SINE-HYPERBOLIC CONVEX VARIABLE STEP-SIZE ALGORITHM

BY

AHMAD KHALIFI

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

TELECOMMUNICATION ENGINEERING

JANUARY 2017

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **AHMAD KHALIFI** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN TELECOMMUNICATION ENGINEERING.**

Dr. Azzedine Zerguine
(Advisor)

Dr. Ali A. Alshaiki
Department Chairman

Dr. Abdelmalek Zidouri
(Member)

Dr. Salam A. Zummo
Dean of Graduate Studies

Dr. Mohamed Deriche
(Member)

23/1/17
Date

To my parents, my Teachers,
my family and my friends

ACKNOWLEDGMENTS

I am thankful to Allah Almighty, for supporting and helping me to complete this thesis. It was a great opportunity to work in a great place, toward a great subject with great professors and colleagues.

I would like to acknowledge King Fahd University of Petroleum and Minerals and the Department of Electrical Engineering for supporting my thesis. I want to thank Professor Azzedine Zerguine, my thesis adviser for his encouragement, advice and suggestions through thesis research. I also want to thank the members of my Committee, Associate Professor Abdelmalek Zidouri, and Associate Professor Mohamed Deriche for their help and support.

Finally, my cordial thanks and appreciations are due to my parent, my family, my friends and colleagues who support me with their prayers, love, and encouragement.

Thank You

Ahmad Khalifi, 20 December 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

LMS	:	Leas Mean Square
RLS	:	Recursive Least Square
LMF	:	Least Mean Fourth
LMMN	:	Least Mean Mixed Norm
SOS	:	Second order-statistics
HOS	:	High order statistic
LHS	:	Least Hyperbolic Sine
EMSE	:	Excess Mean Squared Error
NLMS	:	Normalized Least Mean Squares
ECVSS	:	Exponential Convex Variable Step-Size
SNR	:	Signal to Noise Ratio
VSS	:	Variable Step Size
APA	:	Affine Projection Algorithm

THESIS ABSTRACT

NAME: Ahmad Khalifi

TITLE OF STUDY: Sine-hyperbolic convex Variable step-size algorithm

MAJOR FIELD: Telecommunication Engineering

DATE OF DEGREE: January 2017

In this work a novel family of adaptive filtering algorithms is introduced. They are found by proposing an exclusive hyperbolic sine objective function. These algorithms belong to the variable step size (VSS) class, which are shown to be very successful and of high demand in adaptive filtering theory. Unlike existing VSS algorithms, this new proposed algorithm posses only one tuning parameter. Experimental results show that with sub-optimal selection of the tuning parameter, the algorithm provides a very promising results, especially both stationary and tracking situations. Analytic convergence and steady state error performance analyses are provided to demonstrate the performance. Also, an optimal solution, based on the least hyperbolic sine error, is derived to confirm the convergence of the proposed algorithm towards the Wiener solution.

ملخص الرسالة

الأسم : أحمد محمد صالح خليفي

عنوان الرسالة : خوارزمية جيب الدالة الزاندية المحدبة

التخصص : هندسة الإتصالات

التخرج: ربيع الثاني سنه **1438** هجرية

تقوم الدراسة على إيجاد خوارزمية جديدة لم يسبق أن تم نشرها أو ذكرها في أي مرجع علمي (إعتمادا"على أفضل ما نملك من معلومات) تعمل على زيادة سرعة الأنظمة الذكية على التكيف والتعلم مع البيئة المحيطة بها.

إن سوق العمل يبدي إهتماما" متزايدا" سنة تلو الأخرى بالأنظمة الذكية أو ما يعرف بـ (تعلم الالة) لما أثبتته هذه الأنظمة من قدرة على الاداء المتميز تحت أقصى الظروف مع سرعة إستجابة عالية للمتغيرات المحيطة والتكيف معها.

إن أنظمة الإتصالات الحديثة تعاني من عده عقبات أهمها تغير جودة الأشارة مما يؤثر على جودة الأرسال والأستقبال وهذا التغير اصبح مضطردا" في الأونة الأخيرة نظرا" لكثرة أنظمة الأتصالات الغير متوافقة وزيادة الأعتماد على الأرسال الاسلكي وخير مثال على ذلك الأجهزة المحمولة الذكية وما تتطلبه من إتصال دائم بالشبكة بالرغم ان الجهاز المحمول هو الأخر يتحرك وبسرعة لا بأس بها (كأن يكون في سيارة أو قطار) وما ينتج عن هذه الحركة من تغير في خصائص القناة الاسلكية الموصلة وما يتطلبه الأمر من جهاز الأرسال على إعادة ضبط إعداداته مع هذه المتغيرات للحفاظ على جودة الأرسال.

إن هذه الخوارزمية الجديدة تقوم على جيب الدالة الزاندية المقطع والتي هي في الأساس دالة محدبة تتمتع بالمرونه وقد تم إدخال بعض التحسينات عليها كمعامل التضخم ووضعها ضمن دالة القيمة الصغرى وقد تم دراسة أداء الخوارزمية من خلال برامج محاكاة وكانت النتائج جيدة ولله الحمد والفضل.

CHAPTER 1

INTRODUCTION

1.1 Introduction

When Recursive Least Square (RLS) adaptive method has been introduced in 1980s, a lot of researchers expected that the old adaptive method Least Mean Square (LMS) will be phased-out due to the new capability that RLS has which is fast convergence! However, the LMS algorithm survived this instant and is implemented in so many applications. One of the secret behind that is the simplicity of LMS. It does not require complex computational effort (LMS needs O (M) Vs. O (M^2) operations per iteration for RLS), and robustness.

After that, a lot of researches focused on how to improve LMS algorithm, and several versions of LMS have been introduced in the last thirty years, like normalized least mean squares (NLMS) algorithm, the least mean mixed norm (LMMN) algorithm, the Gauss-Newton algorithm, the sign-error LMS algorithm, the leaky LMS algorithm, the least mean fourth algorithm (LMF), among other algorithms. Those methods deliver an improvement in the convergence speed and some of them also success to have low steady-state error, others success to increase the immunity against different type of noise like LMF. Some methods pay higher price than others in term of increasing computational complexity yet deliver similar performance matching the others who require less computational complexity.

Variable step-size was a new class of LMS where the step-size becomes a time-variable instead of fixed scalar in the standard LMS (or fixed vector in case of affine projection algorithm (APA)). Step-size has a great impact on LMS family algorithms, so the value of the step size must be chosen carefully to guarantee the stability of the algorithm and also to satisfy the performance requirement. Large step size (large comparing to the reciprocal of the input signal power) will speed up the convergence, and will generate certain steady state error level; while small step-size will slow the convergence speed but will lower also the steady state error level. So step-size parameter, (μ) , plays an important role in guiding the performance of the LMS algorithm family. These are conflicted requirements and a compromise solution has to be adopted. NLMS (Normalized Least Mean Square) algorithm in general has gained more focus in real-time applications (voice / video) because of the balance that NLMS delivers among computational cost, tracking ability and end-to-end performance.

A new class called: Variable step-size tried to solve this dilemma. Some versions of this class were introduced in the last twenty-five years. Different versions of variable step-size algorithms delivered different speeds of conversions and different level of steady-state errors. There were new parameters introduced with some of the variable step-size algorithms, which create a lot of annoy in practical usages of those methods. The authors introduced new parameters named α, β, γ,...etc; and gave numerical values to them based on their experimental trials in the lab. Those parameters are not fixed but must be changed to different values if the signal environment changes like the input power signals, the signal to noise ratio, the filter orders, without clear procedures, so it is almost trial and error until you find the best numerical value! This is by itself a barrier to implement these algorithms in broad applications and different environment.

1.2 Literature review

LMS as well as NLMS algorithms families have gained a lot of attention from research studies due to their simplicity and robustness; which facilitate to implement them in so many applications. Their limitations from using fixed step-size encouraged the researchers around the world to find solutions for the two conflict goals, large step-size (within stable range) will lead to high speed convergence, also it will result in high steady state error while small step-size (within stable range) will lead to minimizing the error at steady state zone but slow down the speed of convergence. One of the solutions is to keep changing the step-size value so using larger values if the filter coefficents state is far from the true values, so this will ends-up speeding the convergence speed rate as well as the tracking speed rate. At the moment the algorithm is close enough from the optimum point, smaller values of step-size could be used in order to reach lower level of steady-state error, hence; reproduce enhancement result and better performance. This can be achieved by choosing the right step-size values based on some criteria that deliver acceptable measurement of the adaptation progress. Certain criteria have been used like instantaneous error squared [1], sign changes of successive samples of the gradient[2], attempting to reduce the squared error at each instant [3], or cross correlation of input and error [4]. Some algorithms may perform better than others, however; it is not always easy task to compare them fairly since most of these algorithms require tuning of many parameters.

One of the popular approach that implements variable step-size, which changes its value with time within the standard LMS weight update recursion equation was made by Aboulnasr and Mayyas (1997) [5], where they try to overcome the weaknesses on the previous methods that were published in 19980s and 1990s like high sensitivity to the noise distribution and their high performance over the LMS level of performance is in general attained only in a high SNR signal level. Step-size parameter of the algorithm is adjusted based on squaring the time-average estimate of the auto-correlation among current instantaneous error and previous one. This leads to effective measure of optimal result of the filter coefficients independently of uncorrelated measurement noise. This method showed superior performance among previous methods that was introduced prior to 1997, however, it required introducing some new variables, that need to be determined by trial and error to have the best performance and those parameters are dependable on the signal environment like SNR, and channel length, which represent an obstacle to use this method optimally.

Another famous approach was introduced by Shin, Sayed [6]. The projected weighted error norm vector is the criterion to make sure how close the adaptive filter coefficients from their true values. This method maintains high speed convergence while shows the same steady-state error like the previous methods. However, this method in the publication paper only tested on 30dB SNR, when I try to simulate it on 10dB SNR, the performance degraded significantly! (The author doesn't simulate on 10dB SNR). It is common regarding majority of time-variable step-size algorithms that they may not work very reliably since they depend on several parameters that are not simple to tune in practice [4].

Exponential convex variable step-size (ECVSS) which is a kind of stochastic gradient algorithm [7] was based on a cost function of natural exponential (of the error squared). The step-size values in this method have been guided by exponential function values of the error squared. To improve the convergence; a scaling factor A introduced ($A > 0$), ECVSS outperform LMS (and some LMS variant) in convergence speed rate; however the misadjustment is sensitive to the noise distribution as well as the value of the scaling factor A. Comparing to the traditional LMS algorithm, the ECVSS algorithm requires three multiplications, one comparison and an exponential term per iteration more than LMS. Exponential-value calculations could be avoided by using look up table. Fast convergence of ECVSS algorithm due to large gradients could be a source of instability. To overcome that; bounded gradient introduced where we can take the advantage from the large gradient in term of fast convergence and also bounded the inflation of the gradient to secure the stability.

1.3 Problem Statement

Traditionally, adaptive filter algorithms compromise two conflict outcomes, high speed convergence and low level steady state error. One of the solutions is to use time variant step size. The attempts to achieve that result introduced algorithms that require high computational power and required tuning few parameters based on the level and type of noise along with input signal type and filter length. These two requirements trigger the interest to search for a simple algorithm that can achieve similar or even better results and at the same time maintain low demand on computational power and reduce the tuning parameters in order to simplify the implementation of the algorithm.

1.4 Thesis Objectives

The goal is to find a simple yet robust adaptive algorithm that achieve fast convergence speed and reach to low steady state error level. The current algorithms demand high computational power and required tuning and coordination among few parameters in order to adjust them to the right values that maximize the desired outcomes. Also, we are looking to leverage this new algorithm to create different versions belong to different adaptive algorithms families that can serve in different environments like a special version to work in sub Gaussian noise environment and another special version that can work in mixed noise type environment.

1.5 Thesis Organization

Introduction and literature review start from Chapter one and then followed by four chapters

1 Chapter two about hyperbolic sine algorithm.

2 Chapter three about hyperbolic sine of 4th order algorithm.

3 Chapter four about mixed norm hyperbolic sine algorithm.

4 Chapter five conclusion and future studies.

5 Appendix.

6 References.

CHAPTER 2

HYPERBOLIC SINE ALGORITHM

2.1 Introduction

Most of the gradient algorithms are quadratic-based cost functions because it is simple mathematically and easy to lead to close forms in the analysis. These kind of algorithms are refered to as linear-based or Second order-statistics (SOS) cost functions. LMS [8] and NLMS [9] belong to such a class. Higher order power of adaptation error resulted into High order statistic (HOS) calss of adaprive filters. LMF [10] is an example of this kind of algorithms. HOS algorithms have shown a superior performance over SOS especially in speed of convergece, yet they have a higher misadjstment level, unless the noise is non-Gaussian. The improved speed of convergence of HOS algorithm is due to the steeper error surface which severly penalize the high deviation from the optimum solution. Recently, to improve the SOS speed of convergence and maintain a sufficient level of convergence, new family based on cost function that have exponential dependence on adaptation error [7] [11], has been introduced. recently. These algorithms, have a steeper surface than the quadratic cost function and can be seen as a linear combination of all the even moments. This type of adaptive procedures outperform the LMS in term of convergence speed with increased robustness against the impulsive noise.

In this research, we propose a new cost function, named least hyperbolic sine, which uses the error square as a driving argument. Accordingly a stochastic gradient based algorithm, named as Least Hyperbolic Sine error Squared (LHS) algorithm, is derived. The new algorithm is classified as Variable Step Size (VSS) algorithm and has some gains in terms of speed of convergence, adaption to the sudden changes, less computational cost as compared to the non-linear VSS algorithm, in addition to a less required tuning parameters. In the sequel, the derivation of the algorithm is supported by a thorough analysis to figure out the required conditions for the convergence. The excess mean steady state error (EMSE) is calculated, and the optimal solution in the least hyperbolic sine sense is derived too.

The chapter is organized as follows: the new cost function and the flow of the algorithm derivation is presented in section 2.2. In section 2.3 the optimal solution is derived. Section 2.4 provides the steady state analysis throughout finding a closed form for the EMSE. Section 2.5 outlines the convergence analysis.

Sections 2.6 to 2.9 cast the mean stability conditions followed by the gain within computational cost, then introduce generic step-size, and support the analytic finding through a course of computer simulations. Finally, the conclusions are stated in section 2.10.

The followings are the notations used in this paper: $\mathbf{x}$ denotes a column vector, x is a scalar, $(.)^T$ is the transpose operator. E[.] is the mathematical expectation, and Tr[.] is the trace operator.

2.2 Algorithm Formulation

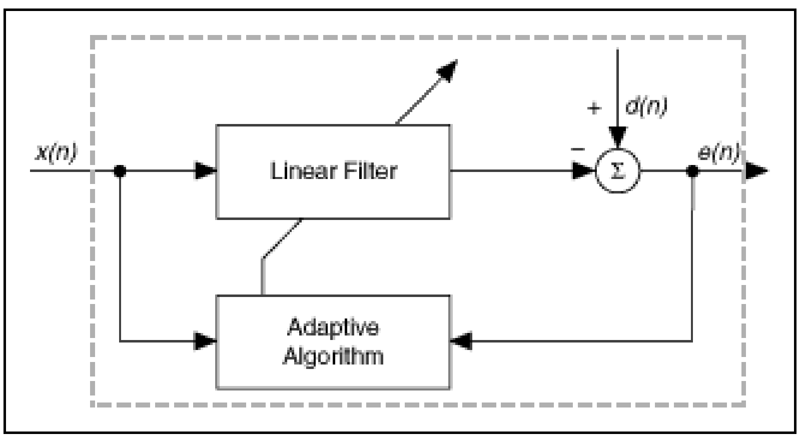

Figure 2.1: Principal diagram of adaptive filter

The considered formulation is applied to the system identification scheme, whereby the proposed algorithm works toward minimizing the hyperbolic sine cost function of the error squared. In reference to Fig. 2.1, the instantaneous error is defined as:

$$e(k) = d(k) - \mathbf{x}_k^T \mathbf{w}_{k-1} \tag{2.1}$$

The desired signal

$$d(k) = \mathbf{x}_k^T \mathbf{w}^o + v(k) \tag{2.2}$$

where $v(k)$ is a zero-mean independent random variable, and $\mathbf{w}^o$ is the optimal time-varying filter coefficients, while $\mathbf{w} = [w_0, w_1, \cdots, w_{M-1}]^T$, is the filter coefficients and M is the filter length. $(.)^T$ is the transpose operator, and $\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-\mathrm{M}+1)]^T$ is the input signal vector.

The new cost function is the hyperbolic sine with the error square argument, defined as

$$J(k) = \sinh(e^2(k)) \tag{2.3}$$

This is a convex and uni-modal function. Its gradient with respect to the filter coefficients yields

$$\Delta_{\mathbf{w}} J(k) = -2e(k)\cosh(e^2(k))\mathbf{x}^T(k) \tag{2.4}$$

$\mathbf{x}(k)$ is the regression vector. To improve the convergence speed, one can introduce a scaling parameter $(A > 0)$ [7], to scale the squared error, in the argument of the hyperbolic sine; so that the modified cost function will be

$$J(k) = \frac{1}{A}\sinh(Ae^2(k)) \tag{2.5}$$

Accordingly, the gradient with the new cost function will be

$$\Delta_{\mathbf{w}} J(k) = -2e(k)\cosh(Ae^2(k))\mathbf{x}^T(k) \tag{2.6}$$

Hence, the stochastic recursive form of the coefficients estimate given as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu e(k)\cosh(Ae^2(k))\mathbf{x}(k) \tag{2.7}$$

We observe that the hyperbolic cosine scales up the step size in case of a high instantaneous error, this lead to a fast convergence. However, this might lead to undesirable negative consequence on the algorithm stability. To utilize the large gradient property and maintain a bounded gradient to preserve the algorithm stability, we use the following selecting function [7]:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu_{min} \min \left[cosh(Ae^2(k)), \frac{\mu_{max}}{\mu_{min}} \right] e(k)\mathbf{x}(k) \tag{2.8}$$

where μ_{max} and μ_{min} are the upper and lower bounds of μ, respectively, while μ is the step size.

