Digital Signal Processing 54 (2016) 35–45

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Stochastic analysis of the Least Mean Kurtosis algorithm for Gaussian inputs

Neil J. Bershad [a,*], José C.M. Bermudez [b,1]

[a] Department of Electrical and Computer Engineering, University of California, Irvine, CA, USA
[b] Department of Electrical and Electronic Engineering, Federal University of Santa Catarina, Trindade, Florianópolis, SC, 88040-970, Brazil

ARTICLE INFO

Article history:
Available online 6 April 2016

Keywords:
Adaptive filters
Analysis
LMK algorithm
Stochastic algorithms

ABSTRACT

The Least Mean Kurtosis (LMK) algorithm was initially proposed as an adaptive algorithm that is robust to the observation noise distribution. Good performances of this algorithm have been shown for non-Gaussian additive measurement noise. However, the complexity of the algorithm imposes difficulties for the development of a reasonably complete theoretical stochastic model for its behavior. The purpose of this paper is to contribute to the development of such a model. We study the stochastic behavior of Least Mean Kurtosis (LMK) algorithm for Gaussian inputs and for additive noises with even probability density functions. Deterministic recursions are derived for the adaptive weight error covariance matrix in a very novel manner, leading to a recursive model for the excess mean square error (EMSE) behavior that is shown to be accurate for Gaussian, uniform and binary noise distributions. The analysis results are then used to compare the performances of LMK with the least mean squares (LMS) and least mean fourth (LMF) algorithms under different circumstances.

1. Introduction

Stochastic gradient based algorithms are largely employed in real-time adaptive filtering for several applications in communications, control engineering, bioengineering and general signal processing [1–4]. The most popular among them is the Least-Mean Square (LMS) algorithm. It is very popular due to its simplicity. Analytical models are available for predicting its behavior under different input conditions, facilitating its design. On the other hand, stochastic gradient adaptive algorithms based on higher order moments of the error signal have been shown to outperform LMS in some important applications [5–10]. The practical use of such algorithms, however, has been largely restricted due to the lack of accurate analytical models to predict their behavior. The Least Mean Kurtosis (LMK) algorithm is one of such algorithms.

The LMK algorithm was initially introduced in [11]. The motivation was to obtain an algorithm robust to the noise distribution. The LMK algorithm as proposed in [11] seeks to minimize the negative of the error signal kurtosis. The kurtosis is related to the fourth order cumulant of the error [1]. Now the kurtosis of zero-mean Gaussian processes is equal to zero. Thus, a kurtosis-based cost function makes LMK algorithm convergence behavior independent of the noise statistics if the measurement noise is Gaussian [11]. However, if the measurement noise is not Gaussian, then the LMK algorithm will respond accordingly and can possibly perform better than an algorithm based on the second order statistics of the error. Moreover, given the linearity property of the cumulants, the LMK algorithm tends to decouple the costs due to the noise and due to the excess estimation error. This property is expected to make the algorithm performance more robust to the noise distribution and has raised interest in its application [12–21].

Though the desirable properties of the LMK algorithm rely on the properties of the kurtosis, its practical implementation is based on a stochastic approximation of the gradient of the cost function and requires a recursive estimation of the conditional mean squared estimation error[2] [11]. The combination of the stochastic approximation and the recursive estimation of the conditioned mean squared error lead to an algorithm whose actual behavior can deviate considerably from the kurtosis-based original concept. Thus, precise analysis is required to determine the algorithm properties. Moreover, even if the cost function is inspired by the minimization of the kurtosis of the estimation error, practical interest and comparison with the performances of other algorithms concentrate on the achievable mean squared estimation error. Recent

* Corresponding author.
E-mail addresses: bershad@ece.uci.edu (N.J. Bershad), j.bermudez@ieee.org (J.C.M. Bermudez).

[1] This work was partly supported by CNPq under grant No. 307071/2013-8.

[2] Conditioned on the adaptive weights.

http://dx.doi.org/10.1016/j.dsp.2016.03.012
1051-2004/

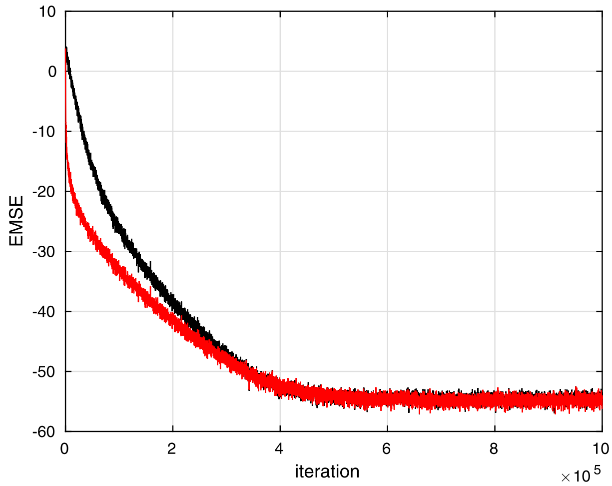

Fig. 1. Example comparing LMK and LMS performances for Gaussian input and Gaussian noise. $\mu_{\text{LMS}} = 2.2 \times 10^{-5}$, $\mu_{\text{LMK}} = 10^{-4}$. LMK converges faster away from the optimal solution and slower close to steady-state. Step sizes designed for equivalent steady-state results. LMS – black (top) curve, LMK – red (bottom) curve. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

