TRANSFORM DOMAIN LMS/F ALGORITHMS, PERFORMANCE ANALYSIS AND APPLICATIONS

BY

MURWAN BASHIR

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

ELECTRICAL ENGINEERING

APRIL 2016

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **MURWAN BASHIR** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**.

Thesis Committee

Dr. Azzedine Zerguine (Adviser)

Dr. Tareq Al-Naffouri (Member)

Dr. Wajih A. Abu Al-Saud (Member)

Dr. Ali Ahmad Al-Shaikhi
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

17/5/16
Date

"Never doubt that a small group of thoughtful, committed citizens can change the world; indeed, it's the only thing that ever has."

Margaret Mead — Cultural Anthropologist

To My Mother, To My Father

ACKNOWLEDGMENTS

I reserve most thanks and appreciation to GOD the Almighty for His countless blessings. I pray that this work and the growth I had at Stanford will be used in His cause.

I would like to express my gratitude and appreciation to my principal supervisor Dr.Azzedine Zerguine, for his continuous push toward independence. I have truly learned to be self-motivated toward my research objectives. I would like to really thank Dr.Tareq Al-Naffouri, for his continuous support through my research experience here at KFUPM, and for his invites to King Abdullah University for Science and Technology (KAUST) . My Experience in KFUPM has been completely reformed after meeting him at the compressive sensing course. Commitment, diligence, perseverance, are few thing I have learned from him, among a longer list. I truly appreciate his thoughtfulness, developed intuition and kindness.

I deeply thank Mr.Mohamed Tamim Alkhodary for giving me the opportunity to join the Center of Energy and Geo-Processing, an opportunity that helped in shaping my experience in KFUPM. I would like to thank as well Mr. Krishnan, the secretary of Electrical Engineering department for his thoughtful advices. I want to thank: Mohamed Eltigany, Mohamed Hasssan, Ahmed Abdelazim, Ahmed Abu-Hassan, Ahmed Hassan, Ammar Jammie, Mohanad Hassan, Khalid Baradieah, Mohamed Abdeen, Mohamed Osman Hussien, Mohamed Abdullah Bokhary, Mohamed Tamim, and many others the space does not allow to mention, for coloring my KFUPM journey, Thank you and I wish all your dreams come true.

Finally, I want to thank my best friends, Mohamed Satti and Mamhoud Raouf for their deep understanding to my future goals and needs, and their continues support and positive thinking. At last, I want to thank my father and my mother for their sacrifice, commitment towards my success and pure love.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS **iv**

LIST OF TABLES **ix**

LIST OF FIGURES **x**

ABSTRACT (ENGLISH) **xii**

ABSTRACT (ARABIC) **xiv**

CHAPTER 1 INTRODUCTION **1**

1.1 Adaptive Filtering Structures: 1

1.1.1 LMS Algorithm 3

1.1.2 LMF Algorithm: 3

1.2 De-Correlation Approaches 4

1.2.1 Input Instantaneous Normalization 4

1.2.2 Transformation and power Normalization: 4

1.2.3 Pre-Whitening filtering approach 8

1.2.4 Performance of the transform domain LMS algorithms . . 8

1.2.5 Stochastic Modeling of TDLMS 9

1.2.6 Variants of Transform domain LMS algorithms 9

CHAPTER 2 TRANSFORM DOMAIN LMF ALGORITHM FOR SPARSE SYSTEM IDENTIFICATION **11**

2.1 Introduction 11

2.2 Regularized Algorithms 11
2.2.1 Sparse LMS 11
2.2.2 Sparse Affine Projection Algorithm (APA) 13
2.2.3 Sparse NLMS Algorithms 13
2.3 Sparse LMF and NLMF Algorithms 13
2.4 Sparse RLS Algorithm 14
2.5 Transform Domain LMS Algorithm and Its Sparse Version: Review 14
2.5.1 Geometrical Insight 14
2.5.2 Transient Analysis of Transform Domain Filters 15
2.5.3 Transient Analysis of Sparse LMS Algorithms 22
2.5.4 New LMF Algorithm 24
2.5.5 Performance Analysis of the TD-ZA-LMF 25
2.5.6 Simulation Results 29
2.5.7 Gaussian Noise Scenario 30
2.5.8 Non-Gaussian Noise Scenario 30
2.6 Conclusion 31

CHAPTER 3 CONVEX COMBINATION FILTERS FOR VARIABLE SPARSE SYSTEM IDENTIFICATION 36

3.1 Why Convex Combination Filters 36
3.2 Convex Combination Filter — Review 37
3.2.1 Steady State Universality 37
3.2.2 Investigation of Transient Universality 38
3.3 Convex Combination of LMF and ZA-LMF for Sparse System Identification 40
3.3.1 Problem Formulation 41
3.3.2 Convex Combination Filter — Review 42
3.3.3 Introduction 42
3.3.4 The universality of the combination filter 43
3.3.5 Computer Simulation 51

3.4 Convex Combination of NLMF and ZA-NLMF algorithms 53
3.5 Convex Combination of TD-LMS and TD-AZ-LMS 53
3.6 abstract 53
3.7 Introduction 53
3.8 Convex Combination Filter 55
3.9 Excess Mean Square Error for Component Filters 56
3.10 Universality of the Convex Filter 58
3.10.1 Non-Sparse System — $J_{ex,2} \leq J_{ex,12} \leq J_{ex,1}$: 59
3.10.2 Semi-Sparse System: 60
3.10.3 Sparse System — $J_{ex,1} \leq J_{ex,12} \leq J_{ex,2}$: 60
3.11 Computer Simulation 60
3.12 Conclusion 64

CHAPTER 4 NON-NEGATIVE SPARSE SYSTEM IDENTIFICATION **65**
4.1 Sparse Non-negative NLMF algorithms 65
4.1.1 Introduction 65
4.1.2 Zero Attractor Non-Negative LMF Algorithm 65
4.1.3 Mean Behavior of the l_1 NNLMF algorithm 67
4.1.4 Simulation 68
4.1.5 Conclusion 68
4.2 Convex Combination of NNLMS and l_1 NNLMS for variable non-negative sparse system identification 68
4.2.1 Introduction 68
4.2.2 Problem Formulation 71
4.3 Computer Simulation 72

CHAPTER 5 CONCLUSION AND FUTURE WORK **73**

REFERENCES **74**

VITAE **83**

LIST OF TABLES

2.1 Zero Attractor values for the different algorithms. 30

LIST OF FIGURES

1.1 Schematic Diagram of the system identification problem. 1

2.1 Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = 0.9$ 15
2.2 Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = 0.9$ and DST transformation 16
2.3 Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = 0.9$ and DCT transformation 16
2.4 Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = -0.9$ and DCT transformation 17
2.5 Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = -0.9$ and DST transformation 17
2.6 The distribution of some elements of the transformed regressors . 19
2.7 MSE Theoretical and Simulation Curves 21
2.8 MSD curves for white Gaussian input with sparsity rate = 2. . . 31
2.9 MSD curves for correlated input with sparsity rate = 2 and uniform measurement noise. 32
2.10 MSD curves for correlated input with sparsity rate = 2 and uniform measurement noise. 33
2.11 MSD curves for correlated input with sparsity rate = 4, with uniform measurement noise. 34
2.12 MSD curves for correlated input with sparsity rate = 4,with uniform measurement noise. 35

3.1 Diagram shows the Convex Combination filter proposed in [1]. . . 37

3.2 Excess Kurtosis versus SNR. Both the input and noise measurement follow Gaussian Distributions. The number of filter taps $M = 10$, with step size $\mu = 0.001$. The experiments are run 10000 times and the estimated Kurtosis is averaged over 50 experiments. We assume ergodicity of the error signal. 41

3.3 MSD curves for LMF, ZA-LMF algorithms and their convex combination. SNR level is 10 dB with uniform measurement noise. . . 52

3.4 Combination factor evolution for the first experiment 54

3.5 Diagram shows the Convex Combination filter proposed in [1]. . . 55

3.6 MSD curves for LMF, ZA-LMF algorithms and their convex combination. SNR level is 10 dB with uniform measurement noise. . . 61

3.7 Combination factor evolution for the second experiment — Correlated Input 63

4.1 MSD curve for Sparse NNLMF, NNLMF algorithms,with White Gaussian Input and White Gaussian noise, and sparsity rate $\frac{1}{32}$ with random support. Simulation parameters are set to $\mu = 0.001$, $\lambda = 0.001$ and $M = 32$. 69

4.2 MSD curve for Sparse NNLMF, NNLMF algorithms,with White Gaussian Input and uniform noise, and sparsity rate $\frac{1}{32}$ with random support. Simulation parameters are set to $\mu = 0.001$, $\lambda = 0.001$ and $M = 32$. 70

4.3 MSD curve for Sparse NNLMS, NNLMS and their convex filter,with White Gaussian Input 72

THESIS ABSTRACT

NAME:	Murwan Bashir
TITLE OF STUDY:	Transform Domain LMS/F Algorithms, Performance Analysis and Applications
MAJOR FIELD:	Electrical Engineering
DATE OF DEGREE:	May 16, 2016

Adaptive Filtering Algorithms are known to suffer in terms of convergence under highly correlated inputs. In this work we investigate the Transform Domain LMF algorithm. The Fourth Moments family of adaptive algorithms is known to have better steady state and convergence properties compared to the LMS family, under Non-Gaussian noise environments. The second scope of the thesis is to study the problem of sparse system identification, that is, when the number of active elements in the linear system under consideration is small. We proposed the Zero-Attractor (ZA-TD-LMF) algorithm to solve the problem under the highly correlated environments and non-Gaussian noise. Furthermore to deal the problem of variable sparsity, we employed the convex combination filter to design an algorithms that can solve the problem of moving sparse elements. The first con-

vex filter, called ZA-LMF and LMF convex, is designed to alleviate the choice of which algorithm we should use (LMF or ZA-LMF) if we do not have previous information about the level of sparsity of the system under consideration. Similarly, a transform Domain LMS and ZA-LMS convex is proposed and studied to endow the designer with more freedom in system designing. The Computer Simulations confirm that the convex is universal and always offer the best performance available and chooses the best performing algorithm.

Acknowledgment

I would like to express my gratitude and appreciation to my principal supervisor Dr.Azzedine Zerguine, for his continuous push toward independence. I have truly learned to be self-motivated toward my research objectives. I would like to really thank Dr.Tareq Al-Naffouri, for his continuous support through my research experience here at KFUPM, and for his invites to King Abdullah University for Science and Technology (KAUST) . My Experience in KFUPM has been completely reformed after meeting him at the compressive sensing course. Commitment, diligence, perseverance, are few thing I have learned from him, among a longer list. I truly appreciate his thoughtfulness, developed intuition and kindness.