2.3 The Optimal Solution

In the algorithm design, one need to guarantee that it will lead to a unique optimal solution so that the algorithm behavior becomes controllable. This is indeed the case with our new proposed hyperbolic sine function.

To this end, the optimal solution is found based on the gradient of the hyperbolic sine cost function, as follows:

$$\Delta_{\mathbf{w}} J(k) = -2e(k) \cosh(Ae^2(k))\mathbf{x}^T(k) = 0 \tag{2.9}$$

To express the later equation in terms of the optimal tap weights $\mathbf{w}^o$, substitute for the $e(k)$ from (2.1) , and after some manipulation, we get

$$\mathbf{x}(k)d(k) \cosh(Ae^2(k)) = \mathbf{x}(k)\mathbf{x}_k^T \mathbf{w}^o \cosh(Ae^2(k)) \tag{2.10}$$

Then, taking the mathematical Expectation of both sides lead to

$$\mathrm{E}\left[\mathbf{x}(k)d(k)\cosh(Ae^2(k))\right] = \mathrm{E}\left[\mathbf{x}(k)\mathbf{x}_k^T\mathbf{w}^o\cosh(Ae^2(k))\right] \tag{2.11}$$

Substitute Taylor power series expansion of cosh function into (2.11) results in

$$\mathbf{P}_{\mathbf{x}d} + \sum_{n=1}^{\infty}\frac{1}{(2n)!}\mathrm{E}[\mathbf{x}(k)d(k)e^{4n}(k)] = \mathbf{R}_{\mathbf{x}}\mathbf{w}^o + \sum_{n=1}^{\infty}\frac{1}{(2n)!}\mathrm{E}[\mathbf{x}(k)\mathbf{x}^T(k)e^{4n}(k)]\mathbf{w}^o \tag{2.12}$$

where $\mathbf{R}_{\mathbf{x}} = \mathrm{E}[\mathbf{x}(k)\mathbf{x}^T(k)]$ is the auto-correlation matrix of the input signal $\mathbf{x}(k)$. and $\mathbf{P}_{\mathbf{x}d} = \mathrm{E}[\mathbf{x}(k)d(k)]$ is the cross-correlation between the input signal $\mathbf{x}(k)$ and the desired signal $d(k)$.

Assuming that both the input vector sequence $\{\mathbf{x}(k)\}$ and the error signal sequence $\{e(k)\}$ to be asymptotically uncorrelated, i.e., $\mathrm{E}[\mathbf{x}(k)\mathbf{x}^T(k)e^{4n}(k)] = \mathbf{R}_{\mathbf{x}}\mathrm{E}[e^{4n}(k)]$. Moreover, since the error signal is small at the steady state scenario, the higher power order error $e(k)$ can be ignored. Hence, these situations result into an expression for the optimal tap weight given as

$$\mathbf{w}^o = \mathbf{R}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}d} \tag{2.13}$$

The optimal solution is the Wiener solution [12],[8] which is the optimal solution for the LMS algorithm, too. Close investigation of the gradient part of (2.8), one can easily justify this similarity. In fact, when the error signal $e(k)$ approaches zero which is the case at the steady-state zone, the hyperbolic cosine can be approximated around the origin as $\cosh(e^2(k)) \approx 1$, hence the cost function effect is in match with the quadratic cost function, which is the LMS algorithm in the standard from.

2.4 Steady State Analysis

It is investigated through deriving an analtical expression for the Excess Mean Squared Error (EMSE). The analysis in this section is based on the energy conservation relation framework [8]-[13]. In addition to the wide sense stationary channel model assumption, the following standard assumptions [8] are introduced:

A1. There is a true values vector $\mathbf{w}^o$ leads to $d(k) = \mathbf{x}^T(k)\mathbf{w}^o + v(k)$

A2. The additive sequence of noise $\{v(k)\}$ is i.i.d. with identical variance $\sigma_v^2 = \mathrm{E}[(v(k))^2]$

A3. The sequence $v(i)$ is independent of the input vector $\mathbf{x}(j)$ for all i,j

A4. The start up values $\mathbf{w}_{-1}$ are independent of all $\{d(j), \mathbf{x}(j), v(j)\}$

A5. The input signal auto-correlation matrix $\mathbf{R_x} = \mathrm{E}[\mathbf{x}(k)\mathbf{x}^T(k)] > 0$

A6. The random variables $\{d(k), \mathbf{x}(k), v(k)\}$ are centralized with zero means

Start by the general relation of the steady-state EMSE [8]

$$\mathrm{S} = \frac{\mu}{2}\frac{\mathrm{N_s}}{\mathrm{D_s}}\mathrm{Tr}[\mathbf{R_x}] \tag{2.14}$$

where $\mathrm{Tr}[\mathbf{R}_{\mathrm{x}}]$ is the trace of the auto-correlation matrix of the input signal. $\mathrm{N_s}$ is defined as:

$$\mathrm{N_s} = \mathrm{E}[\mathrm{f}^2(e(k))] \tag{2.15}$$

also, $\mathrm{D_s}$ is given by

$$\mathrm{D_s} = \frac{\mathrm{E}[e_a(k) \cdot f(e(k))]}{\mathrm{E}[e_a^2(k)]} \tag{2.16}$$

where $e_a(k)$ is apriori error defined as

$$e_a(k) = [\mathbf{w}^o - \mathbf{w}(k)]^T \, \mathbf{x}(k) \tag{2.17}$$

and $f(e)$ can be defined from the proposed recursion equation (2.8) at the steady state zone which can be shorten (due to small error value) as follows:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \underbrace{2\mu_{min}}_{\mu}\underbrace{cosh(Ae^2(k))e(k)}_{f(e)}\,\mathbf{x}(k) \tag{2.18}$$

So $\mathrm{N_s}$ will be written as follows:

$$\mathrm{N_s} = \mathrm{E}[e^2 \cosh^2(Ae^2)] \tag{2.19}$$

For the purpose of brevity, the time index "k" has been deleted. The estimation error e can be written in term of apriori error and noise signal as $e = e_a + v$ [8]

$$\mathrm{N_s} = \mathrm{E}[e_a^2 \cosh^2(Ae^2)] + \sigma_v^2 \mathrm{E}[\cosh^2(Ae^2)] \tag{2.20}$$

where σ_v^2 is the variance of the noise. A step more further toward simplicity by applying the Cauchy-Schwartz inequality.

$$\mathrm{N_s} \leq \sqrt{\mathrm{E}[e_a^4] \cdot \mathrm{E}[\cosh^4(Ae^2)]} + \sigma_v^2 \mathrm{E}[\cosh^2(Ae^2)] \tag{2.21}$$

Furthermore, assume apriori error to be zero-mean Gaussian; we can apply Jensen's inequality to solve the expectation for the hyperbolic cosine function. Thus, N_s can be written in a closed form as:

$$\mathrm{N_s} \leq [\sqrt{3}\mathrm{S} + \sigma_v^2] \cdot \cosh^2(A\mathrm{E}[e_a^2 + \sigma_v^2]) \tag{2.22}$$

$$\mathrm{N_s} \leq [\sqrt{3}\mathrm{S} + \sigma_v^2] \cdot \cosh^2(A[\mathrm{S} + \sigma_v^2]) \tag{2.23}$$

where S $\triangleq \lim_{k\to\infty} \mathrm{E}[e_a^2(k)]$

In the same manner D_s in (2.14) can be written as follows:

$$\mathrm{D_s} = \frac{\mathrm{E}[e_a \cdot e \cdot \cosh(Ae^2)]}{\mathrm{E}[e_a^2]} \tag{2.24}$$

by substituting $e(k) = e_a(k) + v(k)$ we will have:

$$\mathrm{D_s} = \frac{\mathrm{E}[(e_a^2 + e_a \cdot v)\cosh(Ae^2)]}{\mathrm{E}[e_a^2]} \tag{2.25}$$

Since v is independent zero-mean Gaussian noise

$$\mathrm{D_s} = \frac{\mathrm{E}[e_a^2 \cosh(Ae^2)]}{\mathrm{E}[e_a^2]} \tag{2.26}$$

By applying Cauchy-Schwartz inequality:

$$\mathrm{D_s} \leq \frac{\sqrt{\mathrm{E}[e_a^4]\mathrm{E}[\cosh^2(Ae^2)]}}{\mathrm{E}[e_a^2]} \tag{2.27}$$

Followed by Jensen's inequality after assuming apriori error to be zero-mean Gaussian;

$$\mathrm{D_s} \leq \sqrt{3}\,\cosh(A(\mathrm{S} + \sigma_v^2)) \tag{2.28}$$

Excess mean square error at steady state zone can be rephrased in a closed form by shortening the hyperbolic cosine function via its Taylor series expansion and by substituting the values of $\mathrm{N_s}$ and $\mathrm{D_s}$ in (2.14)

$$\mathrm{S} = \frac{\mu_{min}\mathrm{Tr}[\mathbf{R}_x].\sigma_v^2[1 + \frac{A^2}{2}\sigma_v^4]}{\sqrt{3} - \mu_{min}\mathrm{Tr}[\mathbf{R}_x][\sqrt{3} + \frac{\sqrt{3}}{2}A^2\sigma_v^4 + A^2\sigma_v^6]} \tag{2.29}$$

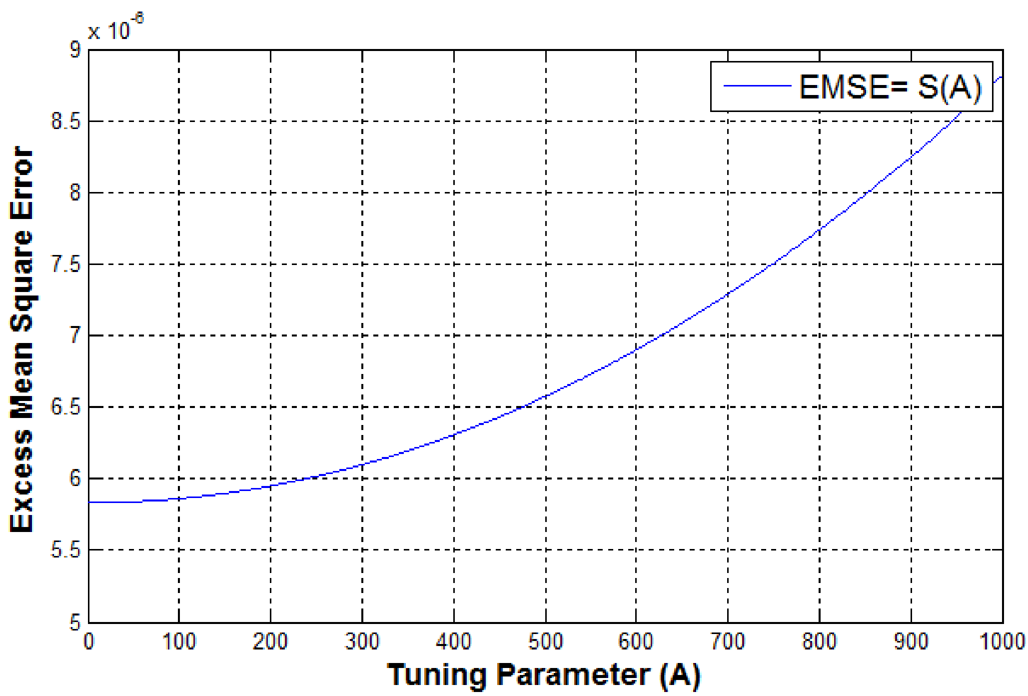

Figure 2.2: Excess Mean Square Error as a function of the tuning parameter (A)

As shown in figure 2.2 for SNR = 30dB, $\mu = 0.01, \mathrm{Tr}[R_x] = 2$; The excess mean square error "EMSE" is increasing as the tuning parameter increases which create a supplementary issue in the implementation. However if the tuning parameter increased up to the level that creates: $\cosh(Ae^2) > \frac{\mu_{max}}{\mu_{min}}$ for all e^2 then the algorithm will behave like LMS with $\mu = \mu_{max}$ all the time. Another impact of large tuning parameter inferred from the simulation experiments that it will cause large fluctuation of the EMSE around its average value.

It can be shown from equation (2.29) that the excess mean square error coincide with that of LMS for A=0. So from now on; we will write μ_{min} as μ_{minLMS} and μ_{max} as μ_{maxLMS} to reflect the relationship between our proposed algorithm and the standard LMS.

2.5 Convergence Analysis

Starting with the general class error adaptive filter:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu f[e(k)]\mathbf{x}(k) \tag{2.30}$$

where

$$f[e(k)] = \frac{2}{\mu}\mu_{minLMS}.min\left[\cosh(Ae^2), \frac{\mu_{maxLMS}}{\mu_{minLMS}}\right]e(k)$$

$f[e(k)]$ doesn't have power Taylor series expansion running to infinity [14], since $f(e)$ at any point doesn't always have the first derivative (Appendix 5.2). However, the approximation:

$$\mathbf{w}(k+1) \approx \mathbf{w}(k) + \mu\{f[e(k)]\mathbf{x}(k) - f'[e(k)]\mathbf{x}(k)\mathbf{w}^T(k)\mathbf{x}(k) + \frac{1}{2}f''[e(k)]\mathbf{x}(k)[\mathbf{w}^T(k)\mathbf{x}(k)]^2\} \tag{2.31}$$

holds in every point except when

$$e(k) = \pm\sqrt{\frac{1}{A}\cosh^{-1}\left(\frac{\mu_{maxLMS}}{\mu_{minLMS}}\right)}$$

So, we can carry on the analysis under the assumption that the noise values are very rare to equal to δ where

$$\delta = \pm\sqrt{\frac{1}{A}\cosh^{-1}\left(\frac{\mu_{maxLMS}}{\mu_{minLMS}}\right)}$$

To find-out the time constant of the proposed algorithm for the general case; we need to evaluate $\mathrm{E}[f'(e(k))]$. We start with the assumption:

$$e(k) \neq \delta$$

then we have

$$f(e) = \begin{cases} \dfrac{2\mu_{minLMS}}{\mu} x \cosh(Ae^2); & |e| < \delta \\ \dfrac{2\mu_{maxLMS}}{\mu} e; & |e| > \delta \end{cases} \tag{2.32}$$

and

$$f'(e) = \begin{cases} \dfrac{2\mu_{minLMS}}{\mu} \left(\cosh(Ae^2) + 2Ae^2 \sinh(Ae^2)\right); & |e| < \delta \\ \dfrac{2\mu_{maxLMS}}{\mu}; & |e| > \delta \end{cases} \tag{2.33}$$

So, the time-constant of the proposed algorithm associated with λ_i (the i^{th} eigenvalue of the auto-correlation matrix $\mathbf{R_x}$ can be determined for small μ from [15] as follows

$$\tau_i = \frac{1}{\mu \mathrm{E}[f'[e(k)]]\lambda_i} \tag{2.34}$$

for the proposed algorithm we have the following two cases:

1. if $A < \dfrac{1}{e^2} \cosh^{-1}\left(\dfrac{\mu_{maxLMS}}{\mu_{minLMS}}\right)$ then

$$\tau_i = \frac{1}{2\mu_{minLMS}\mathrm{E}[\cosh(Ae^2) + Ae^2 \sinh(Ae^2)]\lambda_i} \tag{2.35}$$

2. if $A > \frac{1}{e^2}\cosh^{-1}\left(\frac{\mu_{maxLMS}}{\mu_{minLMS}}\right)$ then

$$\tau_i = \frac{1}{2\mu_{maxLMS}\lambda_i} \tag{2.36}$$

which match LMS case for $\mu = \mu_{maxLMS}$.

Since τ in the first case is smaller than LMS time-constant; the speed of convergence of the proposed algorithm will be better than the convergence speed of LMS and some LMS variants. If the tuning parameter A is not properly chosen then $A < \frac{1}{e^2}\cosh^{-1}\left(\frac{\mu_{maxLMS}}{\mu_{minLMS}}\right)$ may not happen and the proposed algorithm will work like standard LMS with $\mu = \mu_{maxLMS}$ all the time.

As shown in figure 2.3 for a four taps system identification, white Gaussian input signal, SNR = 30dB, and additive white Gaussian noise; the convergence speed has improved with increasing the value of the tuning parameter especially for $A = 10 \rightarrow 100$, while maintain the same steady state error (for SNR =30dB). The impact on steady state error is less sensitive and start being observed for $A > 100$ and this is in consistency with the result on figure 2.2 where EMSE increase significantly as A moves beyond 100. In general; for a range around optimal value of A, we obtain very similar result.

