelerate the convergence rate but with no significant improvement in the steady-state performance.

Inspired by the least sum of exponentials (LSE) [18] and the exponential-error least-mean fourth (EELMF) [20] methods, we propose a class of improved least sum of exponentials (ILSE) algorithms by introducing a scaling factor for performance improvement. The sufficient conditions for mean-square convergence and the theoretical value of steady-state EMSE for the ILSE are derived by utilizing the energy conservation relation. To further improve performance, an efficient variable scaling factor algorithm and kernel methods are incorporated into the ILSE, resulting in the variable scaling factor ILSE (VS-ILSE), kernel ILSE (KILSE), and the kernel variable scaling factor ILSE (KVS-ILSE) algorithm.

2. Improved least sum of exponentials algorithm

2.1. Review of the LSE algorithm

The cost function of LSE [18] is given by

$$J(e(i)) = \cosh[e(i)] = \frac{1}{2}\left(\exp[e(i)] + \exp[-e(i)]\right) \tag{1}$$

where $\cosh(\cdot)$ is the hyperbolic cosine function, and $e(i)$ denotes the prediction error that is the difference between the desired signal $d(i)$ and the actual output $y(i)$.

Let $\boldsymbol{w}(i-1)$ and $\boldsymbol{u}(i)$ denote the *row* weight vector and the *row* input vector, respectively. The output of filter is therefore $y(i) = \boldsymbol{w}(i-1)\boldsymbol{u}(i)^T$. The prediction error can be obtained by

$$e(i) = d(i) - y(i) = d(i) - \boldsymbol{w}(i-1)\boldsymbol{u}(i)^T. \tag{2}$$

Taking the Taylor series expansion of $J(e(i))$ in (1) at the point of zero gives

$$\begin{aligned} J(e(i)) &= \frac{1}{2}\left(\exp[e(i)] + \exp[-e(i)]\right) \\ &= 1 + \frac{1}{2!}e^2(i) + \frac{1}{4!}e^4(i) + \cdots = \sum_{k=0}^{+\infty} \frac{1}{2k!}e^{2k}(i) \end{aligned} \tag{3}$$

where k is the nonnegative integer.

Obviously, the cost function of LSE is a linear combination of all the errors with even moments. Therefore, more information hidden in the errors can be embodied in LSE. According to the steepest descent method, the weight update form of the LSE can be expressed by

$$\boldsymbol{w}(i) = \boldsymbol{w}(i-1) + \eta\ \sinh[e(i)]\boldsymbol{u}(i) \tag{4}$$

where $\sinh(\cdot)$ is the hyperbolic sine function and η denotes the step size.

2.2. Proposed ILSE algorithm

Inspired by [20], we introduce a scaling factor $\lambda > 0$ in (1) to construct a new cost function, i.e.,

$$J(e(i), \lambda) = \frac{1}{\lambda}\cosh[\lambda e(i)] = \frac{1}{2\lambda}\left(\exp[\lambda e(i)] + \exp[-\lambda e(i)]\right). \tag{5}$$

The Taylor series expansion of $J(e(i), \lambda)$ about zero can be expressed by

$$J(e(i), \lambda) = \frac{1}{2\lambda}\left(\exp[\lambda e(i)] + \exp[-\lambda e(i)]\right) = \sum_{k=0}^{+\infty} \lambda^{2k-1}\frac{1}{2k!}e^{2k}(i) \tag{6}$$

Fig. 1 shows the curves of cost functions $J(e(i), \lambda)$ and the corresponding gradient curves with different values of λ. It can be seen from Fig. 1 that, when $e(i)$ is larger, $J(e(i), \lambda)$ with larger λ has a larger gradient, resulting in a faster convergence rate; otherwise, $J(e(i), \lambda)$ with smaller λ has a flat gradient curve, leading to the improvement of filter stability. Hence, λ should have a larger value at the beginning of filtering, leading to a faster initial convergence rate. When the filter approaches its steady state, λ with smaller value will provide a smaller steady-state error. This process can be implemented using variable scale factor strategy.

Take the partial derivative of $J(e(i), \lambda)$ in (5) with respect to $\boldsymbol{w}(i-1)$ as follows:

$$\frac{\partial J(e(i), \lambda)}{\partial \boldsymbol{w}(i-1)} = -\sinh[\lambda e(i)]\boldsymbol{u}(i). \tag{7}$$

The weight update of ILSE is therefore obtained by the steepest descent method, i.e.,

$$\boldsymbol{w}(i) = \boldsymbol{w}(i-1) + \eta\ \sinh[\lambda e(i)]\boldsymbol{u}(i) = \boldsymbol{w}(i-1) + \eta f(e(i))\boldsymbol{u}(i) \tag{8}$$

where $f(e(i)) = \sinh[\lambda e(i)]$ denotes the nonlinear function of the error.

Remark 1. Compared with the Taylor series expansion of LSE in (3), the one of ILSE in (6) has an additional term λ^{2k-1}, which can be regarded as the weighted LSE cost function. The stochastic information of $e(i)$ is influenced by λ directly. The higher-order even

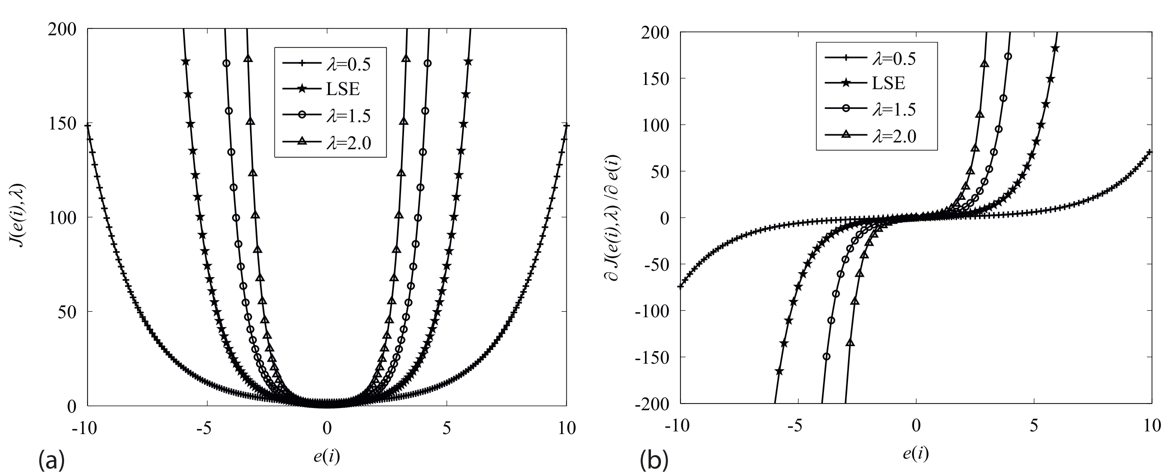

Fig. 1. Curves of cost functions $J(e(i), \lambda)$ and the corresponding gradient curves with different scaling factors: (a) $J(e(i), \lambda)$ versus $e(i)$; (b) gradient versus $e(i)$.