I deeply thank Mr.Mohamed Tamim Alkhodary for giving me the opportunity to join the Center of Energy and Geo-Processing, an opportunity that helped in shaping my experience in KFUPM. I would like to thank as well Mr. Krishnan, the secretary of Electrical Engineering department for his thoughtful advices. I want to thank: Mohamed Eltigany, Mohamed Hasssan, Ahmed Abdelazim, Ahmed Abu-Hassan, Ahmed Hassan, Ammar Jammie, Mohanad Hassan, Khalid Baradieah, Mohamed Abdeen, Mohamed Osman Hussien, Mohamed Abdullah Bokhary, Mohamed Tamim, and many others the space does not allow to mention, for coloring my KFUPM journey, Thank you and I wish all your dreams come true.

Finally, I want to thank my best friends, Mohamed Satti and Mamhoud Raouf for their deep understanding to my future goals and needs, and their continues support and positive thinking. At last, I want to thank my father and my mother for their sacrifice, commitment towards my success and pure love.

CHAPTER 1

INTRODUCTION

1.1 Adaptive Filtering Structures:

Adaptive filters are one of the key elements in most of the communications and real time signal processing systems because of their high efficiency and reliability. It is widely used in channel equalization, echo cancellation, noise cancellation and many other applications. Regardless of the difference in nature between all the systems that uses adaptive filters, there is one common, an input signal and a desired response are used to calculate the error which is used in controlling an adjustable filter coefficients. Generally, there are four types of structures of adaptive filters that covers wide range of applications: system identification, Inverse system modeling, signal predication and noise cancellation [2].

Since the main objective of this thesis is to study sparse system identification structure, we employ system identification setting. Figure 1.1 below show a schematic diagram of the abstract system identification problem.

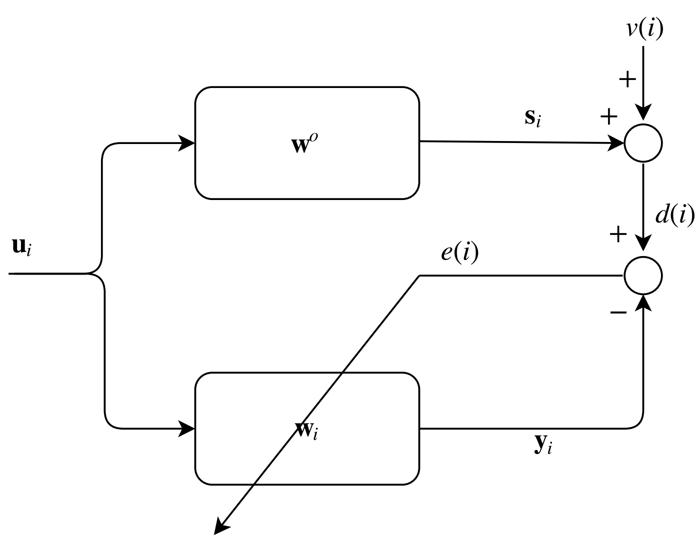

Figure 1.1: Schematic Diagram of the system identification problem.

The input to both systems is $\mathbf{u}_i$. The first system is the unknown $\mathbf{w}^o$ while the other is $\mathbf{w}_i$. Note that we do not have control over $\mathbf{w}^o$, but through $\mathbf{w}_i$ we will be able to recognize the unknown system. $\mathbf{d}_i$ is the desired output which is contaminated with the noise process $v(i)$.

*If $v(i)$ is Gaussian process, we say that the adaptive filter is performing in **Gaussian noise environment**.If else, we say it performs in **Non-Gaussian noise environment**.*

The error signal $e(i) = d(i) - y(i)$ is used to adjust the parameters of $\mathbf{w}_i$. $y(i)$ is the estimated desired signal.

The estimation problem in Fig 1 can be easily casted as an optimization problem. The cost functions used for this purpose are quite abundant, we limit ourselves here to the basic (and famous) costs. For example, the weights vector $\mathbf{w}_i$ can be controlled using Square Error function defined as following:

$$J(i) = \frac{1}{2}E\{e^2(i)\} \tag{1.1}$$

where:

$$e(i) = d(i) - y(i) \tag{1.2}$$

$$d(i) = \mathbf{w}^o\mathbf{u}_i + v(i) \tag{1.3}$$

Note that our objective is to minimize (1.1).

The weights can be solved as well using the fourth power minimization of the error, that is:

$$J(i) = \frac{1}{4}E\{e^4(i)\} \tag{1.4}$$

Several approaches can be followed to minimize both (1.1) and (1.4) recursively. The steepest descent however, has proved its simplicity and robustness through out history, and these functions were not an exception to it. The steepest descent formula is given by:

$$w_i = w_{i-1} + \mu\frac{\partial J(i)}{\partial w_{i-1}} \tag{1.5}$$

By applying (1.5) to (1.1), the resultant steepest descent is of the following form [3]:

$$w_i = w_{i-1} + \mu[\mathbf{R}_{du} - \mathbf{R}_u w_{i-1}] \tag{1.6}$$

where $\mathbf{R}_{du} = E\{d(i)\mathbf{u}_i\}$ is the covariance vector between the input $\mathbf{u}_i$ and the desired output $d(i)$. $\mathbf{R}_u$ is autocorrelation matrix of the input row vector $\mathbf{u}_i$ — assuming the input and the desired output are jointly stationary, and the input is stationary. As $i \to \infty$, $w_i = \mathbf{w}^o$. Note that w_i is deterministic unknown vector.

1.1.1 LMS Algorithm

The steepest descent recursion depends on the second order statistical information of the input and the desired response. Since these moments are practically difficult to acquire, we resort to approximating these moments. The simplest approximation is the instantaneous one, that, $\mathbf{R}_u$ is approximated by $\mathbf{u}_i^*\mathbf{u}$ for example. Following this instantaneous approximation, (1.6) is rendered to [4]:

$$\mathbf{w_i} = \mathbf{w_{i-1}} + \mu \mathbf{u_i} e^*(i) \tag{1.7}$$

Which is called the Least Mean Square (LMS) algorithm. The LMS algorithm is one of the famous adaptive filters algorithms because of its simplicity which suits many practical real time applications.

It may seem at first glance as well this approximation is only beneficial from computational point of view. However, (1.7) adaption depends on streaming data, and if the statistical nature of the data changes with time, the LMS will be able to adapt itself and still give the desired solution. This is not possible with the case of (1.6) since we have to know when the data nature will change — a very difficult pursuit. Note that $\mathbf{w}_i$ is stochastic vector unlike the w_i, which is unknown deterministic vector.

The LMS can be sought as steepest decent algorithm of the following error cost function:

$$J(i) = \frac{1}{2}\, e^2(i) \tag{1.8}$$

1.1.2 LMF Algorithm:

As in (1.8), the cost function in (1.4) can be rendered to (by applying instantaneous statistical approximation — remove the expectation):

$$J(i) = \frac{1}{4}\, e^4(i) \tag{1.9}$$

and the resultant algorithm of minimizing (1.9) is called Least Mean Fourth (LMF) algorithm:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i e^*(i)|e(i)|^2 \tag{1.10}$$

LMF had been proposed as a solution to the poor performance of the LMS under non-Gaussian noise according [5]. However, LMS performs better in the Gaussian noise environment. The common factor though (which is the most related to the future discussion and results) is that both of these steepest decent algorithms suffers from the correlation of the input of the signals [6][3]. The correlated input appears for example, in the Echo Cancellation systems. In these systems, the input is the human speech — which is highly correlated. And since these systems are real time systems, the slowness of the adaptive filters used can significantly reduce the quality of the final echo-free speech. Hence, it becomes mandatory to find solutions to solve this problem of slow convergence to the optimum solution $\mathbf{w}^o$. In the next section, we explore the main approaches used to improve the convergence rate of adaptive algorithms.

1.2 De-Correlation Approaches

The three approaches mainly used to tackle the slow convergence (resulted from highly correlated input) are; instantaneous power normalization, unitary transformations accompanied with power normalization and pre-whitening filtering. At the end of this section we justify the usage of the Transform domain algorithms (which belongs to the approach of unitary transformations).

1.2.1 Input Instantaneous Normalization

Normalized LMS/F Algorithms

The recurrence of Normalized Least Mean Square (NLMS) is given [7][2]:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|_2^2} e(i) \tag{1.11}$$

The power normalization introduced in (1.11) proved to shrink the eigenvalue spread of the autocorrelation matrix of the input $\mathbf{u}_i$, and hence whiten the input.

In [8], the NLMF algorithm is motivated by the NLMS algorithm, and has the following recursion:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|_2^2} e^3(i) \tag{1.12}$$

The algorithm performed better than NLMS in Non-Gaussian Environments. Normalized LMF inherits the LMF features (low Meas Square Error (MSE) in Non-Gaussian noise environment)).

NLMS and NLMF algorithm offer good performance on the correlated environments, however, as the correlation level increase the instantaneous power normalization does not mitigate the correlation and may introduce instabilities especially in the NLMF algorithm. This problem acted as motive for looking for more thorough ways of de-correlation and performing normalization.

1.2.2 Transformation and power Normalization:

The input regressor in this approach is transformed - using suitable transform — to another domain where it has lower correlation. There are many transforms that can do the job, but the optimal (In Mean Square Sense) is Karhunen-Loeve Transform or for short KLT.

Karhunen-Loeve Transform:

The Karhunen-Loeve theorem states that a wide stationary process can be expanded into summation of random variables with orthnormal functions, i.e. it is analogous to the Fourier transform of functions in the deterministic case. Mathematically: assume $x(t)$ a wide stationary random process then:

$$\tilde{x}(t) = \sum_{i=1}^{\infty} \mathbf{C}_n \phi_n(t) \tag{1.13}$$

where $\mathbf{C}_n$ is a random variable given by the following projection:

$$\mathbf{C}_n = \int_0^{\tau} x(t)\phi_n^*(t)dt \tag{1.14}$$

The Karhunen-Loeve expansion is optimal since the following is true :

$$E[|\tilde{x}(t) - x(t)|^2] = 0, \tag{1.15}$$

which implies that the set of the $\phi_n(t)$ are set of orthonormal functions. from the KL theorem we get the following crucial result:

$$E[\mathbf{C}_n\mathbf{C}_m^*] = \begin{cases} \sigma_n^2 & n = m \\ 0 & n \neq m \end{cases}$$

extending the input signal to be a vector, the KL expansion will be a summation of vectors over the same orthonormal set which can be transformed into matrix format and we will have the KL transformation. The KLT based on this simple analysis is optimal in a mean square sense and has perfect de-correlation properties.[7][9]. KLT transform is found to be optimal in another sense, the variance distribution after the transformation [10]. The KLT transform is quite complex, for example the KLT entries of N by N KLT transform for first order Markov process is given by:

$$w_{ij} = [\frac{2}{N + \mu_{j]}}]^{\frac{1}{2}} sin[r_j[(i+1) - \frac{N+1}{2}] + (j+1)\frac{\pi}{2}] \tag{1.16}$$

and $\mu_j = (1-\rho^2)[1 - cos(r_j) + \rho^2]$ and r_j is the real positive root of the transcendental equation:

$$tan(Nr) = \frac{(1-\rho^2)sin(r)}{cos(r) - 2\rho + \rho^2 cos(r)} \tag{1.17}$$

from the above equations we see the strong dependence of the KLT into the correlation factor . This one of the main obstacles in calculating the KLT. for some applications like image processing we can estimate ρ from heuristics. But this is not the case for most of the applications. Not only that, assuming we know

ρ , the calculation can computationally demanding if N is very large, especially in real time applications like echo-cancellation for example. For this reasons, the KLT has been set to be the bench mark for evaluating the sub-optimal transforms. Transforms that is data independent and computationally efficient. The family of Discrete Cosine Transform (DCT), Discrete Since Transform (DST), Discrete Fourier transform (DFT) and Discrete Hilbert Transform are well known for their ability to perform de-correlation.