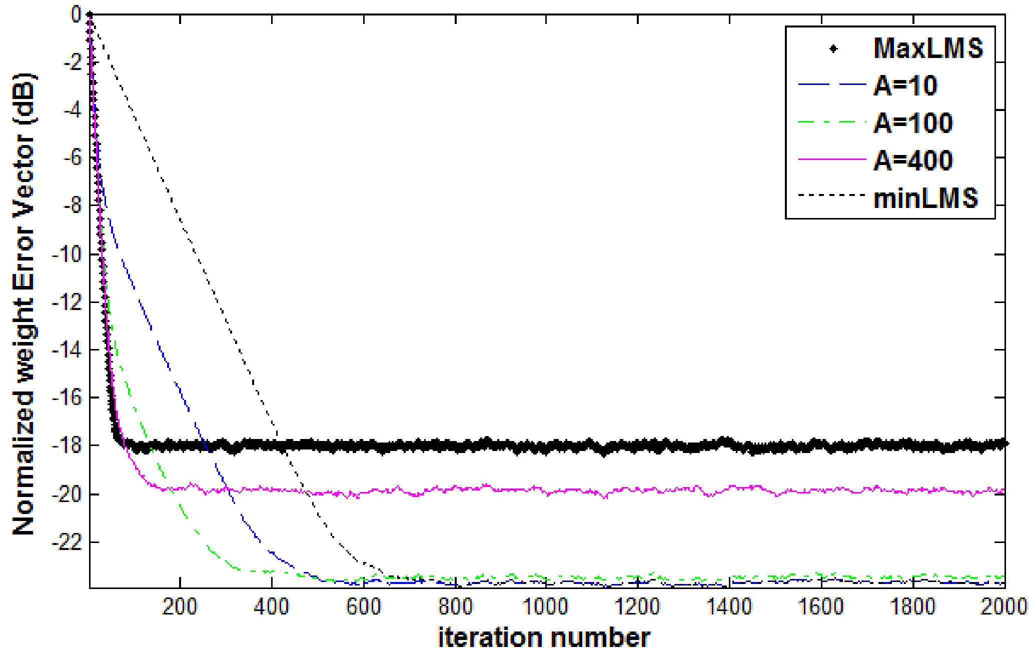

Figure 2.3: Adaptive curves for different tuning parameter values

2.6 Mean Behavior

As common in all gradient decent algorithms, the value of the step-size is critical. To guarantee the stability; the step size value should satisfy some certain bounds.

We can rewrite equation (2.8) as follows:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e(k)\mathbf{x}(k) \tag{2.37}$$

Where $\mu(k)$ can be written as follows:

$$\mu(k) = 2\mu_{minLMS} \min \left[cosh(Ae^2), \frac{\mu_{maxLMS}}{\mu_{minLMS}} \right] \tag{2.38}$$

It is sufficient mentioning that statistical mean value of $\mu(k)$, that is E$[\mu(k)]$

must satisfy the following condition:

$$0 < \mathrm{E}[\mu(k)] < \frac{2}{\lambda_{max}} \tag{2.39}$$

Where λ_{max} is the maximum eigenvalue of the auto-correlation matrix $\mathbf{R_x}$.

For the proposed algorithm, we have the following two cases:

1. if $\cosh(Ae^2) < \left(\frac{\mu_{maxLMS}}{\mu_{minLMS}}\right)$ then

$$\mu(k) = 2\mu_{minLMS} \cdot cosh(Ae^2) \tag{2.40}$$

Taking expectations of equation (2.40), and by using Taylor series we can rewrite $\mathrm{E}[\mu(k)]$ to be:

$$\mathrm{E}[\mu(k)] \geq 2 \cdot \left\{1 + \frac{3}{2}A^2F^2(k) + 3A^2F(k)\sigma_v^2 + \frac{3}{2}A^2\sigma_v^4\right\} \tag{2.41}$$

Where $F = \mathrm{E}[e_a^2]$ is the instantaneous excess mean square error and σ_v^2 is the variance of the noise. So at steady-state and by ignoring S with higher power, where $\mathrm{S} \triangleq \lim_{\mathrm{k}\to\infty} \mathrm{E}[\mathrm{e_a^2(k)}]$, we have a new condition on μ as follows:

$$0 < \mu_{minLMS} < \frac{1}{\{1 + 3A^2 S\sigma_v^2 + \frac{3}{2}A^2\sigma_v^4\}} \tag{2.42}$$

2. if $\cosh(Ae^2) > \left(\frac{\mu_{maxLMS}}{\mu_{minLMS}}\right)$ then

$$\mu(k) = 2\mu_{maxLMS} \tag{2.43}$$

and the new bound will be as follows:

$$0 < \mu_{minLMS} < \mu_{maxLMS} < \frac{1}{\lambda_{max}} \tag{2.44}$$

Which match the LMS case, however we need to choose the μ_{minLMS} first as a lower bound of μ_{maxLMS}.

2.7 Computational Cost

The extra computational load per iteration comparing with standard LMS will be: one comparison, three multiplications and one hyperbolic cosine term. Appropriate look up table can be used to reduce the required computational load rather than calculating hyperbolic cosine by its Taylor series expansion.

Table 2.7 shows the computational complexity of different algorithms, where M is the order of the filter and N is the total number of samples. We assume that the proposed method will use generic μ_{max} rather than fixed one (generic μ_{max} will be explained on the next section). This extra task's cost is already included on the table 2.7

Algorithm	×	+	Comparison	Lookup
Proposed	3N+2MN	MN	1	cosh
Peng[2]	5N+2MN	MN	1	0
MVSS[9]	8N+8	2N+2	2	0
MRVSS[10]	14N+10	4N+2	2	0
ECVSS[4]	3N	0	1	exp

Table 2.1: computational complexity of the algorithms where M is the filter order and N is total samples

It is clear that the proposed algorithm considered low computational demand while maintaining the simplicity by keeping only one tuning parameter. Without the need of the generic step size, the proposed algorithm will match ECVSS [7] on the computational complexity. Both of them are considered almost from the same class of recursive equation.

2.8 Generic upper bound of μ :

In this new proposed algorithm; we are going to introduce μ_{maxLMS} as a generic value rather than a fixed number. By Calculating μ_{maxLMS} in each iteration as follows:

$$\mu_{maxLMS} = \frac{1}{\mathrm{Tr}[\mathbf{R}_x] + \epsilon} \tag{2.45}$$

Epsilon " ϵ " is used to avoid the case when $\mathrm{Tr}[\mathbf{R}_{\mathrm{x}}]$ approaches zero. By using this generic value, we will guarantee the stability of the algorithm and improve the convergence speed. This is also confirmed by our experiments. We run the simulation for the proposed algorithm, one time with a fixed maximum step size and another time with the generic maximum step size; and found that the performance of the generic maximum step size is much better and also delivers more stability against any change in the input signal power while in the case of fixed maximum step size we may loose the stability and need to readjust μ_{maxLMS} manually.

This improvement comes at the expense of little increase on the computational cost (2M multiplications and M additions per iteration) where M is the order of the filter.

2.9 Simulation Results

We run the experiments based on system identification scenarios. The order of the unknown system will be matching the order of the adaptive filter and both of them representing FIR system. We start with zeros intial values for the adaptive filter coefficients. Adding Gaussian noise sequence $v(k)$ to the output of the unknown system. The variance σ_v^2 of the noise sequence $v(k)$ will be selected in each experiment to reflect the desired signal to noise ration (SNR). All experiments are averaged over 200 independent realizations. The quantitative performance measure is the normalized weight error squared vector in dB (also known as misadjustmnet [8]), which is mathematically calculated as follows:

$$\mathcal{M} = 10\log_{10}\left(\frac{\mathrm{E}[||\mathrm{W}^{\mathrm{o}} - \mathrm{W}_{\mathrm{k}}||^2]}{||\mathrm{W}^{\mathrm{o}}||^2}\right) \tag{2.46}$$

Where $\mathrm{W}^{\mathrm{o}} = [\mathrm{w}_0^{\mathrm{o}}, \mathrm{w}_1^{\mathrm{o}}, \cdots \mathrm{w}_{\mathrm{M}-1}^{\mathrm{o}}]^{\mathrm{T}}$ the true values of the unknown system/channel tapes weights. and $\mathrm{W}(\mathrm{k}) = [\mathrm{w}_0(\mathrm{k}), \mathrm{w}_1(\mathrm{k}), \cdots \mathrm{w}_{\mathrm{M}-1}(\mathrm{k})]^{\mathrm{T}}$ the values of the digital filter coefficients at time instant k, and M is the filter order assuming that both of them have the same order, and $[.]^T$ is the transpose of the matrix/vector.

We will run each comparison experiment individually in order to use the similar conditions like the filter order; the noise level, the channel response, the input signal, $\cdots$ etc. as recommended by the counterpart algorithms authors on their publications. This is to achieve fair comparison that can be used in future studies.

2.9.1 Example 1

In this example, we follow the same setup as in Peng and Farhang 2001 paper [16], where the adaptive filter and the unknown system are both of order = 16, the input signal is zero-mean white Gaussian process of unit variance, the desired signal is mixed by additive white Gaussian noise with zero-mean and 0.01 variance, so the SNR = 20dB. Multiplicative Peng algorithm will be used as recommneded by the authors [16] instead of linear Peng counterpart algorithm. The values of the algorithm parameters are: $\alpha = 0.95$ and $\rho = 2 \times 10^{-4}$. The proposed algorithm has a tuning parameter A = 10; and μ_{minLMS} chosen to give the same steady state misadjusmnet while $\mu_{maxLMS} = \frac{1}{\mathrm{Tr}[\mathrm{R_x}]}$ like the maximum μ allowed in Peng algorithm. Sudden change introduced at iteration 4000 where all the coefficients of the unknown system multiplied by -1 in order to test the algorithms ability to track the changes. Peng algorithm initialized with μ_{max} to provide a high initial convergence speed.

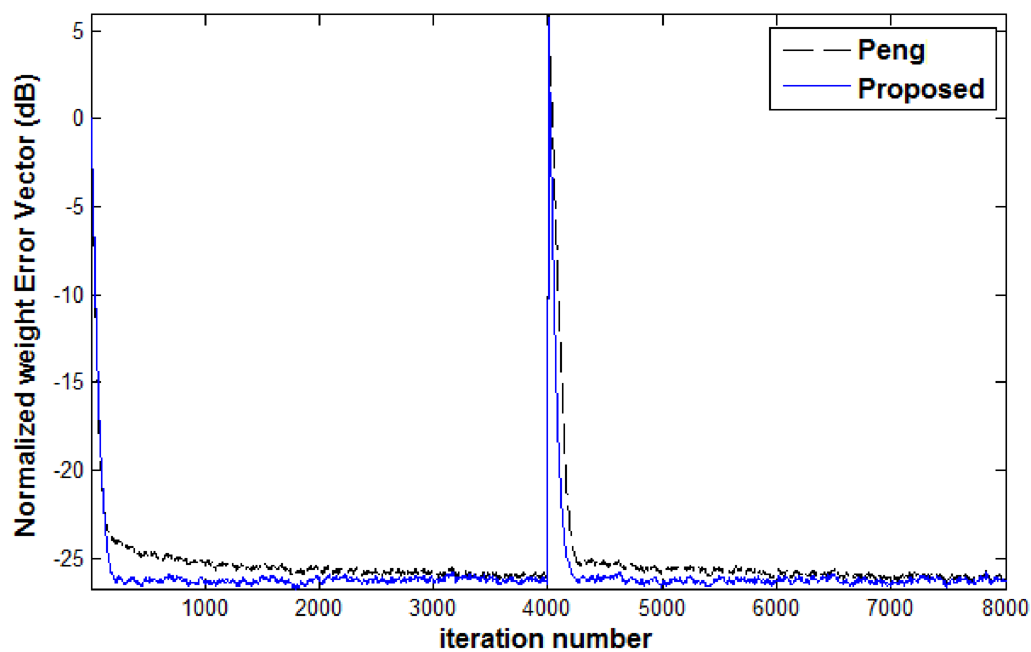

Figure 2.4: Adaptive curves of the proposed algorithm and Peng algorithm for white Gaussian input signal and SNR=20dB

Figure 2.4 shows the adaptive curve of the proposed algorithm and Peng [16].

As we can observe from the figure, the proposed algorithm converges faster and after the sudden change; it shows even better performance in tracking the change and return back to the same steady state misadjustment level. This performance is coming with less computational effort and with less tuning parameters (the proposed algorithm has only one tuning parameter).

2.9.2 Example 2

In this example, the same conditions as in Aboulnaser and Mayyas "MVSS" 1997 paper [5] are used, where the adaptive filter and the unknown system are both of order = 4, the input signal is zero-mean white Gaussian process, the desired signal is mixed by additive white Gaussian noise with zero-mean, and the signal to noise ratio SNR = 30dB. All the parameters of MVSS algorithm asigned like mentioned in [5] where $\alpha = 0.97, \beta = 0.99, \gamma = 1, \mu_{max} = 0.1, \mu_{min} = 5 \times 10^{-4}$. MVSS initialized with μ_{max} to provide a high initial convergence speed. The proposed algorithm has a tuning parameter $A = 120$; and μ_{minLMS} chosen to give the same steady state misadjustmnet level obtained by MVSS algorithm, while $\mu_{maxLMS} = \frac{1}{\mathrm{Tr}[\mathrm{R_x}]}$. At iteration 3000 all the coefficients of the unknown system will be multiplied by -1 in order to test the algorithms ability to track the sudden changes.

Figure 2.5 shows the adaptive curve of the proposed algorithm and MVSS [5]. As can be seen from the figure, the proposed algorithm converges almost at the same speed like MVSS, however after the sudden change; it shows faster tracking

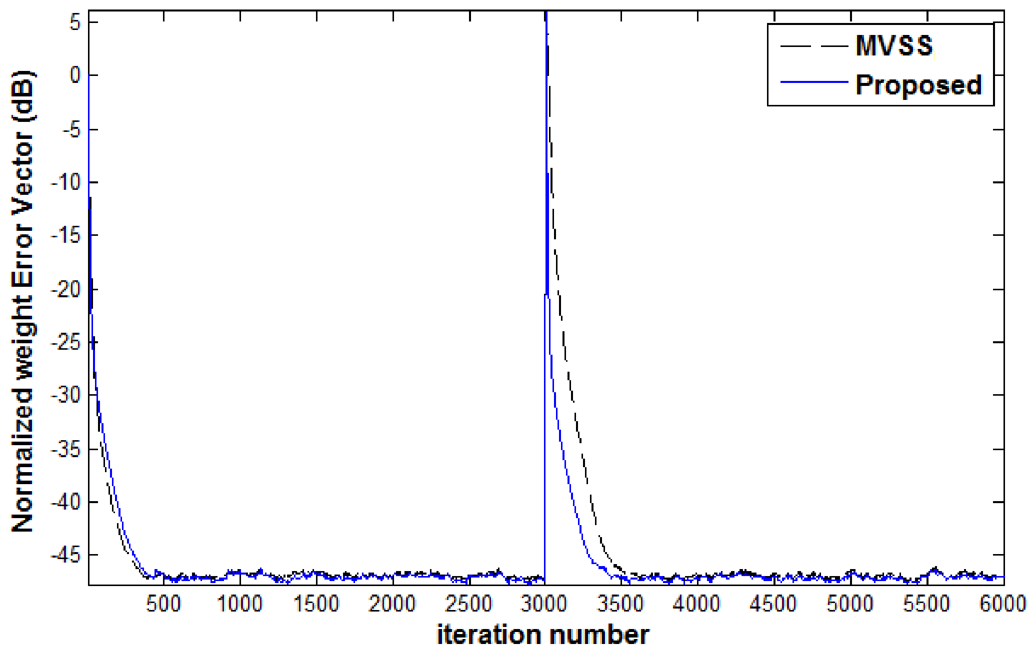

Figure 2.5: Adaptive curves of the proposed algorithm and MVSS algorithm for white Gaussian input signal and SNR=30dB

ability and return back to the same steady state misadjustment level.

Figure 2.6 shows the differences on μ values between the two algorithms for SNR=30dB; and clearly in the tracking zone, the proposed algorithm has the ability to stabilize the step size faster and reach lower steady state misadjustemnet faster than MVSS algorithm.

2.9.3 Example 3

In this example, the same conditions as in Zhao, Man and Khoo "MRVSS" 2007 paper [17] are used, where the adaptive filter and the unknown system are both of order = 4, the input signal is white Gaussian process with zero-mean , the desired signal added to white Gaussian noise with zero-mean, and the signal to noise ratio SNR = 30dB. All the parameters of MRVSS algorithm assigned like

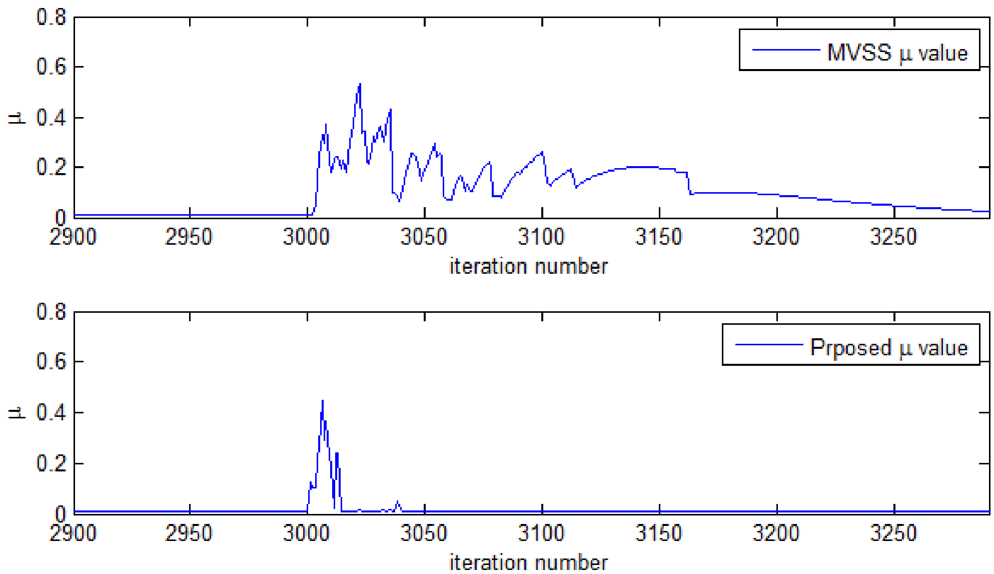

Figure 2.6: μ values at the sudden change point (iteration =3000) for SNR=30dB

mentioned in [17] where $\alpha = 0.97, a = 0.995$, $b = 1 \times 10^{-5}$ and $\mu_{max} = 0.1$. MRVSS initialized with μ_{max} to provide a high initial convergence speed. The proposed algorithm has a tuning parameter $A = 100$; and μ_{minLMS} chosen to give acceptable steady state misadjustmnet level, while μ_{maxLMS} is like the one used in MRVSS algorithm. At iterations 3000, 5000, 7000 and 9000 all the coefficients of the unknown system will be multiplied by -1 in order to test the algorithms ability to track the sudden changes.