results [22] have shown, under reasonable approximations, that the performance surface of the implemented LMK algorithm is indeed different from the negative of the error kurtosis, and that the minimum of the actual cost function is unique and corresponds to the minimum of the mean square error (MSE) performance surface (the Wiener solution). The study in [22] has also shown that LMK tends to converge faster than LMS far from the optimum, and slower than LMS as their solutions approach the Wiener solution. Another important result from [22] is that the actually implemented LMK algorithm may outperform LMS even for Gaussian inputs and Gaussian noise. The simulation results shown in Fig. 1 illustrate this property for zero-mean white Gaussian input with unity variance and zero-mean white Gaussian additive noise with variance 10^{-2}. The step sizes μ_{LMS} and μ_{LMK} were adjusted to obtain the same steady-state performance for both algorithms. In addition, the computational complexity of both algorithms are very similar for high order adaptive filters, as LMK requires $2N + 5$ multiplications and $N + 3$ additions per iteration, while LMS requires $2N + 1$ multiplications and $N + 1$ additions per iteration for a filter of length N.

A first analysis of the behavior of the LMK algorithm for Gaussian inputs and any zero-mean white additive noise with even probability density function (pdf) was presented in [23]. The conditional mean squared error, appearing in the weight update equation, is replaced in [23] with a truncated approximation of the solution for the recursive equation proposed in [11]. The approximation is accurate for small values of the parameter of the first-order autoregressive estimate of $E[e^2(n)]$, which corresponds to a small variance of the estimation error. The resulting model requires the determination of three input correlation matrices, namely $R_i = E[X(n)X^\top(n-i)]$ for $i = 0, 1, 2$. In [24], the steady-state performance of the algorithm has been studied for symmetrical and asymmetrical noise distributions. The study in [24] is restricted to white input signals and utilizes the same type of approximation used in [23] when estimating the conditional second and third order moments of the estimation error.

This paper presents a new analytical model for the LMK algorithm behavior. The model is derived for Gaussian inputs and additive noise with any symmetric pdf as in [23]. Contrary to [23], no approximation is used for the recursive update equation of the mean squared error estimation. The resulting model is valid for any value of the parameter of the autoregressive estimation of the mean squared error. Moreover, it requires only the knowledge of one input correlation matrix.

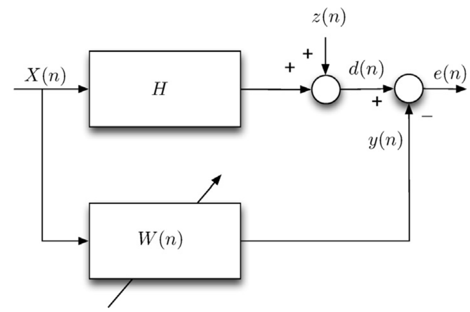

Fig. 2. Block diagram of the problem studied.

Simulation results are shown for three distinct noise distributions (Gaussian, uniform and binary) and for a large set of parameter values. Comparisons of these results with the performance predicted by the derived models illustrate the accuracy of the latter.

The paper is organized as follows. Section 2 reviews the update equation of the LMK algorithm as proposed in [11]. Section 3 presents the stochastic analysis leading to recursive analytical expressions for the mean and mean-square behavior of the adaptive weights. Section 4 provides some insight into the steady state behavior of the algorithm. Section 5 illustrates the accuracy of the theoretical model for three noise distributions. Section 6 compares the EMSE behaviors of the LMK, LMF and LMS algorithms using Monte Carlo simulations. Section 7 discusses slowdown of the LMK and LMF algorithms for low noise powers. Section 8 presents the conclusions.

2. The LMK algorithm

Fig. 2 shows the block diagram of the problem studied. Vector $H = [h_1, \cdots, h_N]^\top$ contains the samples of the optimal solution of the linear estimation problem. $W(n) = [w_1(n), \cdots, w_N(n)]^\top$ is the weight vector of the adaptive transversal FIR linear estimator, to be adapted using the LMK algorithm. The input $x(n)$ is assumed stationary, zero-mean and Gaussian with variance σ_x^2. $X(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^\top$ is the observed data vector. $z(n)$ is the measurement noise, assumed stationary, white, zero-mean, with variance σ_z^2, uncorrelated with any other signal and to have an even probability density function $p_Z(z)$. $y(n)$ is the adaptive filter output and $e(n) = d(n) - y(n)$ is the estimation error to be minimized in the kurtosis sense.

The cost function for the LMK algorithm is the negative kurtosis of $e(n)$, given by [11]

$$J(n) = 3E^2[e^2(n)] - E[e^4(n)] \tag{1}$$

given $W(n)$, or equivalently, given $\mathcal{W} = \{W(0), \cdots, W(n)\}$.

The LMK weight update equation is given by [11]

$$W(n+1) = W(n) - \mu \hat{\nabla} J(n) \tag{2}$$

where

$$-\hat{\nabla} J(n) = 4[3\sigma_e^2(n) - e^2(n)]e(n)X(n) \tag{3}$$

is the stochastic approximation of the gradient of $J(n)$, $X(n)$ is the input vector, $e(n) = z(n) - V^\top(n)X(n)$, $z(n)$ is the noise, $V(n) = W(n) - H$ is the weight error vector and $\sigma_e^2(n)$ is the mean square error conditioned on $W(n)$. Reference [11] proposes to compute $\sigma_e^2(n)$ recursively as

$$\sigma_e^2(n) = \beta\sigma_e^2(n-1) + e^2(n) \tag{4}$$

where $\sigma_e^2(n) = E[e^2(n)|\mathcal{V}]$ with $\mathcal{V} = \{V(0), \cdots, V(n)\}$.