Discrete Cosine Transform:

There are four types of Discrete Cosine Transforms commonly used in the literature, which are:

DCT-I	$[C^{I}_{N+1}]_{nk} = \sqrt{\frac{2}{N}}[\epsilon_n \epsilon_k cos\frac{\pi nk}{N}]$
DCT-II	$[C^{II}_{N}]_{nk} = \sqrt{\frac{2}{N}}[\epsilon_k cos\frac{\pi(2n+1)k}{2N}]$
DCT-III	$[C^{III}_{N}]_{nk} = \sqrt{\frac{2}{N}}[\epsilon_n cos\frac{\pi(2n+1)n}{2N}]$
DCT-IV	$[C^{IV}_{N}]_{nk} = \sqrt{\frac{2}{N}}[cos\frac{\pi(2n+1)(2k+1)}{4N}]$

where $n, k = 0, ..., N-1$ and $\epsilon_p = \frac{1}{\sqrt{2},} p = 0, N$ and zero otherwise.
The DCT-II and its inverse DCT-III proposed by [11] are known for the excellent de- correlation and energy compaction and in some cases can be approximated to the KLT. in [11], the transforms are proposed and proved to perform closely to the KLT for Markov-I process. The same result is proven on [12] using analytical approach. Further study of the DCT diagonalization properties is done in [15],where, eight DCT transformation had been generated form a general from circulant matrix. To study their asymptotic behavior, the weak norm of the difference of the input covariance matrix and the output (after transformation) covariance matrix is used. The result is that the DCTs developed are asymptotically optimal for all finite order Markovian processes. hence, in another word, in this case the DCTs are asymptotically equivalent to Karhunen-Loeve transformation. In [13] comparison is conducted between the Karhunen-Loeve transformation and DCT, for widely correlated data (the correlation factor vary widely), the DCT is found perform closely to KLT in energy compaction and de-correlation.

Discrete Fourier Transform:

Discrete Fourier Transform is defined as following:

$$F_N(i,l) = \frac{1}{\sqrt{N}} e^{\frac{j2\pi l}{N}} \tag{1.18}$$

$fori, l = 0, ..., N-1$
in [14], a comparison is held to measure of the effectiveness of de-correlation for

DFT and DCT transforms in the case of Makov-1 input process. The metric used for the comparison was Hilbert-Schmidt norm for the difference between a chosen Toeplitz matrix and the transformation matrix. it is found that for the first Markovian process the DCT performed (that is, de-correlate much better) than DFT , or at least as it. Since these transformation are generally suboptimal (cannot give perfect de-correlation) residual correlation is expected.

Discrete Sine Transform

There are four types of Discrete Sine Transforms commonly used in the literature, which are:

DST-I	$[C^{I}_{N-1}]_{nk} = \sqrt{\frac{2}{N}}[sin\frac{\pi(n+1)(k+1)}{2N}]$
DST-II	$[C^{IV}_{N}]_{nk} = \sqrt{\frac{2}{N}}[\epsilon_k sin\frac{\pi(2n+1)(2k+1)}{2N}]$
DST-III	$[C^{IV}_{N}]_{nk} = \sqrt{\frac{2}{N}}[\epsilon_n sin\frac{\pi(n+1)(2k+1)}{2N}]$
DST-IV	$[C^{IV}_{N}]_{nk} = \sqrt{\frac{2}{N}}[cos\frac{\pi(2n+1)(k+1)}{2N}]$

where $n, k = 0, ..., N-1$ and $\epsilon_p = \frac{1}{\sqrt{2},} p = 0, N$ and zero otherwise.
DST-II and DST-III are used exchange ably with DCT-II and DCT-III. DST-II is found to be equivalent to KLT when the input processes is Markov-I with negative correlation factor or very small adjacent-element correlation factor [15][16]. In [17] , new transformation is constructed in a way it will be approximated to KLT in the case of the first order process — and work for $\pm\rho$. This transformation is also symmetrical regarding the energy packing.

Power Normalization to enhance the performance

The De-correlation alone will not solve the problem of the poor performance in the correlated environment [18][7][19], because using unitary transformation only removes the correlation and remains the problem of the eigenvalue spread. The unitary transformation performs similarity transformation which is known to preserve the eigen values and eigen value spread. The power normalization concept is used to shrink the eigenvalue spread of the resultant matrix of the de-correlation process. The power normalization hence, is trying to whiten the transformed input [19].

Technique for estimating the power Normalization Matrix

The power normalization technique generally used is exponential smoothing as described in [19][20]. $\mathbf{\Lambda_i}$ is defined as the power normalization matrix with diagonal elements calculated as following:

$$\sigma^2_k(i) = \beta\sigma^2_{k-1}(i) + (1-\beta)|U_k(i)|^2 \tag{1.19}$$

with β as smoothing factor. However it should be noted that this estimation is valid only on the environment where the input is not too non-stationary. The initial values of the powers are commonly based on initial estimate of the input power.

1.2.3 Pre-Whitening filtering approach

Assuming that the second order statistics of the input sequence u(i) is known, the spectral factorization of the input is given by:

$$S_u(z) = \sum_{-\inf}^{\inf} r(k) z^{-k} \tag{1.20}$$

where $r(k) = Eu(i)u^*(i-k)$, and since the spectral factorization has a para-hemeritian property it can be written as:

$$S_u(z) = \sigma_u^2 \frac{\Pi_{i=1}^{m}(z - z_l)(z^{-1} - z_l^*)}{\Pi_{i=1}^{n}(z - p_l)(z^{-1} - p_l^*)} \tag{1.21}$$

from this relation we can clearly see that if we designed a filter with function $\frac{1}{\sigma_u A(z)}$, the output will have unit variance and completely whitened. However this approach is demanding on the level of information required to perform it, since it is all based on the assumption that we know the second order statistics of the input and the order of the process.

Why Transform Domain Adaptive Filtering Approach

From the previous discussion we can conclude clearly that the Transformation and power normalization is the most practical approach to mitigate the effect of the correlated input slow convergence of adaptive filters. By applying unitary transformation that is data independent, we free ourself from the immense knowledge to implement both the KLT transformational and the pre-filtering approach. Moreover, the Unitary transformation will use efficient algorithms to be calculated which will reduce the computational complexity significantly.

To get more insight about these algorithms, we explore the choice of the unitary transformation and its stochastic modeling, and conclude with variants of algorithms.

1.2.4 Performance of the transform domain LMS algorithms

Comparing the two versions of LMS algorithms-time domain versus frequency domain, it is found that they give identical Weiner solution, regardless of the transform used. From computational complexity point of view the TDLMS algorithm increased the complexity of calculations because of the transformation and power normalization [21]. In [22], The Transform Domain LMS algorithm is studied from a different point of view, that is , how the transform used with power normalization can reduce the eigenvalue spread of the input. In [23], the analysis is extended for the second order AR processes. In order for the TDLMS algorithm to perform well, the higher the dimension of the transform the more de-correlation we can get. but this create a real time implementation issue since it will impose large delays. In [24] , an algorithm is proposed to become compatible with rectangular transformation, to overcome this tradeoff and still have high decor relation with small time delay.

1.2.5 Stochastic Modeling of TDLMS

in [25], statistical model is developed for the transform domain LMS algorithm, with regularization factor in the power normalization. The result of the proposed model is the MSE curve for the algorithm. This model is built on easing assumptions: the input vector and the estimated weight vector are statistically independent (independent assumptions), and the power normalization matrix and the numerical autocorrelation of the input are jointly stationary (i.e. the power normalization matrix does not change significantly with time). The model assumes colored Gaussian data as input. The MSE curve is found to be very close to the simulated one, with second order process. The model is developed even more for finer agreement between the simulation and analytical MSE curves in [26] without using the independence assumptions. Also conditions on the step size is obtained. The previous models assume white Gaussian input [19] (and other similar models). Hence, significantly reduces the mathematical analysis.

1.2.6 Variants of Transform domain LMS algorithms

In this As in the case of the LMS, Transform Domain LMS also has its normalized version which proved to be superior to the TDLMS as discussed in [27]. in [28], A new approach is used to study the normalized TDLMS and M-estimate Algorithm TDNLMM[29] , which is better than TDNLMS algorithm in impulsive noise environment. Price theorem is used to obtain decoupled difference equations that describe the mean and mean square behavior. In the work of [30], A new LMS algorithm based on the partial update of the filter coefficients, i.e. to update the only significant coefficients that will reduce the MSE. The choice of the subset group is based on magnitude of the corresponding gradient estimate. The transform domain LMS version of the subset update is proposed in [31], The DCTLMS proved to benefit from the partial update and reduce the computational complexity of the DCTLMS.

In [32][33], variable step TD-LMS is proposed, based on the instantaneous power normalization, error , and a global factor which is the same for each filter confident. The computational complexity increased, however the convergence is very fast compared to the ordinary transform LMS. in [34], A new normalized TDLMS algorithm is proposed to solve the sensitivity of the TDNLMS algorithm to the level of excitation signal (which differs greatly over time). The proposed algorithm uses regularization factor in order to reduce the variance of the estimator. In [35] a new algorithm that uses a variable regularization factor is developed for the case of white noise input. The algorithm showed excellent convergence speed compared to Regularized TDNLMS algorithm for colored Gaussian input.

CHAPTER 2

TRANSFORM DOMAIN LMF ALGORITHM FOR SPARSE SYSTEM IDENTIFICATION

2.1 Introduction

The sparse-aware algorithm advent has been triggered firstly by recognizing that important impulse responses (IR) happens to be sparse by nature, like the room acoustic response wireless channels with few reflector environment and MIMO systems. However because of lack of tools to employ such structure in adaptive filtering and estimation problems, the early work done based on case by case solution, for example numerous algorithms developed for enhancing echo cancelers performance by exploiting this structure [36] [37] [?] [38] [39] [40] [41]the famous proportionate NLMS, which changes the step-size-per-element each iteration and has the potentiality to consider sparsity, by setting small step sizes for the zero elements. The availability of mathematical tolls and their properties introduced by the Compressed Sensing theory, availed the algorithms to be truly sparse aware (the algorithm literally knows that its objective is sparse) as opposed to the PNLMS approach.