Figure 2.7 shows the adaptive curves of the proposed algorithm and MRVSS [17] for SNR=30dB. The proposed method almost match MRVSS in the convergence speed while MRVSS can go to lower steady state misadjusmenet. However this is coming at the expense of tracking ability where it is diminished as more tracking required. Our experiments show that regardless of when the sudden changes occur, as more changes repeated; the ability to track becomes worse and

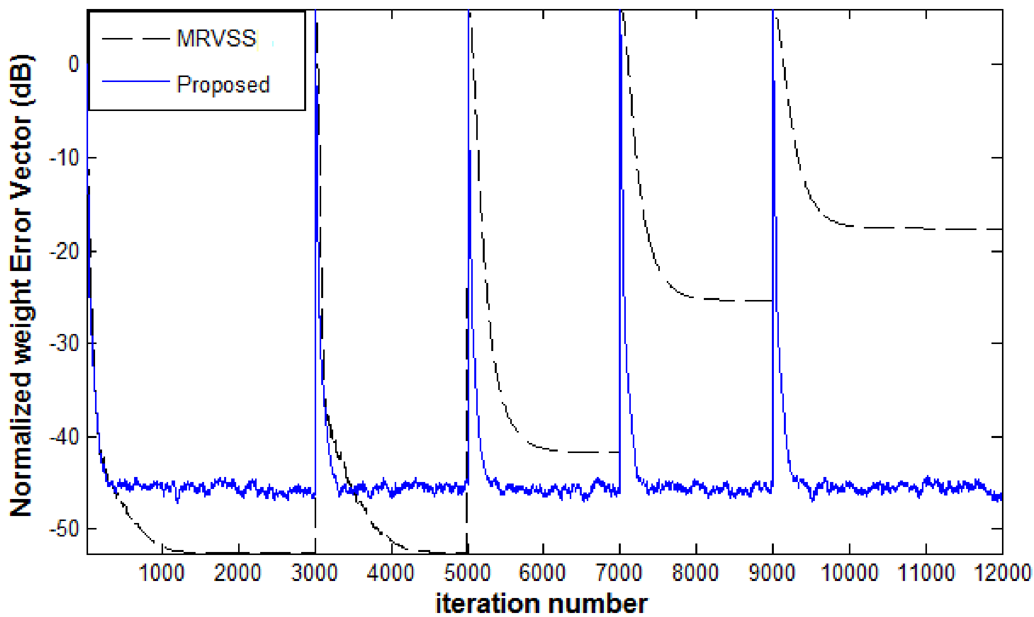

Figure 2.7: Adaptive curves of the proposed algorithm and MRVSS algorithm for white Gaussian input signal and SNR=30dB

worst, while the proposed algorithm has robustness ability to track changes regardless of how many times the changes are repeated. MRVSS algorithm depends on cumulative error which impact heavily by any sudden changes that lead to increase of the instantaneous error and ultimately increase the cumulative error as well.

2.9.4 Example 4

In this example, the same setup as in Rusus and Cowan "ECVSS" 2010 paper [7] is used, where the adaptive filter and the unknown system are both of order = 32, with impulse response $H(z) = \sum_{n=0}^{31} \rho^n Z^{-n}$ and $\rho = 0.80025$; however all the coefficients normalized by $|H(z)|$. The input signal is bipolar sequence from {1,-1} uniform zero-mean random, the desired signal is added to additive white

Gaussian noise with zero-mean, and the signal to noise ratio SNR = 30dB. The A parameter of ECVSS algorithm assigned to be $A = 35$ as we found from our experiments that this is the maximum value that can produce same steady state misadjustmnet as our proposed algorithm, while maintaining the fastest possible convergence speed. The $\mu_{max} = 0.008565$ and $\mu_{min} = 0.0008565$ as recomended on [7].

The proposed algorithm has a tuning parameter $A = 100$; and the same μ_{minLMS} as ECVSS, in order to achieve the same steady state misadjustmnet level. $\mu_{maxLMS} = \frac{1}{\text{Tr}[\mathbf{R}_\text{x}]}$. At iterations 4000 all the coefficients of the unknown system will be multiplied by -1 in order to test the algorithms ability to track the sudden changes.

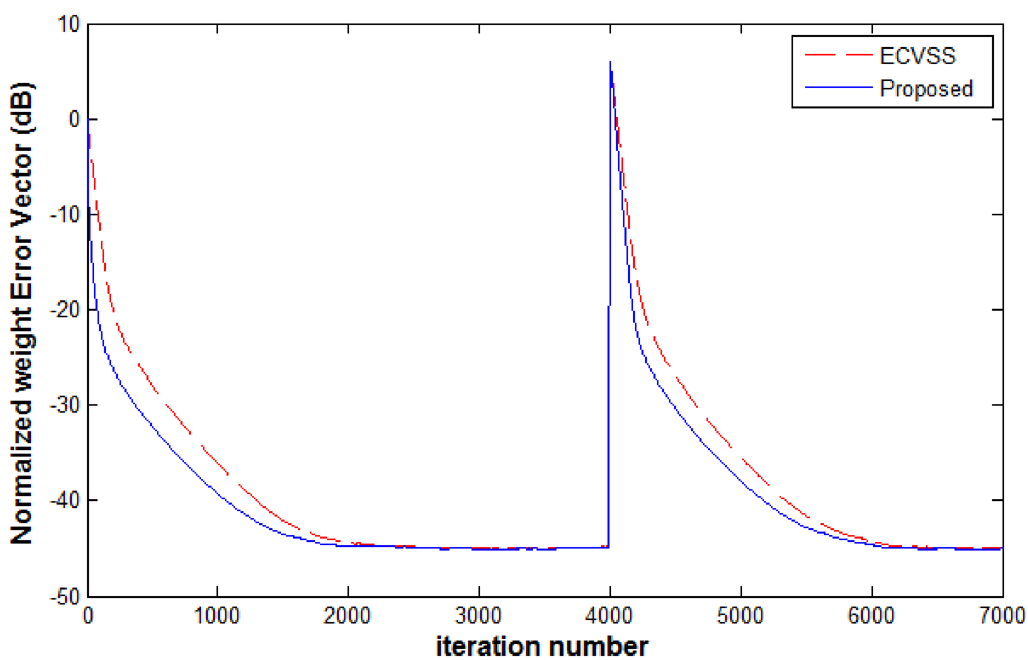

Figure 2.8: Adaptive curves of the proposed algorithm and ECVSS algorithm for white Gaussian input signal and SNR=30dB

Figure 2.8 shows the adaptive curves of the proposed algorithm and ECVSS [7] for SNR=30dB. The proposed method converges faster and tracking faster than ECVSS.

2.10 Conclusion

This chapter introduces a new class of variable step size based on minimizing the cost function of hyperbolic sine. The adaptation error square is the argument that drives the cost function. We try to maintain the simplicity and the robustness of the standard LMS, so the new algorithm has only one tuning parameter and required few computational effort more than the standard LMS; yet producing high performance that match and in some cases outperform those high computational cost; multi-tuning parameters VSS algorithms. The proposed algorithm also outperforms the exponential cost function and shows attractive results in both stationary and abrupt-change situations.

CHAPTER 3

HYPERBOLIC SINE OF 4TH ORDER ALGORITHM

3.1 Introduction

Least mean fourth (LMF) algorithms family use even powers of the instantaneous error as the cost function. This kind of algorithms deliver better compromise between speed of convergence and steady-state error, however there is a stability issue within this group. The stability of LMF algorithms family depends on the input signal power, the noise power, and the initial values of the adaptive filter weights while stability of least mean square (LMS) algorithms family depends only on the input signal power for a specific step size. Normalized LMF removes the dependency on the input signal power, however it doesn't solve the stability issue. LMF algorithm outperforms LMS in the sub-Gaussian environments. Its superior performance lies in the fast convergence speed and lower steady state error especially in low signal to noise ratio (SNR).

Stochastic gradient algorithms based on exponential of a chosen error have been proposed [11] [7]. The exponential cost function has a steeper surface comparing to a linear coordination of even moments. Error square exponential outperforms LMS [7] while fourth order error exponential outperforms LMF [11].

The mixed norm algorithms family employed different error norms in order to achieve better convergence performance. The combination of different norms deliver an extra degree of freedom, however it required optimization mixture between the norms based on prior information of the input signal and noise statistics. Some mixed norm algorithms removed that dependency and showed good performance based on logarithmic cost function [18].

In this work, we propose a new cost function , named least hyperbolic sine, which is non-linearly adapting the error fourth order as a driving argument. Accordingly a stochastic gradient based algorithm, named as Hyperbolic Sine error Fourth (HSF) algorithm, is derived. The new algorithm is classified into the VSS algorithm and has some gains in terms of the speed of convergence, adaption to the sudden changes, less computational cost as compared to the non-linear VSS algorithm in addition to a less required tuning parameters. In the sequel, the derivation of the algorithm is supported by a thorough analysis to figure out the required conditions for the convergence, and the excess mean steady state error (EMSE)

The chapter is organized as follows: the new cost function and the flow of the algorithm derivation is presented in section 3.2. Section 3.3 provides the steady state analysis throughout finding a closed form for the EMSE. Section 3.4 outlines the convergence analysis. Section 3.5 support the analytic finding through a course of computer simulations. Finally, the conclusions are stated in section 3.6.

The followings are the notations used in this chapter: $\mathbf{x}$ denotes a column vector, x is a scalar, $(.)^T$ is the transpose operator. E[.] is the mathematical expectation, and Tr[.] is the trace operator.

3.2 Algorithm Formulation

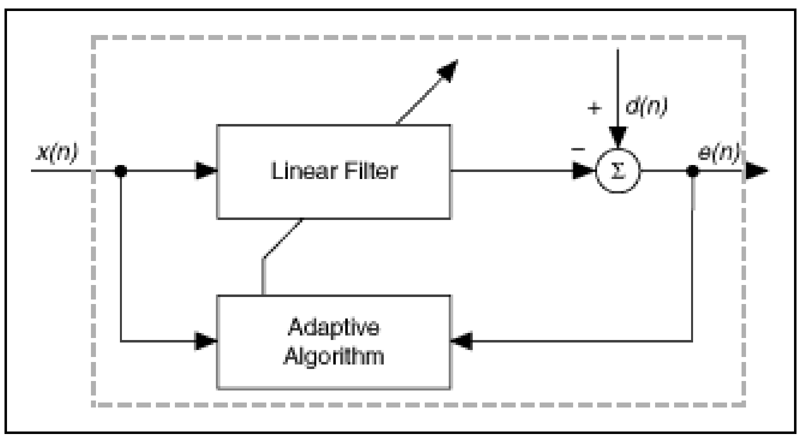

Figure 3.1: Principal diagram of adaptive filter

The considerable formulation is applied on system identification scheme, where the proposed algorithm work toward minimizing hyperbolic sine cost function. As shown in figure 3.1 the instantaneous error is defined as:

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k-1) \tag{3.1}$$

where the desired signal $d(k)$ is given by

$$d(k) = \mathbf{x}^T(k)\mathbf{w}^o + v(k) \tag{3.2}$$

$v(k)$ is a zero-mean independent random variable, and $\mathbf{w}^o$ is the optimal time-varying filter weight coefficients, while $\mathbf{w} = [w_0, w_1, \cdots, w_{M-1}]^T$, is the filter coefficients and M is the filter length. $(.)^T$ stands for the transpose operator, and $\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-M+1)]^T$ is the input signal vector.

The new cost function is hyperbolic sine with the error square argument, defined as

$$J(k) = \sinh(e^4(k)) \tag{3.3}$$

This is a convex and uni-modal function. Its gradient with respect to the filter coefficients yields

$$\Delta_w J(k) = -4e^3(k)\cosh(e^4(k))\mathbf{x}^T(k) \tag{3.4}$$

Where $\mathbf{x}(k)$ is the regression vector. To improve the convergence speed, we will introduce a scaling parameter ($A > 0$) [7], to scale the error to the power four in the argument of the hyperbolic sine; so that the modified cost function will be

$$J(k) = \frac{1}{4A}\sinh(Ae^4(k)) \tag{3.5}$$

while the gradient with the new cost function will be

$$\Delta_w J(k) = -e^3(k)\cosh(Ae^4(k))\mathbf{x}^T(k) \tag{3.6}$$

Hence, the stochastic recursive form of the coefficients estimate is given as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e^3(k)\cosh(Ae^4(k))\mathbf{x}(k) \tag{3.7}$$

We observe that hyperbolic cosine scales up the step size in case of high instantaneous error which will lead to fast convergence. However, this might lead to undesirable negative consequence on the algorithm stability. To utilize the large gradient property and maintain a bounded gradient to preserve the algorithm stability, we use the following selecting function [7]:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu_{min}.min\left[cosh(Ae^4(k)), \frac{\mu_{max}}{\mu_{min}}\right]e^3(k)\mathbf{x}(k) \qquad (3.8)$$

where μ_{max} and μ_{min} are the upper and lower bounds of μ respectively, while μ is the step size of the algorithm.

3.3 Steady State Analysis

It is investigated through deriving an analtical expression for the Excess Mean Squared Error (EMSE). The analysis in this section is based on the energy conservation relation framework [8],[13]. In addition to the wide sense stationary channel model assumption, the following standard assumptions [8] are introduced:

A1. There is a true values vector $\mathbf{w}^o$ leads to $d(k) = \mathbf{x}^T(k)\mathbf{w}^o + v(k)$

A2. The additive sequence of noise $\{v(k)\}$ is i.i.d. with identical variance $\sigma_v^2 = \mathrm{E}[(v^2(k)]$

A3. The sequence $v(i)$ is independent of the input vector $\mathbf{x}(j)$ for all i,j.

A4. The start up values w_{-1} is independent of all $\{d(j), \mathbf{x}(j), v(j)\}$

A5. The input signal auto-correlation matrix $\mathbf{R_x} = \mathrm{E}[\mathbf{x}(k)\mathbf{x}^T(k)] > 0$

A6. The random variables $\{d(k), \mathbf{x}(k), v(k)\}$ have zero means

According to the energy conservation framework [8], the steady state EMSE S is given by

$$\mathrm{S} = \frac{\mu}{2}\frac{\mathrm{N_s}}{\mathrm{D_s}}\mathrm{Tr}[\mathbf{R_x}] \tag{3.9}$$

where $\mathrm{Tr}[\mathbf{R_x}]$ is the trace of the auto-correlation matrix of the input signal. $\mathrm{N_s}$ is defined as:

$$\mathrm{N_s} = \mathrm{E}[f^2(e(k))] \tag{3.10}$$

and $\mathrm{D_s}$ is defined as:

$$\mathrm{D_s} = \frac{\mathrm{E}[e_a(k) \cdot f(e(k))]}{\mathrm{E}[e_a^2(k)]} \tag{3.11}$$

where $e_a(k)$ is apriori error defined as

$$e_a(k) = [\mathbf{w}^o - \mathbf{w}(k)]^T \mathbf{x}(k) \tag{3.12}$$

and $f(e)$, at the steady state zone, is defined from equation (3.8) as follows:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \underbrace{\mu_{min}}_{\mu} \underbrace{cosh(Ae^4(k))e^3(k)}_{f(e)} \mathbf{x}(k) \tag{3.13}$$

Accordingly, $\mathrm{N_s}$ becomes

$$\mathrm{N_s} = \mathrm{E}[e^6(k)\cosh^2(Ae^4(k))] \tag{3.14}$$

Due to working on the steady-state analysis, from now on we drop the time index k. The estimation error e can be written in term of apriori error and noise signal as ($e = e_a + v$) [8], accordingly, $\mathrm{N_s}$ becomes

$$\begin{aligned} \mathrm{N_s} = \mathrm{E}[e_a^6\cosh^2(Ae^4)] + 15\sigma_v^2\mathrm{E}[e_a^4\cosh^2(Ae^4) \\ +15\sigma_v^4\mathrm{E}[e_a^2\cosh^2(Ae^4) + \sigma_v^6\mathrm{E}[\cosh^2(Ae^4)] \end{aligned} \tag{3.15}$$

where σ_v^2 is the variance of the noise. By applying the Cauchy-Schwartz inequality, (3.15) is further simplified as

$$\begin{aligned} \mathrm{N_s} \leq \sqrt{\mathrm{E}[e^12_a]\cdot\mathrm{E}[\cosh^4(Ae^4)]} + 15\sigma_v^2\sqrt{\mathrm{E}[e_a^8]\cdot\mathrm{E}[\cosh^4(Ae^4)]} \\ +45\sigma_v^4\sqrt{\mathrm{E}[e_a^4]\cdot\mathrm{E}[\cosh^4(Ae^4)]} + 15\sigma_v^6\mathrm{E}[\cosh^2(Ae^4)] \end{aligned} \tag{3.16}$$