2.2 Regularized Algorithms

2.2.1 Sparse LMS

The regularization firstly introduced to the LMS algorithm by [42]-motivated by the LASSO-, where two version of regularizations used. The first one is the l_1 norm which appears as a sign function on the recursion. The term acts as a Zero Attractor (ZA), which simply attracts the filter coefficients to zero, including the non-zero elements. The second logarithmic (with absolute value) regularization function is proposed to distribute the attraction power unequally, force the zero elements to zero and release the non-zeros from the attraction power and it is called Re-wighted ZA (RZA). The second algorithm (RZA-LMS) performed much better because it introduces less bias as compared to (ZA-LMS).The RZA term is approximated by piecewise linear functions to make the slope of the decent steeper, which resulted in faster convergence [43].

The Steady State Analysis for ZA-LMS algorithm is given in [44], assuming the input is white Gaussian.The analysis resulted in convergence limits and a rule of on choosing the regularization factor.

The l_0 norm is basically the optimum norm to exploit sparsity (as suggested by the compressive sensing theory), but minimization of the l_0 is NP hard problem which motivated the l_1. However,in [45] the l_0 sparse aware LMS is implemented using exponential something of l_0 given in [46] . The analysis of the algorithm resulted in rules of choosing the regularization parameter and the smoothing factor of the l_0 approximation[47].

The l_p norm is introduced in [**?**], in [48] the Re-wighted l_1 is introduced,which results in an attraction term stronger than RZA. The Leaky RZA-LMS is introduced in [49].

Set theoretic approach based on weighted projection on l_1 norm is discussed in [50] [51], where each data pair come is used to create set that the solution exist in, and recursive projection leads to the solution.

Adaptive Sparse Signals Estimation

Sparse-Aware algorithms are employed in estimating the sparse signals. In [52], three algorithms introduced; l_0 LMS, Exponential Forgetting Factor LMS (EFF-LMS), and Zero Attraction Projection (ZAP-LMS). The algorithms splits the estimation into two recursive steps. The first step is LMS based (it can exploit sparsity as well) while second step enforces sparsity in the solution.

Sparse LMS slow Convergence remedies

The convergence speed of the LMS is inversely proportional to the correlation level of regressors, and its sparse aware versions inherit this disadvantage. Transform domain LMS with ZA and RZA proposed by [53] to secure faster convergence when the input is correlated. The variable step size algorithms VSS is another valid approach for rising the convergence performance, the VSS-ZA-LMSa and VSS-RZA-LMS introduced in [54][55], where the step size recursion depends on the squared estimation error.

Support Identification Algorithms

By splitting the estimation od sparse vectors into two optimization problem, the first is estimating the locations (gains at each location) and the second to estimate the amplitude, the authors in [56] presented two recursions update alternatively, the algorithm performance is very close to the oracle LMS (the know-it-all algorithm) and much better than sparse-LMS algorithms, because clearly this algorithm double check at each iteration which means faster convergence.

Block Sparsity Structure

The block sparsity is structure can appear in many situation. The sparse algorithms presented accounts for sparsity but id does not take block structure into consideration. In [57], the authors presented mixed norm, namely $l_{1,2}$. Two algorithms showed and called Group ZA-LMS (GZA-LMS) and its reweighed version called (GRZA-LMS).

Variable Sparsity Structure

The structure of the system can shift over time from sparse to completely non-sparse, in this case ZA-LMS for example performance deteriorate significantly, in [58] convex combination filter composed of LMS and ZA-LMS. The filter proved to be universal-perform at least as the best of the two.

2.2.2 Sparse Affine Projection Algorithm (APA)

The ZA-APA and RZA-APA are introduced in [59] and the stability analysis is developed in [60], and based on MSD minimization a rule on regularization factor is deduced. Soft thresholding technique based on local approximation of the l_p is introduced in [61]. The l_0 APA is introduced in [62], the performance is proved by simulation only.

2.2.3 Sparse NLMS Algorithms

The NLMS sparse [62].The VSS-NLMS based on the based on l_0 and l_1 are introduced in [63][64] respectively and the transient analysis and stability of ZA-NLMS is studied on [65]. In [66] an exact NLMS l_p based algorithm is derived based on the fact that the l_p is non-convex (we can not just take the derivative) and hence the recursion result is exact and Quasi Newton, and the resulted NLMS is reported to be stable.

2.3 Sparse LMF and NLMF Algorithms

LMF has its sparse versions introduced in [67] , where ZA-LMF, RZA-LMF ,l_p-LMF and l_0-LMF are proved to perform significantly better than the LMF and LMS at low SNR environment. The NLMF performance is know to be much better than LMF in the case of correlated input, and sparse versions inherit this property as well. In [68] ZA-NLMF introduced and the RZA-NLMF is proposed by [69], and the same algorithm with used for estimating sparse signals in [70][71]. The l_p norm NLMF introduced in [72] with application to the MIMO channel estimation. The ZA-NLMF and RZA-NLMF algorithms are generalized by the regularized weighted l_1 algorithm introduced in [73].

2.4 Sparse RLS Algorithm

The Least Square Method (LS) is extended to its l_1 norm regularized version to explore the sparsity of the MIMO channels estimation [74]. The RLS is equipped with the l_1 and l_0 norm in and the l_p-RLS in[75]. Greedy implementation that has comparable computational complexity to recursive approach is proposed in [76]. The RLS is used as well to solve the problem of block sparsity using mixed norms in [77] . The RLS is used with l_1 to estimate sparse signal introduced in [78].

This chapter deals with the LMF algorithms and their transform domain version in a sparse system identification (SSI) setting. The SSI has gained its popularity over the years because it has been shown , in many research area, that the impulse response naturally have very few insignificant components compared to its length,i.e sparse.

2.5 Transform Domain LMS Algorithm and Its Sparse Version: Review

2.5.1 Geometrical Insight

Let us first understand how the transformation and power normalization help in increasing the convergence speed. We can treat (2.1) and (2.2) as LMS with input $\mathbf{v}_s(i) = \frac{\mathbf{v}_s(i)}{\|\mathbf{v}_s(i)\|^2}$ and $\mathbf{v}_c(i) = \frac{\mathbf{v}_c(i)}{\|\mathbf{v}_c(i)\|^2}$. The Correlation Matrices of these inputs can be approximated as following:

$$\mathbf{R}_s = E\mathbf{v}_s\mathbf{v}_s^T \approx E\mathbf{\Lambda}_s^{-1}(i)\mathbf{T}_s\mathbf{R}\mathbf{T}_s^T \tag{2.1}$$

$$\mathbf{R}_c = E\mathbf{\Lambda}_c^{-1}(i)\mathbf{T}_c\mathbf{R}\mathbf{T}_c^T \tag{2.2}$$

Where $\mathbf{R} = E\mathbf{u}^T\mathbf{u}$. In order to evaluate the effect of TPN (Transformation and Power Normalization), we study the eigenvalues of $\mathbf{R}_s$ and $\mathbf{R}_c$ with respect to the non-transformed eigenvalues of $\mathbf{R}$. And because of the difficulty of eigenvalue analysis of the toeplitz matrices, we resort to limiting theorems of eigenvalues, namely Gershgorin's theorem

Gershgorin's Theorem:

Let A be a complex $n \times n$ matrix, with entries a_{ij}, For $i \in \{1, \dots, n\}$ let $R_i = \sum_{j \neq i} |a_{ij}|$ be the sum of the absolute values of the non-diagonal entries in the i-th row. Let $D(a_{ii}, R_i)$ be the closed disc centered at a_{ii} with radius R_i. Such a disc is called a Gershgorin disc. Then:

Every eigenvalue of A lies within at least one of the Gershgorin discs $\boldsymbol{D(a_{ii}, R_i)}$.

Let us assume we have the following 4×4 AR(1) Autocorrelation matrix

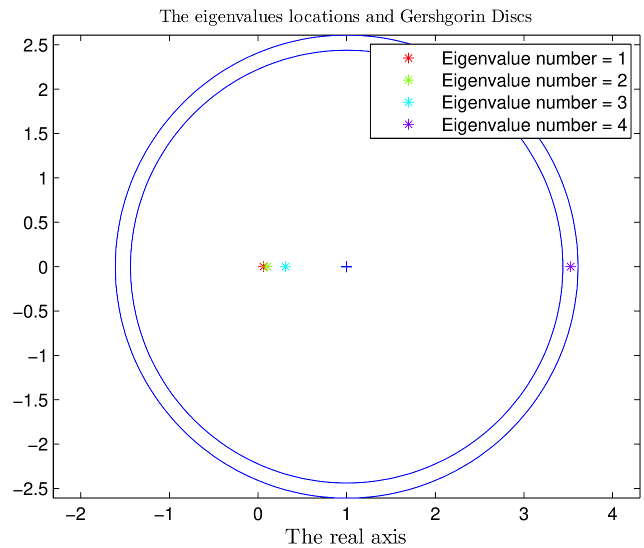

Figure 2.1: Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = 0.9$

$\mathbf{R}$, with pole ρ:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \tag{2.3}$$

The Gershgorin's discs of this matrix is depicted in Fig (2.1). For the Adaptive filters to have the fastest convergence, we wish $\mathbf{R}$ to be an identity matrix, with centers at $(\sigma_u^2, 0)$ and Radius 0. We can clearly see that now how the matrix in (2.3) worsen the convergence. The matrix has very large radii, and high eigenvalues spread. Now, for the case of (2.1), where we have transformed normalized input with DST, the Gershorgorin's discs are depicted in Fig(2.2). The DST matrix decreased the radii, compared to the original matrix $\mathbf{R}$, and based on the theorem, this means the non-diagonal elements are smaller, which means whiter autocorrelation matrix. Fig (2.3), it shows how the DCT is superior to the DST for this process, where the non-diagonal elements are significantly diminished. For AR(1), with $\rho = -0.9$, Fig (2.4) and Fig (2.5) depict how the DST is better than DCT, for negative poles.

2.5.2 Transient Analysis of Transform Domain Filters

In this section we conduct a customized transient analysis of th DCT-LMS algorithms. Th same procedure is replicated for the DST-LMS. Based on the results of the Gersgorin's theorem, we can see that, for DCT-LMS for example, the AR(1) with positive pole, the transformation almost completely de-correlates the input, and results in Gaussian regressors with diagonal autocorrelation matrix, and dif-

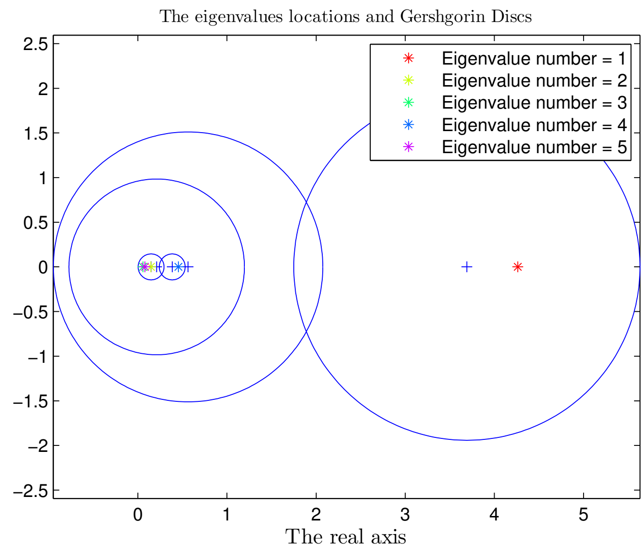

Figure 2.2: Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = 0.9$ and DST transformation

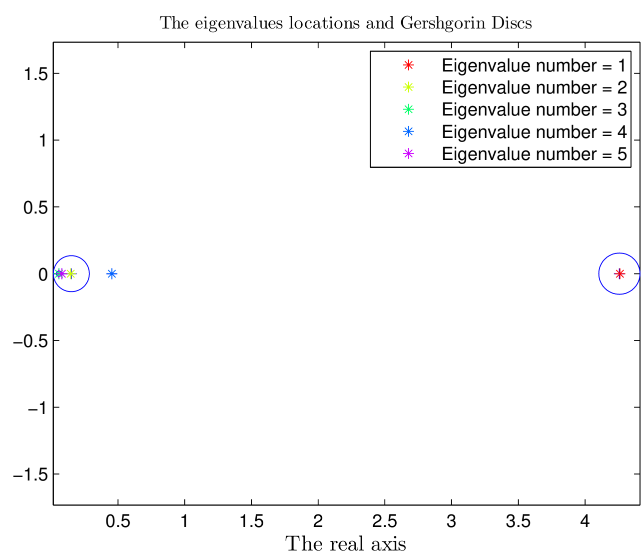

Figure 2.3: Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = 0.9$ and DCT transformation

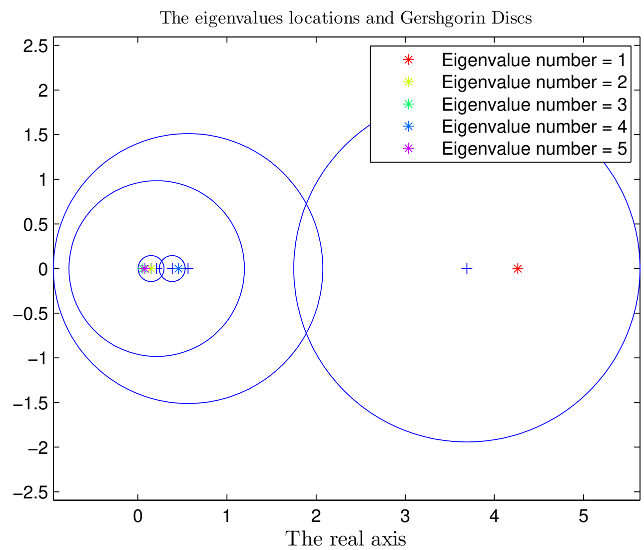

Figure 2.4: Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = -0.9$ and DCT transformation

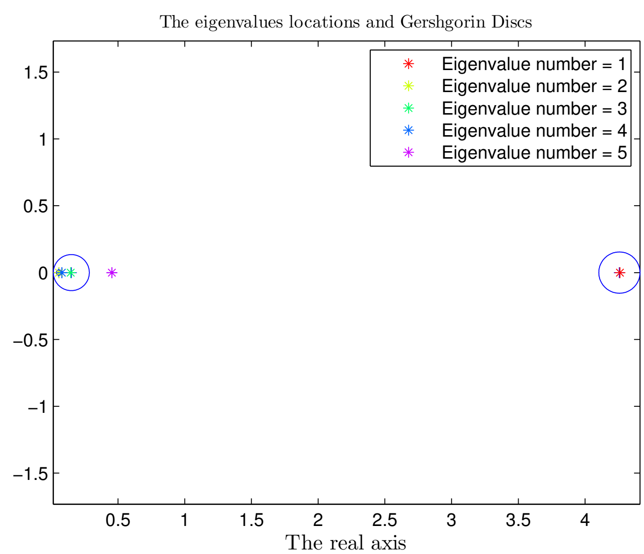

Figure 2.5: Gershgorin's Discs and locations of Eigenvalues of AR(1), with $\rho = -0.9$ and DST transformation

ferent entries in the diagonal, i.e. colored uncorrelated Gaussian. Figure (2.6) explains the claim of Gaussianty as well. We launch the analysis by the energy conservation relation for Matrix data non-linearities, where the LMS recursion is defined as following:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{H}[\mathbf{u}_i]\mathbf{u}_i^T e(i) \tag{2.4}$$

$$e(i) = d(i) - \mathbf{u}_i \mathbf{w}_{i-1} \tag{2.5}$$

the Energy conservation relation is then:

$$\|\mathbf{u}_i\|^2_{H\Sigma H}\|\tilde{\mathbf{w}}_i\|^2_2 + \|e_a^{H\Sigma}(i)\|^2 = \|\mathbf{u}_i\|^2_{H\Sigma H}\|\tilde{\mathbf{w}}_{i-1}\|^2_2 + \|e_p^{H\Sigma}(i)\|^2 \tag{2.6}$$

and the variance relation is given by:

$$E\|\tilde{\mathbf{w}}_i\|^2_{\Sigma} = E\|\tilde{\mathbf{w}}_i\|^2_{\bar{\Sigma}} + \mu^2\sigma_v^2 E\|\mathbf{u}_i\|^2_{H\Sigma H} \tag{2.7}$$

$$\bar{\mathbf{\Sigma}} = \Sigma - \mu\Sigma E(H[\mathbf{u}_i]\mathbf{u}_i^T\mathbf{u}_i) - \mu E(H[\mathbf{u}_i]\mathbf{u}_i^T\mathbf{u}_i)\Sigma \tag{2.8}$$

where are are using here the small step size approximation. The matrix data non linearity is defined here by the power normalization matrix as following:

$$H[\mathbf{u}_i] = \mathbf{\Lambda}^{-1}(i) \tag{2.9}$$

where $\mathbf{\Lambda}(i)$ is a diagonal matrix with elements given by the following exponential windowing setting:

$$p_k(i) = \beta p_k(i-1) + (1-\beta)|u_k(i)|^2 \tag{2.10}$$

and β is the smoothing parameter. In this customized analysis,we benefit from the fact the transformed regressor is Gaussian in calculating the following expectations appear in variance relation given by (2.7) and (2.8),

$$\underbrace{E(H[\mathbf{u}_i]\mathbf{u}_i^T\mathbf{u}_i)}_{A}, \underbrace{E\|\mathbf{u}_i\|^2_{H\Sigma H}}_{B} \tag{2.11}$$

employing the separation principle , by assuming that the input signal is not too non-stationary, hence the power normalization matrix will not vary significantly with time, compared to the input signal.
$\underline{E(H[\mathbf{u}_i]\mathbf{u}_i^T\mathbf{u}_i)}$ Applying separation principle, we have:

$$A = E\mathbf{\Lambda}_i^{-1} E\mathbf{u}_i^T\mathbf{u}_i \tag{2.12}$$

We tackle $E\mathbf{\Lambda}_i^{-1}$ element by element, since it is a diagonal matrix,as following:

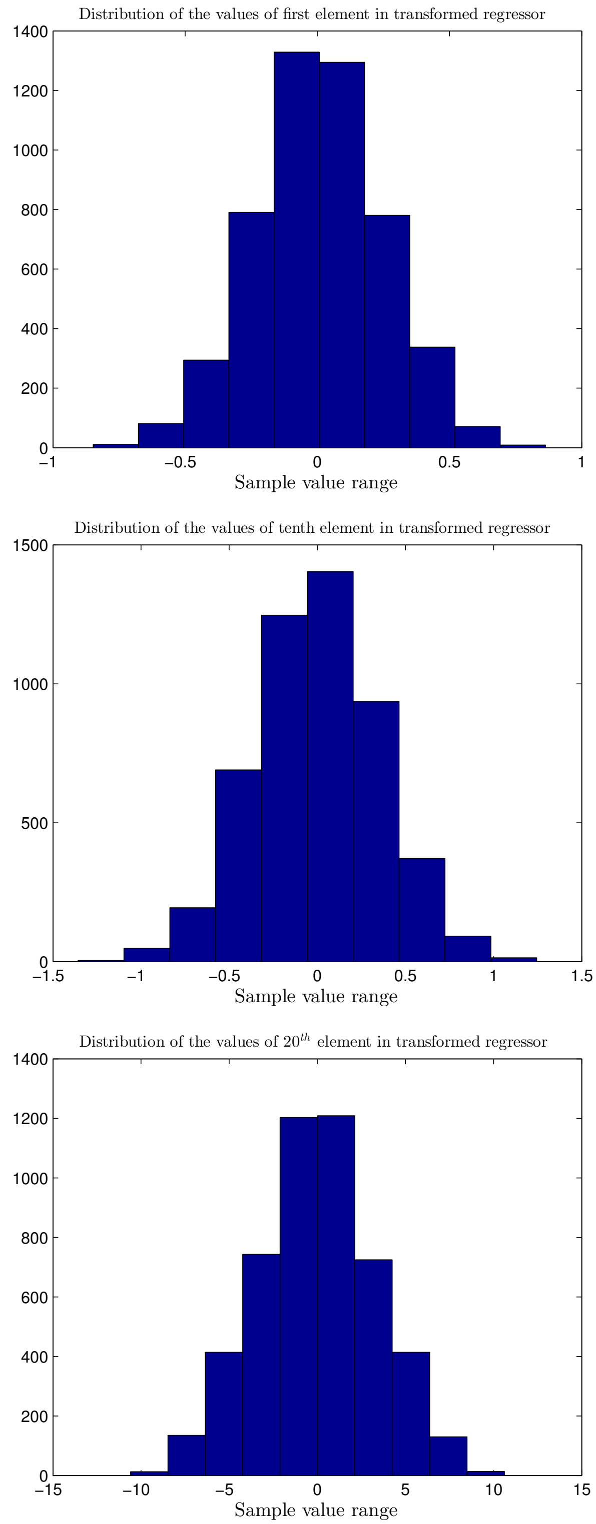

Figure 2.6: The distribution of some elements of the transformed regressors

$$
\begin{aligned}
E\mathbf{\Lambda}_i^{-1} &= \begin{pmatrix} E\frac{1}{p_1(i)} & 0 & 0 & 0 \\ 0 & E\frac{1}{p_2(i)} & 0 & 0 \\ \vdots & & & \vdots \\ 0 & 0 & E\frac{1}{p_{M-1}(i)} & 0 \\ 0 & 0 & 0 & E\frac{1}{p_M(i)} \end{pmatrix} \\
&\approx \begin{pmatrix} \frac{1}{Ep_1(i)} & 0 & 0 & 0 \\ 0 & E\frac{1}{Ep_2(i)} & 0 & 0 \\ \vdots & & & \vdots \\ 0 & 0 & E\frac{1}{Ep_{M-1}(i)} & 0 \\ 0 & 0 & 0 & E\frac{1}{Ep_M(i)} \end{pmatrix} \qquad (2.13)
\end{aligned}
$$

iterating (2.9), with zero initial conditions:

$$
Ep_k(i) = (1-\beta)\sum_{j=0}^{i} \beta^j E|u_(i-j)|^2 = \sigma_k^2(1-\beta^{i+1}) \qquad (2.14)
$$