Furthermore, with the same prior conditions and via Jensen's inequality can be written in a closed form as:

$$\mathrm{N_s} \leq [15\sqrt{105}S^2\sigma_v^2 + 45\sqrt{3}\sigma_v^4 S + 15\sigma_v^6] \times \cosh^2(A\mathrm{E}[e_a^4 + 6e_a^2\sigma_v^2 + \sigma_v^4]) \tag{3.17}$$

defining $\mathrm{S} \triangleq \lim_{k\to\infty} \mathrm{E}[e_a^2(k)]$ [8], lead to

$$\mathrm{N_s} \leq [15\sqrt{105}S^2\sigma_v^2 + 45\sqrt{3}\sigma_v^4 S + 15\sigma_v^6] \times \cosh^2(A[3S^2 + 6S\sigma_v^2 + 3\sigma_v^4]) \tag{3.18}$$

In a similar way, $\mathrm{D_s}$ in (3.9) can be written as follows:

$$\mathrm{D_s} = \frac{\mathrm{E}[e_a \cdot e^3 \cdot \cosh(Ae^4)]}{\mathrm{E}[e_a^2]} \tag{3.19}$$

Substitute $e(k) = e_a(k) + v(k)$ into (3.19) gives

$$\mathrm{D_s} = \frac{\mathrm{E}[(e_a^4 + 3e_a^3 v + 3e_a^2 v^2 + e_a v^3)\cosh(Ae^4)]}{\mathrm{E}[e_a^2]} \tag{3.20}$$

Based on the assumptions (**A**1-**A**6), one can easily shows that e_a is a zero-mean Gaussian variable and independent of the noise v [8], hence

$$\mathrm{D_s} = \frac{\mathrm{E}[(e_a^4 + 3e_a^2 v^2)\cosh(Ae^2)]}{\mathrm{E}[e_a^2]} \tag{3.21}$$

Again, applying Cauchy-Schwartz inequality gives

$$\mathrm{D_s} \leq \frac{\sqrt{\mathrm{E}[e_a^8 + 6e_a^6 v^2 + 9e_a^4 v^4]\mathrm{E}[\cosh^2(Ae^4)]}}{\mathrm{E}[e_a^2]} \tag{3.22}$$

Also, applying Jensen's inequality to (3.22) leads to

$$\mathrm{D_s} \leq (\sqrt{105}S + \sqrt{90S\sigma_v^2} + 9\sigma_v^2)\ \cosh(A[3S^2 + 6S\sigma_v^2 + 3\sigma_v^4]) \tag{3.23}$$

Eventually, using Taylor series expansion of the hyperbolic cosine function, an approximate closed form expression for the steady state EMSE in (3.9) is written as

$$\mathrm{S} = \frac{7.5\mu_{min}Tr(\mathbf{R_x}).\sigma_v^4[1 + \frac{9}{2}A^2\sigma_v^8]}{9 - 0.5\mu_{min}Tr(\mathbf{R_x})[45\sqrt{3}\sigma_v^2 + \zeta A^2\sigma_v^{10}]} \tag{3.24}$$

Where $\zeta = \frac{45}{2} \times 9\sqrt{3} + (\frac{36}{2} \times 15)$. The following remarks are outlined out of the derived EMSE in (3.24):

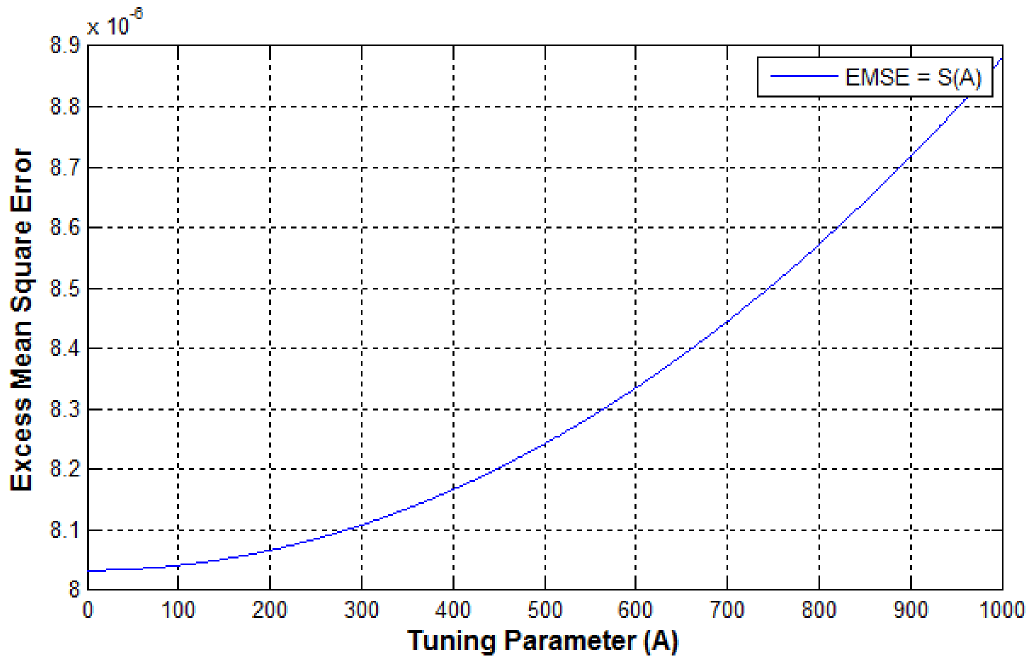

Figure 3.2: Excess Mean Square Error versus the tuning parameter (A)

- The EMSE depends on the even powers of the noise power.

- The EMSE is also depending on the tuning parameter A and it is usually coupled with the high order even power of the noise variance σ_v^2. To demonstrate the consequence of A on the proposed algorithm performance, the following experiment is carried out.

  In figure 3.2, the SNR $=$ 20dB, $\mu = 0.003$, and $\text{Tr}[\mathbf{R_x}] = 32$, the EMSE is shown to be increasing as the tuning parameter does. This creates a supplementary issue at the implementation. Another impact for large A on the algorithm performance is that it causes a large fluctuation of the EMSE around its average value.

- If the tuning parameter A increases such that $\cosh(Ae^4) > \mu_{max}/\mu_{min}$ for all e^4, then the algorithm will behave like the LMF algorithm with a fixed

$\mu = \mu_{max}$ all the time.

- It can be shown from (3.24) that the EMSE of the proposed algorithm becomes equal to the EMSE of the LMF by setting $A = 0$. Henceforth, we will write μ_{min} as μ_{minLMF} and μ_{max} as μ_{maxLMF} to reflect the relationship between our proposed algorithm and the standard LMF algorithm.

3.4 Convergence Analysis

The update equation (3.8) belongs to the general update equation of the error adaptive algorithm [8]:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu f[e(k)]\mathbf{x}(k) \tag{3.25}$$

where

$$f[e(k)] = \frac{1}{\mu}\mu_{minLMF}.min\left[\cosh(Ae^4), \frac{\mu_{maxLMF}}{\mu_{minLMF}}\right] e^3(k) \tag{3.26}$$

Due to the lack of differentiability of the min function in (3.26), $f(e)$ first derivative doesn't always exist at any point (Appendix 5.2), and hence $f[e(k)]$ doesn't have power Taylor series expansion running to infinity [14]. However, the approximation

$$\mathbf{w}(k+1) \approx \mathbf{w}(k) + \mu\{f[e(k)]\mathbf{x}(k) - f'[e(k)]\mathbf{x}(k)\mathbf{w}^T(k)\mathbf{x}(k) + \frac{1}{2}f''[e(k)]\mathbf{x}(k)[\mathbf{w}^T(k)\mathbf{x}(k)]^2\} \tag{3.27}$$

holds in every point except when $e(k) = \pm\delta$, where

$$\delta = \sqrt[4]{\frac{1}{A}\cosh^{-1}\left(\frac{\mu_{maxLMF}}{\mu_{minLMF}}\right)}$$

Therefore, we carry on the analysis under the assumption that the noise values are very rare to become equal to δ.

3.4.1 Convergence speed

According to [15], for a small step size μ, the time-constant of the proposed algorithm linked with $\lambda_i(\mathbf{R_x})$ (the i^{th} eigenvalue of the auto-correlation matrix $\mathbf{R_x}$) is given by

$$\tau_i = \frac{1}{\mu \mathrm{E}[f'[e(k)]]\lambda_i(\mathbf{R_x})} \tag{3.28}$$

Now, assuming $e(k) \neq \pm\delta$, then

$$f(e) = \begin{cases} \frac{\mu_{minLMF}}{\mu} e^3 \cosh(Ae^4); & |e| < \delta \\ \frac{\mu_{maxLMF}}{\mu} e^3; & |e| > \delta \end{cases} \tag{3.29}$$

Accordingly,

$$f'(e) = \begin{cases} \frac{\mu_{minLMF}}{\mu} \left(3e^2 \cosh(Ae^4) + 4Ae^6 \sinh(Ae^4)\right) ; |e|< \delta \\ \frac{\mu_{maxLMF}}{\mu}(3e^2); \qquad |e|> \delta \end{cases} \tag{3.30}$$

Eventually, based on (3.28) and (3.30), the proposed algorithm have the following two cases:

1. if $A < \frac{1}{e^4} \cosh^{-1}\left(\frac{\mu_{maxLMF}}{\mu_{minLMF}}\right)$ then

$$\tau_i = \frac{1}{\mu_{minLMF} \mathrm{E}[3e^2 \cosh(Ae^4) + 4Ae^6 \sinh(Ae^4)]\lambda_i(\mathbf{R_x})} \tag{3.31}$$

2. if $A > \frac{1}{e^4} \cosh^{-1}\left(\frac{\mu_{maxLMF}}{\mu_{minLMF}}\right)$ then

$$\tau_i = \frac{1}{\mu_{maxLMF} \mathrm{E}[3e^2]\lambda_i(\mathbf{R_x})} \tag{3.32}$$

which match LMF case for $\mu = \mu_{maxLMF}$.

Since τ in the first case is smaller than LMF time-constant; the convergence of the proposed algorithm will be faster than the convergence of LMF and some LMF variants. If the tuning parameter A is not properly chosen then $A < \frac{1}{e^4} \cosh^{-1}\left(\frac{\mu_{maxLMF}}{\mu_{minLMF}}\right)$ may not happen and the proposed algorithm will work like standard LMF with firmed $\mu = \mu_{maxLMF}$ all the time.

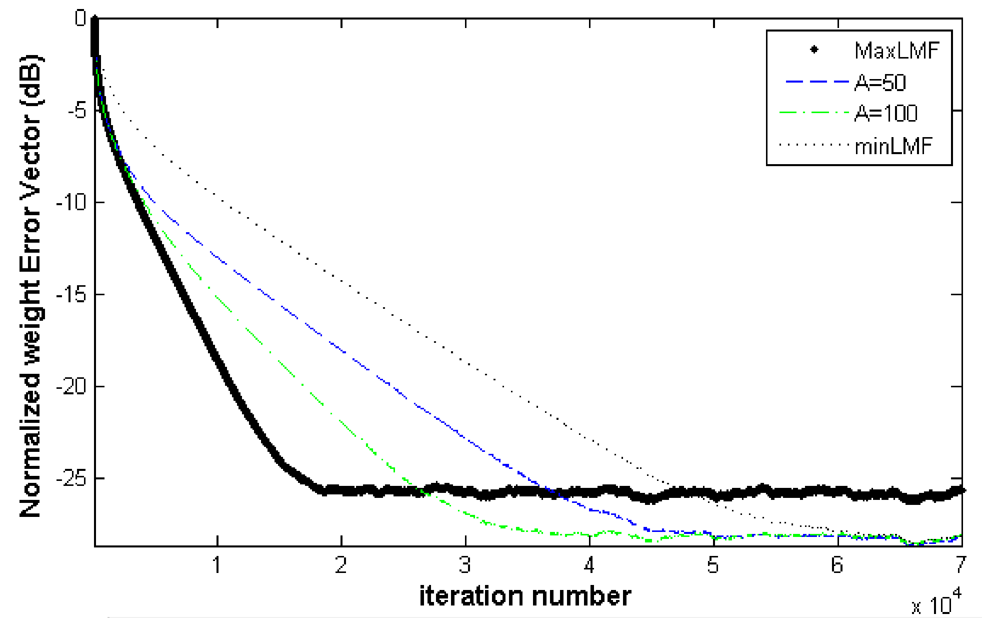

Figure 3.3: Adaptive curves for different tuning parameter values, SNR=10dB

Figure 3.3 demonstrates the consequence of varying A over the convergence speed, the experiment setup has a white Gaussian input signal, and additive white Gaussian noise with SNR = 10 dB. The convergence speed is noticed to be improved with the tuning parameter, and this is achieved with a maintained consistent steady state error. In particular, when A ranges between 10 and 50. Furthermore, the impact of A has a compromise consequence on both the speed of convergence and the steady state error, especially for $A > 300$, whereby the EMSE increases significantly as A moves beyond 300. In general, for a certain range around A, we obtain a quite similar result.

3.4.2 Mean Behavior

As common in all gradient decent algorithms, the value of the step-size is critical. To guarantee the stability; the step size value should satisfy some certain bounds.

Rewrite equation (3.8) as follows:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e^3(k)\mathbf{x}(k) \tag{3.33}$$

where $\mu(k)$ is given as

$$\mu(k) = \mu_{minLMF} \min\left[\cosh(Ae^4(k)), \frac{\mu_{maxLMF}}{\mu_{minLMF}}\right] \tag{3.34}$$

It is sufficient mentioning that statistical mean value of $\mu(k)$, (i.e., $\mathrm{E}[\mu(k)]$), must satisfy the following condition [8][19]

$$0 < \mathrm{E}[\mu(k)] < \frac{2}{3\sigma_v^2\lambda_{max}(\mathbf{R_x})} \tag{3.35}$$

Based on (3.34), we have the following two cases:

1. if $\cosh(Ae^4(k)) < \left(\frac{\mu_{maxLMF}}{\mu_{minLMF}}\right)$ then

$$\mu(k) = \mu_{minLMF} \cdot cosh(Ae^4(k)) \tag{3.36}$$

Taking expectations of (3.36), and using the Taylor series expansion, we can rewrite $\mathrm{E}[\mu(k)]$ as

$$\mathrm{E}[\mu(k)] \geq \mu_{minLMF}\left\{1+\frac{9(70)}{2}A^2F^2(k)\sigma_v^4+\frac{15(28)}{2}A^2F(k)\sigma_v^6+\frac{105}{2}A^2\sigma_v^8\right\} \tag{3.37}$$

Where $F = \mathrm{E}[e_a^2]$. At steady state, one can drop the high order powers of S. This implies to a new bound on μ as follows:

$$0 < \mu_{minLMF} < \frac{2}{3\sigma_v^2\ \lambda_{max}(\mathbf{R_x})\{1+315A^2\mathrm{S}\sigma_v^6+52.5A^2\sigma_v^8\}} \tag{3.38}$$

2. if $\cosh(Ae^4(k)) > \left(\frac{\mu_{maxLMF}}{\mu_{minLMF}}\right)$ then

$$\mu(k) = \mu_{maxLMF} \tag{3.39}$$

and the new bound will be as follows:

$$0 < \mu_{minLMF} < \mu_{maxLMF} < \frac{2}{3\sigma_v^2\lambda_{max}(\mathbf{R_x})} \tag{3.40}$$

Which match the LMF case, however we need to choose the μ_{minLMF} first as a lower bound than μ_{maxLMF}.

3.5 Simulation Results

We will run the experiments based on system identification scenarios. The order of the unknown system will be matching the order of the adaptive filter and both of them representing FIR system. We start with zeros intial values for the adaptive filter coefficients. Adding Gaussian noise sequence $v(k)$ to the output of the unknown system. The variance σ_v^2 of the noise sequence $v(k)$ will be selected in each experiment to reflect the desired signal to noise ration (SNR). Each experiments run 200 times and averaged to have average results. The quantitative performance measure is the normalized weight error squared vector in dB (also known as misadjustmnet [8]), which is mathematically calculated as follows:

$$\mathcal{M} = 10\log_{10}\left(\frac{\mathrm{E}[||\mathbf{w}^o - \mathbf{w}(k)||^2]}{||\mathbf{w}^o||^2}\right)$$

where $\mathbf{w}^o = [w_0^o, w_1^o, \cdots w_{M-1}^o]^T$ the true values of the unknown system/channel tapes weight.

In the ensuing, to achieve fair comparison, we will run different experiments to compare the proposed HSF algorithm with others based on the simulation environment used in the reference of the compared algorithm.

3.5.1 Example 1

In this example, we would like to compare our new proposed algorithm with the standard LMF. The adaptive filter and the unknown system are both of order = 5, The input signal is uniform zero-mean random bipolar sequence from {1,-1}, the desired signal is corrupted by sub-Gaussian noise, and the signal to noise ratio SNR = 10dB. $\mu = 0.001$. The proposed algorithm will have a scaling parameter $A = 100$ and $\mu_{max} = 0.01$ while μ_{min} will be like the LMF μ, so both of them will have the same steady state $\mathcal{M}$.