A renders:

$$
\mathbf{A}(i) = \frac{1}{1-\beta^{i+1}}\mathbf{K}^{-1}\mathbf{R}_u \qquad (2.15)
$$

Following the same procedure, for term B:

$$
\begin{aligned}
\mathbf{B}(i) &= E\|\mathbf{u}_i\|^2_{EH^2\Sigma} = \mathbf{R}_u EH^2(i)\Sigma \qquad (2.16) \\
&= \frac{1+\beta}{3(1-\beta)(1-\beta^{2i+1})}\mathbf{K}^{-2}\mathbf{R}_u \qquad (2.17)
\end{aligned}
$$

using vector notation, equations (37)-(38) read:

$$
\begin{aligned}
E\|\tilde{\mathbf{w}}_i\|^2_{\sigma} &= E\|\tilde{\mathbf{w}}_i\|^2_{\bar{\sigma}} + \mu^2\sigma_v^2(\lambda^T\mathbf{B}(i)\sigma) \qquad (2.18) \\
\bar{\sigma} &= (\mathbf{I} - 2\mu A)\sigma = \mathbf{F}\sigma \qquad (2.19)
\end{aligned}
$$

the state-space recursion then is given by:

$$
\mathcal{W}_i = \mathcal{F}_i\mathcal{W}_{i-1} + \mu^2\sigma_v^2\mathcal{Y}_i \qquad (2.20)
$$

Where: [1]

$$\mathcal{W}_i = \begin{pmatrix} E\|\tilde{\mathbf{w}}(i)\|_{\sigma} \\ E\|\tilde{\mathbf{w}}(i)\|_{\mathbf{F}\sigma} \\ E\|\tilde{\mathbf{w}}(i)\|_{\mathbf{F}^2\sigma} \\ \vdots \\ E\|\tilde{\mathbf{w}}c(i)\|_{\mathbf{F}^(M-1)\sigma} \end{pmatrix}, \left[\mathcal{Y}\right]_k = \lambda^T \mathbf{F}^{k-1}\mathbf{B}(i)\sigma \tag{2.21}$$

$$\mathcal{F} = \begin{pmatrix} 0 & 1 & & & & \\ 0 & 0 & 1 & & & \\ 0 & 0 & 0 & 1 & & \\ \vdots & & & & & \\ 0 & 0 & 0 & & & 1 \\ -a_0 & -a_1 & -a_2 & -a_3 & \dots & -a_{Ma-1} \end{pmatrix} \tag{2.22}$$

and a_k are the coefficients of the characteristics polynomial of $\mathbf{F}$. The steady-state performance measure can be constructed from equations (37)-(38) directly,

$$MSD = \mu^2\sigma_v^2(\lambda^T\mathbf{B}(\infty)(\mathbf{I}-\mathbf{F})^{-1}\mathbf{q}) \tag{2.23}$$

$$EMSE = \mu^2\sigma_v^2(\lambda^T\mathbf{B}(\infty)(\mathbf{I}-\mathbf{F})^{-1}\lambda) \tag{2.24}$$

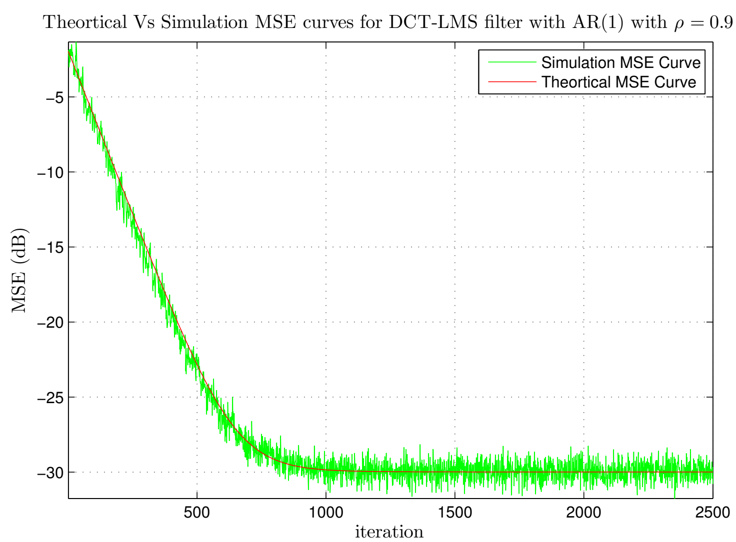

Figure 2.7: MSE Theoretical and Simulation Curves

[1]Note that, the state matrix changes with time, but after a while the change is extremely small, and we can treat the matrix as constant Matrix , and drop the time change notations, we use this fact for simulation. Similarly for $\mathcal{Y}_i$.

2.5.3 Transient Analysis of Sparse LMS Algorithms

In order to understand the behavior of the sparse LMS algorithm, it is mandatory to see how the missadjustment error vector ($\tilde{\mathbf{w}}_i = \mathbf{w}^o - \mathbf{w}_i$) evolves with time and how the sparsity enforcing term affects stability conditions and steady state level. We launch the analysis by the following recursion:

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu \mathbf{v}_i^T e(i) + \mathbf{s}_i \tag{2.25}$$

We follow an energy conservation argument [2], by using a weighted matrix Σ:

$$E\|\tilde{\mathbf{w}}_i\|_\Sigma^2 = \left(\tilde{\mathbf{w}}_{i-1} - \mu \mathbf{v}_i^T e(i) + \mathbf{s}_i\right)^T \Sigma \left(\tilde{\mathbf{w}}_{i-1} - \mu \mathbf{v}_i^T e(i) + \mathbf{s}_i\right)$$

Which, after expanding, becomes:

$$\begin{aligned} E\|\tilde{\mathbf{w}}_i\|_\Sigma^2 &= E\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 - \underbrace{\mu E \tilde{\mathbf{w}_{i-1}}^T \mathbf{v}_i^T e(i)}_{A} + \tilde{\mathbf{w}_{i-1}}^T \Sigma \mathbf{s}_i - \underbrace{\mu E \mathbf{v}_i^T \Sigma \tilde{\mathbf{w}_{i-1}} e(i)}_{B} \\ &+ \underbrace{\mu^2 E\|\mathbf{v}_i\|_2^2 e^2(i)}_{C} - \underbrace{\mu E \mathbf{v}_i \Sigma \mathbf{s}_i e(i)}_{D} + E\mathbf{s}_i^T \Sigma \tilde{\mathbf{w}_{i-1}} - \underbrace{\mu E \mathbf{s}_i \Sigma \mathbf{v}_i^T e(i)}_{E} + E\|\mathbf{s}_i\|_2^2 \end{aligned} \tag{2.26}$$

to conclude the relation, we progressively deal with terms A,B,C,D and E as following.
<u>Term A</u>

$$A = -\mu E\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma \mathbf{v}_i^T \mathbf{u}}^2$$

where we used the fac that $e(i) = \mathbf{u}_i^T \tilde{\mathbf{w}}_i + v(i)$, where $v(i)$ is the measurement noise modeled as white Gaussian and $\mathbf{u}_i$ is the non-transformed regressor. By invoking the independence assumption , term A renders:

$$A = -\mu E\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma E \mathbf{v}_i^T \mathbf{u}}^2$$

Following the same arguemnt, term B reads:

$$B = -\mu E\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma E \mathbf{u}_i^T \mathbf{v}}^2$$

[2]We can not follow the Energy conservation argument directly (except for the steady state calculations) because of the sparsity term which affects the flow of energy from iteration to iteration,not necceassry in a simiar way to the original energy conservation relation.

C,D and E summarize to:

$$\begin{aligned} C &= \mu^2 E\|\tilde{\mathbf{w}}_{i-1}\|_{E\|\mathbf{v}_i\|^2_\Sigma \mathbf{u}_i^T \mathbf{u}_i^2} + \mu^2\sigma_v^2 Tr(\mathbf{R}_v) \\ D &= \mu E \tilde{\mathbf{w}}_{i-1}^T \mathbf{u}_i^T \mathbf{v}_i \Sigma \mathbf{s}_i \\ E &= -\mu E \mathbf{s}_i^T \Sigma \mathbf{v}_i^T \mathbf{u}_i \tilde{\mathbf{w}}_{i-1} \end{aligned}$$

introducing A,B,C,D and E into (1.29) produce:

$$\begin{aligned} E\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma &= E\|\tilde{\mathbf{w}}_{i-1}\|_{\grave{\Sigma}} + E\mathbf{s}_i^T \mathbf{K}\tilde{\mathbf{w}}_{i-1} + E\|\mathbf{s}_i\|^2_\Sigma + \mu^2\sigma_v^2 Tr(\mathbf{R}_v) \\ \grave{\Sigma} &= \Sigma - \mu\Sigma E\mathbf{v}_i^T\mathbf{u}_i - \mu E\mathbf{u}_i^T\mathbf{v}_i\Sigma + \mu^2 E\|\mathbf{v}_i\|^2_\Sigma \mathbf{u}_i^T\mathbf{u}_i^2 \\ \mathbf{K} &= \Sigma - \mu\Sigma E\mathbf{v}_i^T\mathbf{u}_i - \mu E\mathbf{u}_i^T\mathbf{v}_i\Sigma \end{aligned} \tag{2.27}$$

assuming the step size is small enough, hence $\mathbf{K} = \grave{\Sigma}$. In order to diagonalize the set of equations in (2.27), we use $\mathbf{U}$ as an arbitrary unitary transformation and the matrices are redefined as following: $\bar{\mathbf{w}}_{i-1} = \mathbf{U}\tilde{\mathbf{w}}_{i-1}$, $\bar{\mathbf{u}}_i = \mathbf{U}\mathbf{u}_{i-1}$, $\bar{\Sigma} = \mathbf{U}^T\grave{\Sigma}\mathbf{U}$ and $\bar{s}_i = \mathbf{U}^T\mathbf{s}_i$. Employing the vector notation $\bar{\sigma} = vec(\bar{\Sigma})$, we have:

$$E\|\bar{\mathbf{w}}_i\|^2_{\bar{\sigma}} = E\|\bar{\mathbf{w}}_{i-1}\|^2_{\bar{\mathbf{F}}\bar{\sigma}} + E(\bar{\mathbf{s}}_i \odot \bar{\mathbf{w}}_{i-1})^T\bar{\sigma} + E\|\ \bar{\mathbf{s}}_i\|^2_{\bar{\sigma}} + \mu^2\sigma_v^2(\lambda_{\bar{v}}\bar{\sigma}) \tag{2.28}$$

where $\bar{\mathbf{F}} = \mathbf{I} - 2\mu\mathbf{R}_{\bar{w}\bar{v}}$. (2.28) can be introduced in a recursive state space model similar to the one deduced in [3] as following:

$$\mathcal{W}_i = \mathcal{F}\mathcal{W}_{i-1} + \mu^2\sigma_v^2\mathcal{Y} + \mathcal{S}_i + \mathcal{J}_i \tag{2.29}$$

where:

$$\left[\mathcal{Y}_k\right] = \mathbf{F}^{k-1}\bar{\sigma},\ \left[\mathcal{S}(k)\right] = E\|\bar{\mathbf{s}}_i\|^2_{\mathbf{F}^{k-1}\bar{\sigma}},\ \left[\mathcal{J}(k)\right] = E(\bar{\mathbf{s}}_i \odot \bar{\mathbf{w}}_{i-1})^T\mathbf{F}^{k-1}\bar{\sigma} \tag{2.30}$$

and the EMSE (or MSD, depending on the $\bar{\sigma}$ choice) is given by the first entry in $\mathcal{W}_i$.

Steady State EMSE and MSD

At the steady state, (2.28) reduces further into:

$$E\|\bar{\mathbf{w}}_\infty\|^2_{(\mathbf{I}-\bar{\mathbf{F}})\bar{\sigma}} = \mu^2\sigma_v^2(\lambda_v^T\bar{\sigma}) + E\|\mathbf{s}_\infty\|^2_{\bar{\sigma}} + E(\bar{\mathbf{s}}_\infty \odot \bar{\mathbf{w}}_\infty)^T\bar{\sigma} \tag{2.31}$$

By choosing $\bar{\sigma} = (\mathbf{I} - \bar{\mathbf{F}})^{-1}\lambda_{\bar{v}}$, we yield $EMSE(\infty)$ and for $\bar{\sigma} = ones(M,1)$, $MSD(\infty)$ is deduced. As an example, let us consider the ZA-LMS algorithm, where $\mathbf{u}_i = \mathbf{v}_i$ (without TPN) and $\mathbf{s}_i = \rho sgn(\mathbf{w}_{i-1})$. The $MSD(\infty)$ as a function of ρ becomes:

$$MSD_{ZA-LMS}(\infty) = \mu^2\sigma_v^2(\lambda_u^T\mathbf{q}) + M\rho^2 + \rho\sum_{k=1}^{M} Esgn(w_\infty(k))w_\infty(k) \qquad (2.32)$$

We infer from this relation that in addition to the Excess LMS error, the sparsity term introduces an excess error by itself ($M\rho^2$), and at the same time reduce the bias by the other term (which later we prove it is negativity). The optimum ρ, for the ZA-LMS is then:

$$\rho_{opt} = -\frac{\sum_{k=1}^{M} Esgn(w_\infty(k))w_\infty(k)}{2M} \qquad (2.33)$$

The MSD in (2.32) seems to be larger than the MSD without the sparse term in the regular algorithms. However, we should not that the kast term

which is clearly a function of the filter length. Before moving on, let us summarize. To the point, we have explained the transform domain algorithm and its convergence behavior. We have shed some light as well on the sparse adaptive algorithms, how the sparse term $\mathbf{s}_i$ affects the steady state. With these insights, we explain the proposed algorithm.

2.5.4 New LMF Algorithm

The Transform Domain LMF Algorithm

Consider a system identification scenario (as in figure 1.1) with input $\mathbf{u}_i$ and the desired output $d(i)$ defined by

$$d(i) = \mathbf{w}^o\mathbf{u}_i + n(i) \qquad (2.34)$$

where $\mathbf{w}^o$ is the optimal filter with length N. The transformed input vector is defined as

$$\mathbf{x}_i = \mathbf{u}_i\mathbf{T} \qquad (2.35)$$

where $\mathbf{T}$ is an $N \times N$ transformation matrix and the transformed input is $\mathbf{x}_i$. Both $\mathbf{x}_i$ and $\mathbf{u}_i$ are of length N. The transform domain LMF algorithm is given by

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu\mathbf{\Lambda}_i^{-1}\mathbf{x}_i^T e^3(i) \qquad (2.36)$$

where

$$\mathbf{w}_i = \mathbf{T}^T\bar{\mathbf{w}}_i \qquad (2.37)$$

where $\bar{\mathbf{w}}_i$ is the time domain weight vector and $\mathbf{w}_i$ is the transformed domain weight vector. μ is the step size, $e(i)$ is the error, the difference between the desired output and the output of the adaptive system, and $\mathbf{\Lambda}$ is the power normalization matrix. Clearly (3) does not exploit any sparsity information in its recursion. In order to exploit sparsity, we will alter the cost function to include a sparsity-aware term and then apply the general gradient algorithm formula:

$$\mathbf{w}_i = \mathbf{w}_{i-1} - \mu \mathbf{\Lambda}_i^{-1} \frac{\partial J}{\partial \mathbf{w}_i^T} \tag{2.38}$$

to yield the proposed TD-ZA-LMF algorithm.

Zero Attractor TDLMF Algorithm

The cost function of the zero attractor is given by

$$J_{ZA} = \frac{1}{4} e^4(i) + \lambda_{ZA} \|\mathbf{T}\mathbf{w}_i\|_1 \tag{2.39}$$

where λ_{ZA} is the zero attraction force (Lagrangian multiplier). Note that we are transforming back the coefficients in order to exploit sparsity, since the transformed weight vector is not sparse. Now,

$$\frac{\partial J_{ZA}}{\partial \mathbf{w}_i^T} = -e^3(i)\mathbf{x}_i^T + \lambda_{ZA} \mathbf{T}^T sgn(\mathbf{T}\mathbf{w}_i) \tag{2.40}$$

substituting (7) into (5) gives

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{\Lambda}^{-1} \mathbf{x}_i^T e^3(i) - \rho_{ZA} \mathbf{\Lambda}_i^{-1} \mathbf{T}^T sgn(\mathbf{T}\mathbf{w}_i)$$

where $\rho_{ZA} = \mu \lambda_{ZA}$. In contrast to the algorithm introduced in [79], this algorithm is designed for Non-Gaussian and correlated inputs. Moreover, it has the ability to attract all the filter coefficients to zero.

2.5.5 Performance Analysis of the TD-ZA-LMF

Convergence Analysis

In this section, we study the convergence in the mean of the TD-LMF sparse-aware algorithm. We launch the analysis by using the following general recursion:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{v}_i^T e^{2k-1}(i) - \mathbf{s}_i \tag{2.41}$$

where $\mathbf{s}_i$ is the sparsity penalty term and $k \in \mathbb{N}^*$, note that $k = 1$ and $k = 2$ result in the LMS algorithm and LMF algorithm, respectively.

Defining the weight error vector by $\mathbf{z}_i = \mathbf{w}^o - \mathbf{w}_{i-1}$ and employing the relation between the $\mathbf{z}_i$ and $e(i)$, that is,

$$e(i) = n(i) - \mathbf{z}_i^T \mathbf{v}_i \tag{2.42}$$

equation (8), reads:

$$\mathbf{z}_i = \mathbf{z}_{i-1} + \mu \mathbf{z}_i^T \{n(i) - \mathbf{z}_i^T \mathbf{v}_i\}^{2k-1}(i) - \mathbf{s}_i \tag{2.43}$$

Ignoring the high power of the $\mathbf{v}_i^T \mathbf{u}_i$ and applying statistical expectation to (10) results in

$$\begin{aligned} E[\mathbf{z}_i] &= E[\mathbf{z}_{i-1}] + \mu E[\mathbf{v}_i n_i^{2k-1}] \\ &- \mu(2k-1) E[n_i^{2k-2} \mathbf{v}_i \mathbf{v}_i^T \mathbf{z}_i] - E[\mathbf{s}_i] \end{aligned} \tag{2.44}$$

By modeling the measurement noise as a white Gaussian process, $E[\mathbf{v}_i n_i^{2k-1}] = 0$. The noise and regressor, using independence assumption, can be assumed to be independent from each other at steady state. The higher the noise plant power, the lower the SNR which implies higher $E[\mathbf{v}_i]$ value. But for very small step size, the effect of $E[\mathbf{v}_i]$ will be decreasing regardless of the noise plant power and the regressor power at steady state. Hence this independence assumption is valid for a small step size scenario. This assumption suggests the following relation:

$$\begin{aligned} E[n_i^{2k-2} \mathbf{v}_i \mathbf{v}_i^T \mathbf{z}_i] &= E[n_i^{2k-2}] E[\mathbf{v}_i \mathbf{v}_i^T] E[\mathbf{z}_i] \\ &= E[n_i^{2k-2}] \mathbf{R}_v E[\mathbf{z}_i] \end{aligned} \tag{2.45}$$

Finally, taking into account (12), (11) looks like:

$$E[\mathbf{z}_i] = \left\{ \mathbf{I} - \mu(2k-1) E[n_i^{2k-2}] \mathbf{R}_v E[\mathbf{z}_{i-1}] \right\} - E[\mathbf{s}_i] \tag{2.46}$$

Clearly the sparsity contribution, $E[\mathbf{s}_i]$, does not affect the convergence. Equation (13) is quite similar to that of the LMS convergence relation introduced in [3], and hence a necessary condition for the stability of (8) in the mean is that the step size, μ, should satisfy the following:

$$0 < \mu < \frac{2}{(2k-1) E[n_i^{2k-2}] Tr(\mathbf{R}_v)} \tag{2.47}$$

and in the case of TD-ZA-LMF, $k = 2$, this range should be:

$$0 < \mu < \frac{2}{\sigma_n^2 Tr(\mathbf{R}_v)} \tag{2.48}$$

where σ_n^2 is the variance of noise. and from [80], $Tr(\mathbf{R}_v) = M$, where M is the number of filter coefficients. The condition in (15) becomes:

$$0 < \mu < \frac{2}{\sigma_n^2 M} \tag{2.49}$$

where for sparse filters, M is known to be large, which reflects in a small step size.