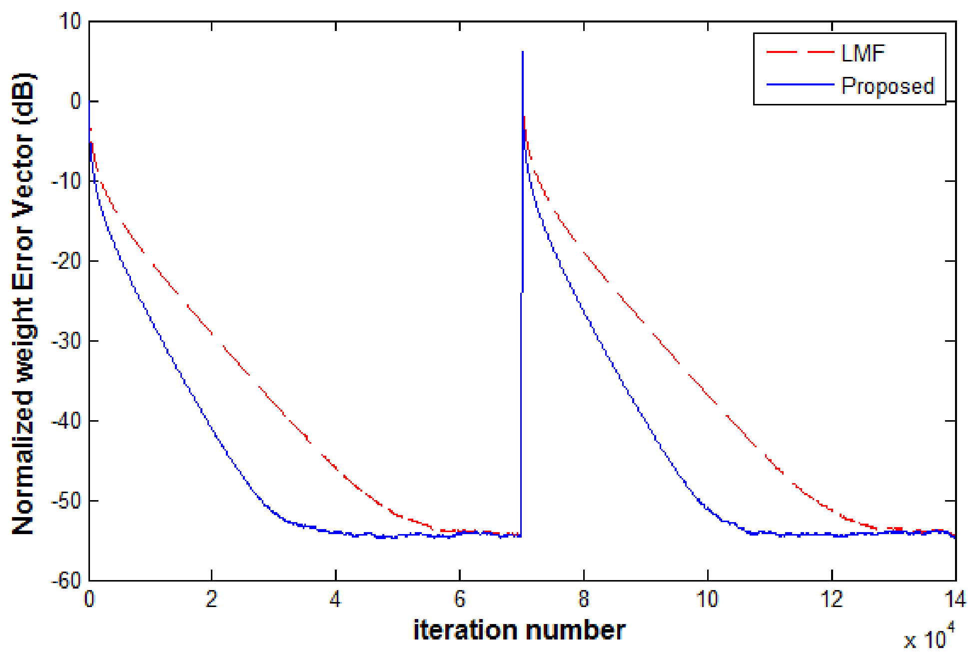

Figure 3.4: Adaptive curves of the proposed algorithm and LMF algorithm for sub-Gaussian noise; bipolar input signal and SNR=10dB

Figure 3.4 depicts the adaptive curve of the proposed algorithm and the LMF algorithm. As can be seen from the figure, the proposed algorithm converges faster to the same steady state $\mathcal{M}$.

3.5.2 Example 2

In this example, we follow the same setup as in Umair, Asad and Zerguine EELMF 2014 [11], where the adaptive filter and the unknown system are both of order = 5, the input signal is bipolar $\{-1, 1\}$. The desired signal is corrupted by sub-Gaussian noise with zero-mean. SNR = 10dB. From our experemants, the maximum scaling parameter for EELMF [11] that we can use while maintaining the stability is k = 0.14. The step size is $\mu = 0.001$. Our proposed algorithm has a tuning parameter A = 100 and μ_{minLMF} chosen to be like EELMF to give the same steady state $\mathcal{M}$, while $\mu_{maxLMF} = 0.01$.

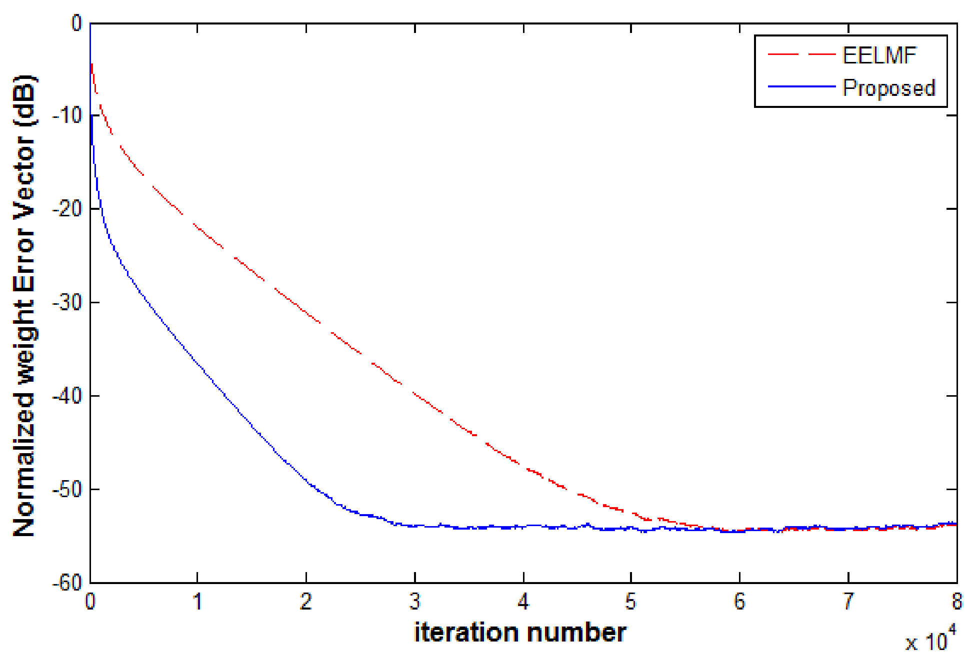

Figure 3.5: Adaptive curves of the proposed algorithm and EELMF algorithm for sub-Gaussian noise; bipolar input signal, SNR=10dB, and EELMF k=0.14

Figure 3.5 depicts the adaptive curve of the proposed algorithm and EELMF [11]. As can be seen from the figure, the proposed algorithm converges faster to the same steady state $\mathcal{M}$.

Tracking ability is a challenging feature within LMF family. At start-up, the initial values of the filter coefficients are zeros, so the instantaneous error has a certain value while if there is a sudden change in the coefficients while the filter running on the steady state zone (like multiply the filter coefficients by (-1)), the instantaneous error value will be duplicated comparing to the initial instantaneous error at start-up. This could drive the algorithm to diverge.

In order to test the tracking ability, we run the same setup like the previous experiment. For EELMF [11], we use k= 0.009 (this is the maximum value we found from our experiments that maintain the stability of EELMF algorithm) while the proposed algorithm didn't require any modification. At iteration 7000. we multiply all the filter coefficients by -1.

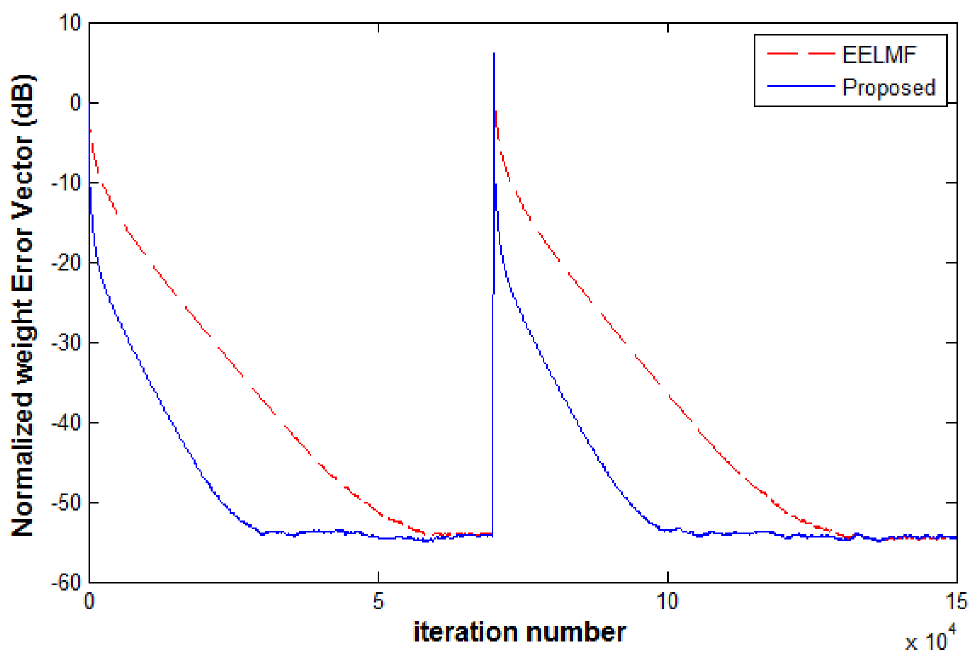

Figure 3.6: Adaptive curves of the proposed algorithm and EELMF algorithm for sub-Gaussian noise, bipolar input signal, SNR=10dB, and EELMF k=0.009

Figure 3.6 depicts the adaptive curve of the proposed algorithm and EELMF

[11]. As can be seen from the figure, the proposed algorithm converges faster and shows faster tracking ability and return back to the same steady state $\mathcal{M}$ level.

3.5.3 Example 3

In this example, the counter part algorithm (LMS-LMF Type II, Zerguine, Cowan, Bettayeb, 1997) [10] is using two different μ which make it similar to our proposed algorithm. The adaptive filter and the unknown system are both of order = 16, the input signal is white-Gaussian with zero mean. The desired signal is corrupted by sub-Gaussian noise with zero-mean. SNR = 10dB. For LMS-LMF type two algorithm we choose $\mu_1 = 0.03$ and $\mu_2 = 0.001$ as those values showed better performance in our experiments. The proposed algorithm tuning factor A =100 and $\mu_{maxLMF} = 0.002$ and $\mu_{minLMF} = 0.001$.

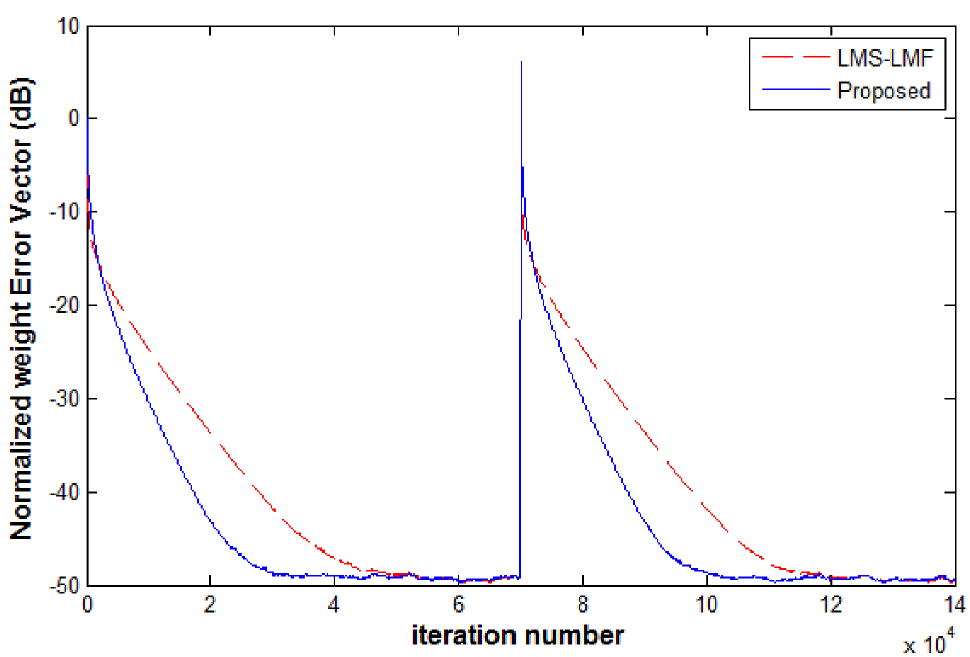

Figure 3.7: Adaptive curves of the proposed algorithm and LMS-LMF algorithm for sub-Gaussian noise, white-Gaussian input signal, and SNR=10dB

Figure 3.7 depicts the adaptive curve of the proposed algorithm and LMS-LMF [10]. As can be seen from the figure, the proposed algorithm converges faster and shows faster tracking ability and return back to the same steady state $\mathcal{M}$ level.

3.6 Conclusion

This chapter introduces a new class of stochastic gradient variable step size algorithm. This is achieved by introducing a new hyperbolic sine cost function. The adaptation error fourth is the argument that drives the cost function. The new HSF algorithm maintains the simplicity and the robustness of the standard LMF, by having one tuning parameter only with few computational cost. However, the proposed HSF enjoys an improved performance over other LMF family algorithms counterparts. In particular, the proposed algorithm outperforms the exponentiation-based algorithms and shows attractive results in both the stationary and the abrupt-change scenarios.

CHAPTER 4

MIXED NORM HYPERBOLIC SINE ALGORITHM

4.1 Introduction

Mixed norm adaptive algorithms represent a new family that is based on mixing error norms. So, they combine the advantages of different error forms. Least Mean Mixed Norm (LMMN) algorithm [8] combined the relative strong stability as well as the well-behavior from the least mean square (LMS) algorithm and fast convergence and lower steady state error from least mean fourth (LMF) algorithm.

The combination of different norms deliver an extra degree of freedom, however it required optimization mixture between the norms based on prior information of the input signal and noise statistics. Some mixed norm algorithms removed that dependancy and showed good performance based on logarithmic cost function

[18].

Direct combination between LMS and LMF has been introduced by [10] which gain a superior performance due to mixing the two algorithms without needs to prior knowledge on noise power nor distribution. Indirect combination based on weight parameter is the well-known method, however the mix is not always optimal and depends on the noise distribution environment.

In this work, we propose a new cost function , named as a Least Hyperbolic Sine (LHS), which is non-linearly adapting the second and forth error moments as a driving argument. Accordingly a stochastic gradient based algorithm, named as Hyperbolic Sine error Mix (HSM) algorithm, is derived. The new algorithm is classified into the VSS algorithm and has some gains in terms of the speed of convergence, adaption to the sudden changes, less computational cost as compared to the non-linear VSS algorithm in addition to a less required tuning parameters. In the sequel, the derivation of the algorithm is supported by a thorough analysis to figure out the required conditions for the convergence, and the excess mean steady state error (EMSE)

The chapter is organized as follows: the new cost function and the flow of the algorithm derivation and the generic step size are presented in section 4.2. Section 4.3 provides the steady state analysis throughout finding a closed form for the EMSE. Section 4.4 outlines the convergence analysis. Section 4.5 supports the analytic finding through a course of computer simulations. Finally, the conclusions are stated in section 4.6.

The followings are the notations used in this chapter: $\mathbf{x}$ denotes a column vector, x is a scalar, $(.)^T$ is the transpose operator. E[.] is the mathematical expectation, and Tr[.] is the trace operator.

4.2 Algorithm Formulation

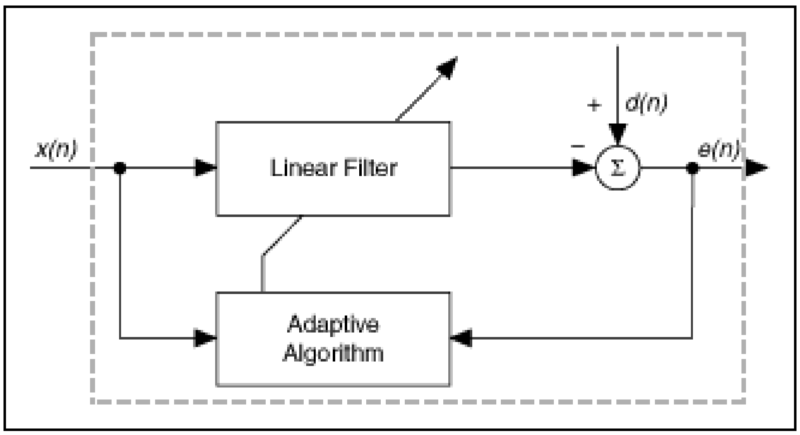

Figure 4.1: Principal diagram of adaptive filter

4.2.1 Algorithm derivation

The considerable formulation is applied on system identification scheme, where the proposed algorithm work toward minimizing hyperbolic sine cost function.
As shown in figure 4.1 the instantaneous error defined as:

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k-1) \tag{4.1}$$

where the desired signal $d(k)$ is given by

$$d(k) = \mathbf{x}^T(k)\mathbf{w}^o + v(k) \tag{4.2}$$

$v(k)$ is a zero-mean independent random variable, and $\mathbf{w}^o$ is the optimal time-varying filter weight coefficients, while $\mathbf{w} = [w_0, w_1, \cdots, w_{M-1}]^T$, is the filter coefficients and M is the filter length. $(.)^T$ stands for the transpose operator, and $\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-M+1)]^T$ is the input signal vector.

The new cost function is hyperbolic sine with the error raised to the power four argument, defined as

$$J(k) = \sinh(e^4(k)) \tag{4.3}$$

This is a convex and uni-modal function. Its gradient with respect to the filter coefficients yields

$$\Delta_w J(k) = -4e^3(k)\cosh(e^4(k))\mathbf{x}^T(k) \tag{4.4}$$

Where $\mathbf{x}(k)$ is the regression vector. To improve the convergence speed, we will introduce a scaling parameter ($A > 0$) [7], to scale the error fourth order, in the argument of the hyperbolic sine; so that the modified cost function will be

$$J(k) = \frac{1}{4A}\sinh(Ae^4(k)) \tag{4.5}$$

while the gradient with the new cost function will be

$$\Delta_w J(k) = -e^3(k)\cosh(Ae^4(k))\mathbf{x}^T(k) \tag{4.6}$$

Hence, the stochastic recursive form of the coefficients estimate given as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e^3(k)\cosh(Ae^4(k))\mathbf{x}(k) \tag{4.7}$$

We observe that hyperbolic cosine scales up the step size in case of high instantaneous error which will lead to fast convergence. However, this might lead to undesirable negative consequence on the algorithm stability. To utilize the large gradient property and maintain a bounded gradient to preserve the algorithm stability, we use the following selecting function [7]:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu_{min}.min\left[cosh(Ae^4(k)), \frac{\mu_{max}}{\mu_{min}}\right]e^3(k)\mathbf{x}(k) \tag{4.8}$$

where μ_{max} and μ_{min} are the upper and lower bounds of μ respectively.

4.2.2 Generic upper bound of μ :

In this new proposed algorithm, we are introducing μ_{max} as a generic value rather than a fixed number. One can evaluate μ_{max} at each iteration as follows:

$$\mu_{max} = \frac{1}{\mathrm{Tr}[\mathbf{R}_x] + \epsilon} \cdot \frac{1}{1 + e.^2(k)} \tag{4.9}$$

where $\epsilon \ll 1$ is used to avoid the case of zero denominator. By using this generic value, we guarantee the stability of the algorithms and improve the convergence speed. Stability is an issue for LMF family algorithms and by introducing error normalization, we equipped the proposed algorithm with the stability of LMS and also enjoy the lower steady state error achieved by LMF algorithms family. This is confirmed throughout by experimental validation. Whereby, we run the simulation for the proposed algorithm, one time with a fixed maximum step size and another with the generic maximum step size. This kind of generic value allows the algorithm to adapt against the abrupt changes in the signal power. This improvement comes at the expense of little increase on the computational cost: $2M$ multiplications and M additions per iteration.

4.3 Steady State Analysis

Steady atate analysis is investigated through deriving an analtical expression for the Excess Mean Squared Error (EMSE). The analysis in this section is based on the energy conservation relation framework [8]-[13]. In addition to the wide sense stationary channel model assumption, the following standard assumptions [8] are introduced:

A1. There is a true values vector $\mathbf{w}^o$ leads to $d(k) = \mathbf{x}^T(k)\mathbf{w}^o + v(k)$

A2. The additive sequence of noise $\{v(k)\}$ is i.i.d. with variance $\sigma_v^2 = \mathrm{E}[(v^2(k)]$

A3. The sequence $v(i)$ is independent of the input vector $\mathbf{x}(j)$ for all i,j.

A4. The start up values w_{-1} is independent of all $\{d(j), \mathbf{x}(j), v(j)\}$

A5. The input signal auto-correlation matrix $\mathbf{R_x} = \mathrm{E}[\mathbf{x}(k)\mathbf{x}^T(k)] > 0$

A6. The random variables $\{d(k), \mathbf{x}(k), v(k)\}$ have zero means

According to the energy conservation framework [8], the steady state EMSE S is given by

$$\mathrm{S} = \frac{\mu}{2}\frac{\mathrm{N_s}}{\mathrm{D_s}}\mathrm{Tr}[\mathbf{R_x}] \tag{4.10}$$

where $\mathrm{Tr}[\mathbf{R_x}]$ is the trace of the auto-correlation matrix of the input signal. $\mathrm{N_s}$ is defined as:

$$\mathrm{N_s} = \mathrm{E}[f^2(e(k))] \tag{4.11}$$

and $\mathrm{D_s}$ is defined as:

$$\mathrm{D_s} = \frac{\mathrm{E}[e_a(k) \cdot f(e(k))]}{\mathrm{E}[e_a^2(k)]} \tag{4.12}$$

where $e_a(k)$ is the apriori error defined as

$$e_a(k) = [\mathbf{w}^o - \mathbf{w}(k)]^T \, \mathbf{x}(k) \tag{4.13}$$

and $f(e)$, at the steady state zone, is defined from equation (4.8) as follows:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \underbrace{\mu_{min}}_{\mu} \underbrace{cosh(Ae^4(k))e^3(k)}_{f(e)} \mathbf{x}(k) \tag{4.14}$$

Accordingly, $\mathrm{N_s}$ becomes

$$\mathrm{N_s} = \mathrm{E}[e^6(k) \cosh^2(Ae^4(k))] \tag{4.15}$$

Due to working on the steady-state analysis, from now on we drop the time index k. The estimation error e can be written in term of apriori error and noise signal as $(e = e_a + v)$ [8], accordingly, $\mathrm{N_s}$ becomes

$$\begin{aligned}\mathrm{N_s} = \mathrm{E}[e_a^6 \cosh^2(Ae^4)] + 15\sigma_v^2 \mathrm{E}[e_a^4 \cosh^2(Ae^4)\\ +15\sigma_v^4 \mathrm{E}[e_a^2 \cosh^2(Ae^4) + \sigma_v^6 \mathrm{E}[\cosh^2(Ae^4)]\end{aligned} \tag{4.16}$$

where σ_v^2 is the variance of the noise. A step more further toward simplicity by applying Cauchy-Schwartz inequality, (4.16) is further simplified as

$$\begin{aligned}\mathrm{N_s} \leq \sqrt{\mathrm{E}[e^12_a] \cdot \mathrm{E}[\cosh^4(Ae^4)]} + 15\sigma_v^2 \sqrt{\mathrm{E}[e_a^8] \cdot \mathrm{E}[\cosh^4(Ae^4)]}\\ +45\sigma_v^4 \sqrt{\mathrm{E}[e_a^4] \cdot \mathrm{E}[\cosh^4(Ae^4)]} + 15\sigma_v^6 \mathrm{E}[\cosh^2(Ae^4)]\end{aligned} \tag{4.17}$$

Furthermore, with the same prior assumptions and by applying Jensen's inequality, we can rewrite the expression in a closed form as:

$$\mathrm{N_s} \leq [15\sqrt{105}S^2\sigma_v^2 + 45\sqrt{3}\sigma_v^4 S + 15\sigma_v^6] \times \cosh^2(A\mathrm{E}[e_a^4 + 6e_a^2\sigma_v^2 + \sigma_v^4]) \tag{4.18}$$

defining $\mathrm{S} \triangleq \lim_{k\to\infty} \mathrm{E}[e_a^2(k)]$ [8], lead to

$$\mathrm{N_s} \leq [15\sqrt{105}S^2\sigma_v^2 + 45\sqrt{3}\sigma_v^4 S + 15\sigma_v^6] \times \cosh^2(A[3S^2 + 6S\sigma_v^2 + 3\sigma_v^4]) \quad (4.19)$$

In a similar way, $\mathrm{D_s}$ in (4.10) can be written as follows:

$$\mathrm{D_s} = \frac{\mathrm{E}[e_a \cdot e^3 \cdot \cosh(Ae^4)]}{\mathrm{E}[e_a^2]} \quad (4.20)$$

Substitute $e(k) = e_a(k) + v(k)$ into (4.20) gives

$$\mathrm{D_s} = \frac{\mathrm{E}[(e_a^4 + 3e_a^3 v + 3e_a^2 v^2 + e_a v^3)\cosh(Ae^4)]}{\mathrm{E}[e_a^2]} \quad (4.21)$$

Based on the assumptions (**A**1-**A**6), one can easily shows that e_a is a zero-mean Gaussian variable and independent of the noise v [8], hence

$$\mathrm{D_s} = \frac{\mathrm{E}[(e_a^4 + 3e_a^2 v^2)\cosh(Ae^2)]}{\mathrm{E}[e_a^2]} \quad (4.22)$$

Again, applying Cauchy-Schwartz inequality gives

$$\mathrm{D_s} \leq \frac{\sqrt{\mathrm{E}[e_a^8 + 6e_a^6 v^2 + 9e_a^4 v^4]\mathrm{E}[\cosh^2(Ae^4)]}}{\mathrm{E}[e_a^2]} \tag{4.23}$$

Also, applying Jensen's inequality to (4.23) leads to

$$\mathrm{D_s} \leq (\sqrt{105}S + \sqrt{90S\sigma_v^2} + 9\sigma_v^2)\ \cosh(A[3S^2 + 6S\sigma_v^2 + 3\sigma_v^4]) \tag{4.24}$$

Eventually, using Taylor series expansion of the hyperbolic cosine function, an approximate closed form expression for the steady state EMSE in (4.10) is written as

$$\mathrm{S} = \frac{7.5\mu_{min} Tr(\mathbf{R_x}).\sigma_v^4[1 + \frac{9}{2}A^2\sigma_v^8]}{9 - 0.5\mu_{min} Tr(\mathbf{R_x})[45\sqrt{3}\sigma_v^2 + \zeta A^2 \sigma_v^{10}]} \tag{4.25}$$

Where $\zeta = \frac{45}{2} \times 9\sqrt{3} + (\frac{36}{2} \times 15)$. The following remarks are outlined out of the derived EMSE in (4.25):

- The EMSE depends on the even powers of the noise power.
- The EMSE is also depends on the tuning parameter A and it is usually coupled with the high order even power of the noise variance σ_v^2. To demonstrate the consequence of A on the proposed algorithm performance, the following experiment is carried out.

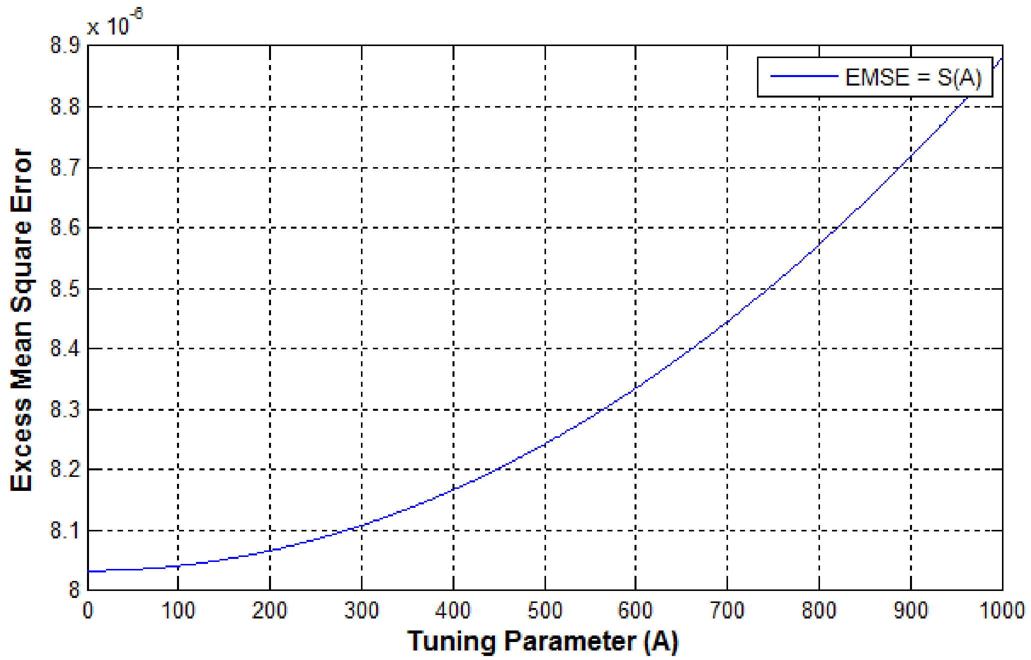

Figure 4.2: Excess Mean Square Error versus the tuning parameter (A)

In figure 4.2, the SNR = 20dB, $\mu = 0.003$, and $\text{Tr}[\mathbf{R_x}] = 32$, the EMSE is shown to be increasing as the tuning parameter does. This creates a supplementary issue at the implementation. Another impact for large A on the algorithm performance is that it causes a large fluctuation of the EMSE around its average value.

- If the tuning parameter A increases such that $\cosh(Ae^4) > \mu_{max}/\mu_{min}$ for all e^4, then the algorithm will behave like the LMF algorithm with a fixed $\mu = \mu_{max}$ all the time.

4.4 Convergence Analysis

The update equation (4.8) belongs to the general update equation of the error adaptive algorithm [8]:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu f[e(k)]\mathbf{x}(k) \tag{4.26}$$

where

$$f[e(k)] = \frac{1}{\mu}\mu_{min}.min\left[\cosh(Ae^4), \frac{\mu_{max}}{\mu_{min}}\right] e^3(k) \tag{4.27}$$

Due to the lack of differentiability of the min function in (4.27), $f(e)$ first derivative doesn't always exist have at any point (Appendix 5.2), and hence $f[e(k)]$ doesn't have power Taylor series expansion running to infinity [14]. However, the approximation

$$\mathbf{w}(k+1) \approx \mathbf{w}(k) + \mu\{f[e(k)]\mathbf{x}(k) - f'[e(k)]\mathbf{x}(k)\mathbf{w}^T(k)\mathbf{x}(k) + \frac{1}{2}f''[e(k)]\mathbf{x}(k)[\mathbf{w}^T(k)\mathbf{x}(k)]^2\} \tag{4.28}$$

holds in every point except when $e(k) = \pm\delta$, where

$$\delta = \sqrt[4]{\frac{1}{A}\cosh^{-1}\left(\frac{\mu_{max}}{\mu_{min}}\right)}$$

Therefore, we carry on the analysis under the assumption that the noise values are very rare to become equal to δ.

4.4.1 Convergence speed

According to [15], for a small step size μ, the time-constant of the proposed algorithm associated with $\lambda_i(\mathbf{R_x})$ (the i^{th} eigenvalue of the auto-correlation matrix $\mathbf{R_x}$) is given by

$$\tau_i = \frac{1}{\mu \mathrm{E}[f'[e(k)]]\lambda_i(\mathbf{R_x})} \tag{4.29}$$

Now, assuming $e(k) \neq \pm\delta$, then

$$f(e) = \begin{cases} \frac{\mu_{min}}{\mu} e^3 \cosh(Ae^4); & |e| < \delta \\ \frac{\mu_{max}}{\mu} e^3; & |e| > \delta \end{cases} \tag{4.30}$$

Accordingly,

$$f'(e) = \begin{cases} \frac{\mu_{min}}{\mu} \left(3e^2 \cosh(Ae^4) + 4Ae^6 \sinh(Ae^4)\right); & |e| < \delta \\ \frac{\mu_{max}}{\mu} (3e^2); & |e| > \delta \end{cases} \tag{4.31}$$

Eventually, based on (4.29) and (4.31), the proposed algorithm have the following two cases:

1. if $A < \frac{1}{e^4} \cosh^{-1}\left(\frac{\mu_{max}}{\mu_{min}}\right)$ then

$$\tau_i = \frac{1}{\mu_{min}\mathrm{E}[3e^2\cosh(Ae^4) + 4Ae^6\sinh(Ae^4)]\lambda_i(\mathbf{R_x})} \tag{4.32}$$

2. if $A > \frac{1}{e^4}\cosh^{-1}\left(\frac{\mu_{max}}{\mu_{min}}\right)$ then

$$\tau_i = \frac{1}{\mu_{max}\mathrm{E}[3e^2]\lambda_i(\mathbf{R_x})} \tag{4.33}$$

which match LMF considering μ_{max} given as in equation 4.9

Since τ in the first case is smaller than LMS/LMF time-constant; the convergence of the proposed algorithm will be faster than the convergence of LMS/LMF and some of their variants. If the tuning parameter A is not properly chosen then $A < \frac{1}{e^4}\cosh^{-1}\left(\frac{\mu_{max}}{\mu_{min}}\right)$ may not happen and the proposed algorithm will work like standard LMS/LMF with firmed μ_{max} because the recursive equation becomes:

$$w(k+1) = w(k) + \mu_{max}\frac{e^3(k)}{1+e^2(k)}\mathbf{x}(k) \tag{4.34}$$

So; if the error is small, the algorithm will work like LMF while if the error is large then the algorithm will work like LMS.

Figure 4.3 demonstrates the consequence of varying A over the convergence speed, the experiment setup has a white Gaussian input signal, and additive

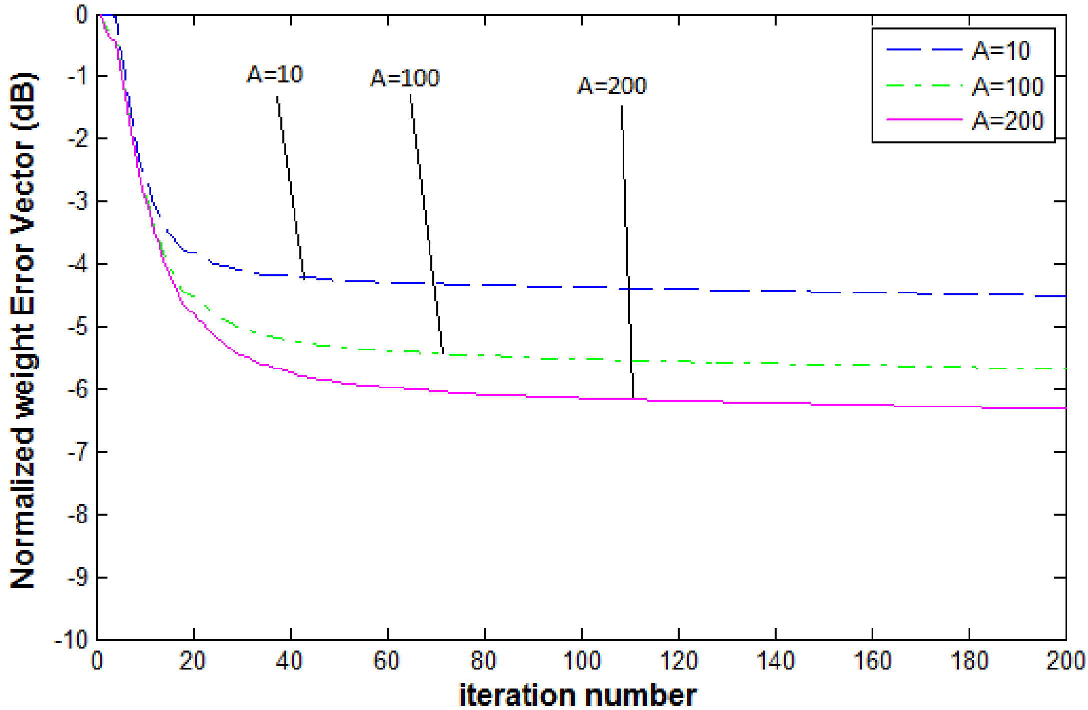

Figure 4.3: Adaptive curves for different tuning parameter values, SNR=10dB

white Gaussian noise with SNR = 10 dB. The convergence speed is noticed to be improved with the increase in the tuning parameter, and this is achieved while maintained consistent steady state error. In particular, when A ranges between 10 and 50. Furthermore, the impact of A has a compromise consequence on both the speed of convergence and the steady state error, especially for $A > 300$, whereby the EMSE increase significantly as A moves beyond 300. In general, for a certain range around A, we obtain a quite similar result.

4.4.2 Mean Behavior

As common in all gradient decent algorithms, the value of the step-size is critical. To guarantee the stability; the step size value should satisfy some certain bounds.

Rewrite equation (4.8):

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)e^3(k)\mathbf{x}(k) \tag{4.35}$$

where $\mu(k)$ is given as

$$\mu(k) = \mu_{min} \min \left[\cosh(Ae^4(k)), \frac{\mu_{max}}{\mu_{min}}\right] \tag{4.36}$$

It is sufficient to mention that the mean value of $\mu(k)$, i.e., $\mathrm{E}[\mu(k)]$, must satisfy the condition [8][19]

$$0 < \mathrm{E}[\mu(k)] < \frac{2}{3\sigma_v^2 \lambda_{max}(\mathbf{R_x})} \tag{4.37}$$

Based on (4.36), we have the following two cases:

1. if $\cosh(Ae^4(k)) < \left(\frac{\mu_{max}}{\mu_{min}}\right)$ then

$$\mu(k) = \mu_{min} \cdot cosh(Ae^4(k)) \tag{4.38}$$

Taking expectations of (4.38), and using the Taylor series expansion, we rewrite $\mathrm{E}[\mu(k)]$ as

$$\mathrm{E}[\mu(k)] \geq \mu_{min} \left\{1 + \frac{9(70)}{2} A^2 F^2(k) \sigma_v^4 + \frac{15(28)}{2} A^2 F(k) \sigma_v^6 + \frac{105}{2} A^2 \sigma_v^8 \right\} \tag{4.39}$$

Where $F = \mathrm{E}[e_a^2]$. At steady state, one can drop the high order powers of S. This implies to a new bound on $\mu_{min} > 0$ as follows:

$$\mu_{min} < \frac{2}{3\sigma_v^2\ \lambda_{max}(\mathbf{R_x})\{1 + 315A^2\mathrm{S}\sigma_v^6 + 52.5A^2\sigma_v^8\}} \tag{4.40}$$

2. if $\cosh(Ae^4(k)) > \left(\frac{\mu_{max}}{\mu_{min}}\right)$ then

$$\mu(k) = \mu_{max} \tag{4.41}$$

However; we are using generic μ which also normalized the error, so we can rewrite $\mu(k)$ as follows

$$\mu(k) = \frac{\mu_{max}}{1 + e^2(k)} \tag{4.42}$$

Hence; taking the expectations

$$\mathrm{E}[\mu(k)] = \mathrm{E}[\frac{\mu_{max}}{1 + e^2(k)}] \tag{4.43}$$

and by assuming Independence between the noise and the input signal we can approximate as follows:

$$\mathrm{E}[\mu(k)] = \frac{\mu_{max}}{1+\sigma_v^2} \tag{4.44}$$

and the new bound will be as follows:

$$0 < \mu_{min} < \mu_{max} < \frac{2(1+\sigma_v^2)}{3\sigma_v^2 \lambda_{max}(\mathbf{R_x})} \tag{4.45}$$

However we need to choose the μ_{min} first as a lower bound than μ_{max}.

4.5 Simulation Results

We run the experiments based on system identification scenarios. The order of the unknown system will be matching the order of the adaptive filter and both of them representing FIR system. Filter coefficients start up values will be zeros. There will be a mix between the output of the unknown system and white Gaussian zero mean noise sequence $v(k)$. The variance σ_v^2 of the noise sequence is selected in accordance with the desired SNR. Each experiment will run 200 times and averaged to have average results. The quantitative performance measure is the normalized weight error squared vector in dB (also known as misadjustmnet [8]), which is mathematically calculated as follows:

$$\mathcal{M} = 10\log_{10}\left(\frac{\mathrm{E}[||\mathbf{w}^o - \mathbf{w}(k)||^2]}{||\mathbf{w}^o||^2}\right)$$

where $\mathbf{w}^o = [w_0^o, w_1^o, \cdots w_{M-1}^o]^T$ the true values of the unknown system/channel tapes weight.

4.5.1 Example 1

In this example, we would like to compare our new proposed algorithm with the logarithmic cost function algorithm [18] and with LMS-LMF type II [10] as both of them considered mixed norm algorithms. The system order = 5, The input signal is uniform bipolar sequence from {1,-1} zero-mean random , the desired signal is mixed with non-Gaussian noise, and the signal to noise ratio SNR = 10dB. $\mu = 0.001$. The proposed algorithm will have scaling parameter $A = 100$ and μ tuned for all the three algorithms to have the same steady state $\mathcal{M}$ in order to have fare comparison. For LMS-LMF type II, we have μ_1 which chosen to maximize the convergence speed and μ_2 to reach the same steady state $\mathcal{M}$ as the other algorithms.

Figure 4.4 shows the adaptive curve of the proposed algorithm and the other counterpart algorithms. As can be seen from the figure, the proposed algorithm converges faster and demonstrate faster tracking ability and return back to the

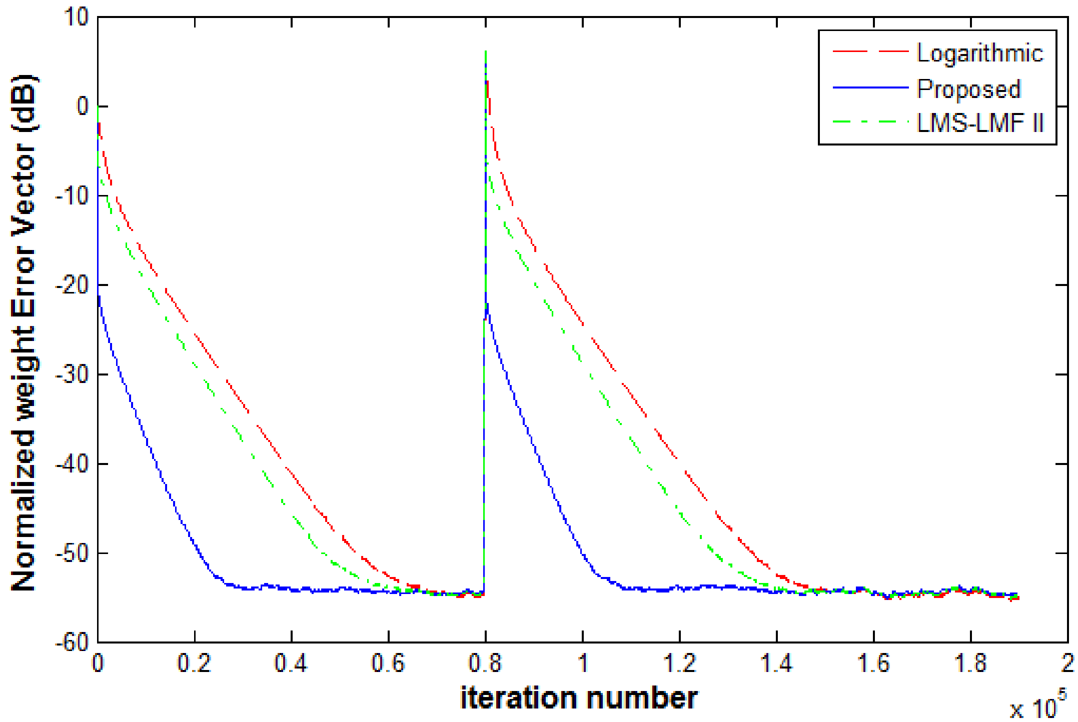

Figure 4.4: Adaptive curves of the proposed algorithm and other counterpart algorithms for sub-Gaussian noise; bipolar input signal and SNR=10dB

same steady state $\mathcal{M}$ level.

4.5.2 Example 2

In this example, we are going to demonstrate the robustness of the proposed algorithm with three different type of noise distributions (Gaussian, Uniform, and Laplacian). All of them have the same noise power.

Figure 4.5 depicts the adaptive curve of the proposed algorithm under different kind of noise. The best performance is achieved under uniform distribution noise environment. This is can be explained due to higher order moments that the proposed algorithm depends on especially in the steady state region.

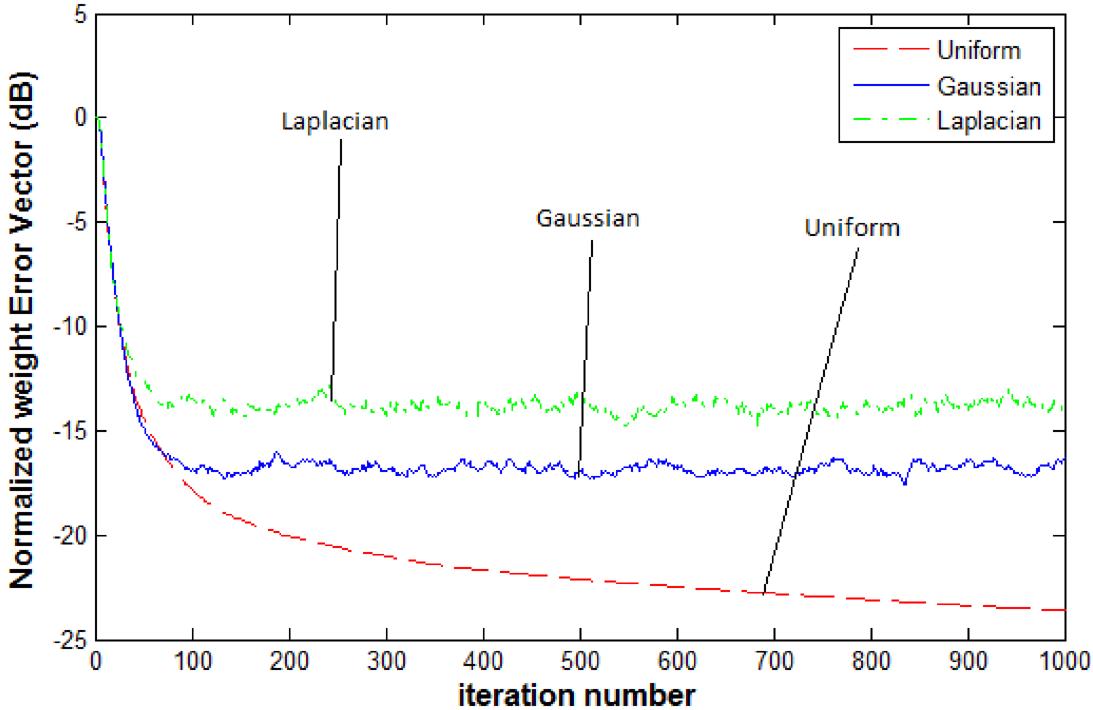

Figure 4.5: Adaptive curves of the proposed algorithm under different noise distributions

4.6 Conclusion

This chapter introduces a new class of stochastic gradient variable step size algorithm. This is achieved by introducing a new hyperbolic sine cost function. The adaptation error fourth is the argument that drives the cost function with generic μ that also normalized the error for better performance. The new algorithm maintain the simplicity and the robustness of the standard LMF, as well as the stability from LMS. It has only one tuning parameter. However, the algorithm enjoys an improved performance over others. In particular, the proposed algorithm outperforms the logarithmic-based algorithms and shows attractive results in stationary environment as well as during sudden changes scenarios.

CHAPTER 5

CONCLUSION AND FUTURE STUDIES

5.1 Conclusion

In this work, a new cost function has been introduced which is hyperbolic sine of the error. Steepest decent method is the context where the new cost function implemented. By squaring the error, we produce a new LMS-type algorithm that outperform best in class LMS-family algorithms in different environments and with different levels of signal to noise ratio and filter tap length. By raising the error to the fourth degree, we introduced a new LMF-family algorithm that outperform the best in class LMF-family algorithms. By normalizing the error, we introduced a new mixed-norm family algorithm that also outperform best in class mixed-norm family algorithms. The new cost function has only one tuning parameter, so we maintain the simplicity in the implementation. It required less computational effort comparing with similar algorithms from the same family and yet deliver better results.

Generic upper bound for the step size is a new idea that introduced in this work. It delivers flexible upper bound that can adapt with the different input signal power and channel/filter length. It required few computational effort per iteration, but it worth that effort in term of improving the performance and increase the capability of the algorithm. It can be used with any adaptive algorithm. Our experimental simulations showed good improvement in the performance, so we can look to the generic upper bound as an enhancement technique that increase the performance.

5.2 Future Studies

The new proposed cost function as well as the concept of generic step size upper bound can be leverage to produce new adaptive algorithm versions belong to sign error least mean square algorithm family, normalized least mean square error algorithm family, and normalized least mean fourth order algorithms family, to name a few. The concept of generic step size upper bound can be used alone with different adaptive algorithms, so we can look to it as an enhancement element that can be added to many existing algorithms in order to enhance their performance.

The new cost function could be modified to fit in many optimization problems, and we can look to it as an engine that have flexibility to be modified to fit in many applications. We can build on this thesis work to find other implementation work in adaptive control field, in electrical power filed, and many others.

Appendix (A)

Proof of no first derivative exist

We observe that if $h(x)$ is continues in x_0, then

$$\lim_{x \to x_0} \frac{h(x)sign(g(x)) - h(x_0)sign(g(x_0))}{x - x_0} \tag{1}$$

exists if and only if $g(x) - g(x_0)$ has the same sign in the neighborhood of x_0.

let $f : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = x\, min\{cosh(Ax^2), B\} \tag{2}$$

where $A > 0\, and\, B > 1$ ṡince

$$min\{a, b\} = \frac{a + b - |a - b|}{2} \tag{3}$$

and $x = |x|sign(x)$ for all real x, we have

$$f(x) = \frac{x\{\cosh(Ax^2) + B - [\cosh(Ax^2) - B]sign[\cosh(Ax^2) - B]\}}{2} \qquad (4)$$

$$= \frac{x\{\cosh(Ax^2) + B\}}{2} - \frac{x[\cosh(Ax^2) - B]sign[\cosh(Ax^2) - B]\}}{2} \qquad (5)$$

where x passes through $\sqrt{\frac{1}{A}\cosh^{-1}(B)} = x_0$, the term $[\cosh(Ax^2) - B]$ modifies its sign, thus $f(x)$ doesn't have a derivative in x_0

REFERENCES

[1] R. H. Kwong and E. W. Johnston, “A variable step size lms algorithm,” *IEEE Transactions on signal processing*, vol. 40, no. 7, pp. 1633–1642, 1992.

[2] R. Harris, D. Chabries, and F. Bishop, “A variable step size (vss) algorithm,” *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 499–510, 1986.

[3] V. J. Mathews and Z. Xie, “A stochastic gradient adaptive filter with gradient adaptive step size,” *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2075–2087, 1993.

[4] T. Shan and T. Kailath, “Adaptive algorithms with an automatic gain control feature,” *IEEE Transactions on circuits and systems*, vol. 35, no. 1, pp. 122–127, 1988.

[5] T. Aboulnasr and K. Mayyas, “A robust variable step-size lms-type algorithm: analysis and simulations,” *IEEE Transactions on signal processing*, vol. 45, no. 3, pp. 631–639, 1997.

[6] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size nlms and affine projection algorithms," *IEEE signal processing letters*, vol. 11, no. 2, pp. 132–135, 2004.

[7] C. Rusu and C. F. Cowan, "The exponentiated convex variable step-size (ecvss) algorithm," *Signal Processing*, vol. 90, no. 9, pp. 2784–2791, 2010.

[8] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

[9] N. Bershad, "Analysis of the normalized lms algorithm with gaussian inputs," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 793–806, 1986.

[10] A. Zerguine, M. Bettayeb, and C. F. Cowan, "A hybrid lms-lmf scheme for echo cancellation," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 3. IEEE, 1997, pp. 2313–2316.

[11] U. bin Mansoor, S. M. Asad, and A. Zerguine, "Stochastic gradient algorithm based on an improved higher order exponentiated error cost function," in *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2014, pp. 900–903.

[12] S. Haykin and B. Widrow, *Least-mean-square adaptive filters*. John Wiley & Sons, 2003, vol. 31.

[13] T. Y. Al-Naffouri and A. H. Sayed, "Adaptive filters with error nonlinearities: Mean-square analysis and optimum design," *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 1, pp. 192–205, 2001.

[14] C. Rusu and C. F. Cowan, "On some performances of the exponentiated convex variable step-size (ecvss) algorithm," in *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2013, pp. 360–365.

[15] S. C. Douglas and T.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1335–1351, 1994.

[16] W.-P. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size lms algorithms," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 805–810, 2001.

[17] Z. Shengkui, M. Zhihong, and K. Suiyang, "A fast variable step-size lms algorithm with system identification," in *2007 2nd IEEE Conference on Industrial Electronics and Applications*. IEEE, 2007, pp. 2340–2345.

[18] M. O. Sayin, N. D. Vanli, and S. S. Kozat, "A novel family of adaptive filtering algorithms based on the logarithmic cost," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4411–4424, 2014.

[19] S. M. Asad, A. Zerguine, and M. Moinuddin, "On the convergence analysis of a variable step-size lmf algorithm of the quotient form," in *2010 IEEE*

International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010, pp. 3722–3725.

Vitae

- Name: Ahmad Khalifi

- Nationality: Yemen

- Date of Birth: 1976

- Email: *g201003120@kfupm.edu.sa or ahmad2100@gmail.com*

- Permenant Address: Aden, Shake Othman Zone.

- Academic Background :

  Ahmad Khalifi received the B.Sc. degree from Jordan University of Science and Technology, Irbid, Jordan (2001), He is currently pursuing M.Sc. degree from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, all in telecommunication engineering. He has MBA degree from KFUPM too. He has worked with few international companies in communication and information technology fields. His current research interests include adaptive filtering, signal processing and channel estimation techniques.