Steady State Analysis

The steady-state analysis of the proposed algorithm is derived in this section. For the case of the TD-ZA-LMF, the recursion is governed by

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{v}_i^T e^3(i) - \mathbf{s}_i \tag{2.50}$$

where $\mathbf{v}_i^T = \mathbf{\Lambda}^{-1}\mathbf{x}_i^T$. Using energy conservation relation arguments leads to the following [3]:

$$E\|\widetilde{\mathbf{w}}_i\|_2 + E\frac{|e_a(i)|^2}{\|\mathbf{v}_i\|_2^2} = E\|\widetilde{\mathbf{w}}_{i-1}\|_2 + E\frac{|e_p(i)|^2}{\|\mathbf{v}_i\|_2^2} \tag{2.51}$$

where $\widetilde{\mathbf{w}_i} = \mathbf{w}_o - \mathbf{w}_i$ is the weight error vector, $e_a(i) = \mathbf{v}_i\widetilde{\mathbf{w}}_{i-1}$ and $e_p(i) = \mathbf{v}_i\widetilde{\mathbf{w}}_i$ are the posterior and priori errors, respectively. At steady state the following condition is always valid:

$$E\|\widetilde{\mathbf{w}}_i\|_2^2 = E\|\widetilde{\mathbf{w}}_{i-1}\|_2^2 \tag{2.52}$$

the Excess Mean Square Error (EMSE) is defined as

$$EMSE = E|e_a(i)|^2 \tag{2.53}$$

Hence equation (9) becomes:

$$E\frac{|e_a(i)|^2}{\|\mathbf{v}_i\|_2^2} = E\frac{|e_p(i)|^2}{\|\mathbf{v}_i\|_2^2} \tag{2.54}$$

the priori and a posteriori errors can be found to be related through the following recursion:

$$e_a(i) = e_p(i) - \mu\|\mathbf{v}_i\|^2 e^3(i) + \mathbf{v}_i\mathbf{s}_i \tag{2.55}$$

Then:

$$\begin{aligned} E\frac{|e_p(i)|^2}{\|\mathbf{v}_i\|_2^2} \quad &= \quad E\frac{|e_a(i)|^2}{\|\mathbf{v}_i\|_2^2} + \underbrace{E\mu^2 e^6(i)\|\mathbf{v}_i\|_2^2}_{A} + \underbrace{E\frac{\|\mathbf{s}_i\|^2_{\mathbf{v}_i\mathbf{v}_i^T}}{\|\mathbf{v}_i\|_2^2}}_{B} \\ &- \quad \underbrace{2\mu E e_a(i)e^3(i)}_{C} + \underbrace{e_a(i)\frac{\mathbf{v}_i\mathbf{s}_i}{\|\mathbf{v}_i\|_2^2}}_{D} \\ &- \quad \underbrace{\mu e^3(i)\mathbf{v}_i\mathbf{s}_i}_{E} \end{aligned} \tag{2.56}$$

In order to simplify (222) the following adopted assumptions are used:

A1: At steady state the error is independent from the input, and hence the a posterior error.

A2: The measurement noise is independent of the adaption error, the input and the sparsity enforcing term.

Moreover, we will also resort to the separation principle assumption, which states that, at steady-state the adaption error is independent of the input. Also, we will use the following relation expectation approximation [81]:

$$E\frac{\mathbf{v}_i^T\mathbf{v}_i}{\|\mathbf{v}_i\|_2^2} \approx \frac{E\mathbf{v}_i^T\mathbf{v}_i}{E\|\mathbf{v}_i\|_2^2} = \frac{\mathbf{R}_v}{Tr(\mathbf{R}_v)} \tag{2.57}$$

This approximation is quite helpful in simplifying the analysis, and suitable for long filters, which is the case for sparse filters. It should be noted that the TD-LMF algorithm inherits the same step size conditions for convergence.

We start the steady-state analysis by exploring each term of the right hand side of (222). By applying the separation principle, the term A reads:

$$A = \mu^2 Ee^6(i)E\|\mathbf{v}_i\|_2^2 \tag{2.58}$$

Using the following relation between the error, a posterior error and noise:

$$e(i) = e_a(i) + n(i) \tag{2.59}$$

the term A results in the following:

$$A \approx 45\mu^2 Tr(\mathbf{R}_v)\sigma_n^4 Ee_a(i)^2 + 15\mu^2 Tr(\mathbf{R}_v)\sigma_n^6 \tag{2.60}$$

The sparsity enforcing at steady-state is independent from the regressors. However, for the case of the transform domain, we will assume further that the input is not too non-stationary, and hence the power normalization matrix is almost constant and by employing the rational expectation approximation, the term B lands in:

$$E\frac{\|\mathbf{s}_i\|^2_{\mathbf{v}_i\mathbf{v}_i^T}}{\|\mathbf{v}_i\|_2^2} = E\frac{\|\mathbf{s}_i\|_{\mathbf{R}_v}}{Tr(\mathbf{R}_v)} \tag{2.61}$$

Using **A2** and ignoring the high power terms, C approximates to

$$C \approx -6\mu\sigma_n^2 Ee_a(i)^2 \tag{2.62}$$

Also, the term D can be shown to be results in

$$\begin{aligned} D &= Ee_a(i)\frac{\mathbf{v}_i\mathbf{s}_i}{\|\mathbf{v}_i\|_2^2} \\ &= E\frac{\tilde{\mathbf{w}}_i^T\mathbf{v}_i^T\mathbf{v}_i\mathbf{s}_i}{\|\mathbf{v}_i\|_2^2} \\ &= E\frac{\tilde{\mathbf{w}}_i^T\mathbf{R}_v\mathbf{s}_i}{Tr(\mathbf{R}_v)} \end{aligned} \tag{2.63}$$

Finally, applying the same procedure for term D, E gives:

$$E = -6\mu\sigma_n^2 \tilde{\mathbf{w}}_i^T \mathbf{R}_v \mathbf{s}_i \tag{2.64}$$

Now, substituting the different terms in (222), the $EMSE$ reads

$$\begin{aligned} \eta &= \underbrace{\frac{5\mu\sigma_n^4 Tr(\mathbf{R}_v)}{2 - 9\mu\sigma_n^2 Tr(\mathbf{R}_v)}}_{EMSE_{TD-LMF}} \\ &+ \gamma \underbrace{\left\{ E\frac{\|\mathbf{s}_i\|_{\mathbf{R}_v}}{Tr(\mathbf{R}_v)} - \tilde{\mathbf{w}}_i Tr(\mathbf{R}_v)[6\mu\sigma_n^2 - \frac{1}{Tr(\mathbf{R}_v)}]\mathbf{s}_i \right\}}_{EMSE_{sparse}} \end{aligned} \tag{2.65}$$

where

$$\gamma = \frac{1}{6\mu\sigma_n^2 - 45\mu^2 Tr(\mathbf{R}_v)\sigma_n^4} \tag{2.66}$$

If the transformation and power normalization matrices are set to $\mathbf{I}$, the $EMSE$ for the TD-LMF falls back to that of the $EMSE$ for the LMF given in [3].

To get more insight of equation (31), we study the $EMSE$ for the two set of elements in the vector $\mathbf{w}_i$, namely, the non-zero (NZ) elements and zero (Z) elements. The $EMSE$ exerted for the non-zero elements is

$$EMSE_{m\in NZ} = \frac{5\mu\sigma_n^4 Tr(\mathbf{R}_v)}{2 - 9\mu\sigma_n^2 Tr(\mathbf{R}_v)} + \gamma E\frac{\|\mathbf{s}_i\|_{\mathbf{R}_v}}{Tr(\mathbf{R}_v)} \tag{2.67}$$

where for this group of weights, the sparsity enforcing term $\mathbf{s}_i$ is independent from the error $\tilde{\mathbf{w}}_i$. While, for the zero elements, the $EMSE$ is

$$\begin{aligned} EMSE_{m\in Z} &= \frac{5\mu\sigma_n^4 Tr(\mathbf{R}_v)}{2 - 9\mu\sigma_n^2 Tr(\mathbf{R}_v)} \\ &- E\tilde{\mathbf{w}}_i Tr(\mathbf{R}_v)(6\mu\sigma_n^2 - \frac{1}{Tr(\mathbf{R}_v)})\mathbf{s}_i \end{aligned} \tag{2.68}$$

Since for this group, the sparsity term $\mathbf{s}_i$ is expected to be zero, then from (32) and (33), we can see the tension in the EMSE between the NZ and Z elements. When the sparsity rate is low (i.e., $|Z| >> |NZ|$), the overall $EMSE$ will be lower than for the non-sparse-aware algorithm, which in this case the TD-LMF.

2.5.6 Simulation Results

In this section, the performance of the proposed TD-ZA-LMF, ZA-LMF, ZA-LMS and TD-ZA-LMS are evaluated using Monte Carlo simulations. The results are averaged over 500 runs. The transform used is the Discrete Cosine Transform

(DCT-II). The SNR is set to 5 dB, ($\sigma_n^2 = 0.31$) and $M = 32$, which sets the step size range to $0 < \mu < 0.098$. We choose $\mu = 0.005$.

2.5.7 Gaussian Noise Scenario

Two experiments are conducted for the input signal, Gaussian input and correlated one — under low SNR regime with measurement noise modeled as white Gaussian. For the correlated input, a first-order filter with transfer function $\frac{1}{1+0.9z^{-1}}$ is sued to generate the correlated samples. The system under identification has sparsity $m = 2$, with random location for the non-zero elements.

The zero-attraction parameters are set to allow all algorithms to reach the same level of mean-square-deviation (MSD), which allow a fair comparison of the algorithms. The values are give in Table (2.1).

Algorithm	First Experiment	Second Experiment
ZA-LMS	5.5×10^{-5}	1×10^{-5}
ZA-LMF	5×10^{-6}	9×10^{-6}
TD-ZA-LMS	4.5×10^{-5}	3.1×10^{-5}
TD-ZA-LMF	9×10^{-6}	9×10^{-6}

Table 2.1: Zero Attractor values for the different algorithms.

As can be seen from Fig.4, the LMF based algorithms reached the steady state faster compared to the LMS based ones. Because of the bias introduced by the power normalization matrix, the TD-ZA-LMF is slightly slower than the ZA-LMF.

Figure 5 depicts the ZA-LMF [79] algorithm is better than the ZA-LMS algorithm when the SNR is low. However, both algorithms suffer from very slow convergence due to the high correlated input. More importantly, The transform domain algorithms, though, are comparatively faster in convergence and lower in steady state level, and the TD-ZA-LMF is the best performing. This proves the power of the hybridisation of the transform domain and the sparse-aware ability in enhancing the convergence behavior of the LMF algorithm at low SNR.

2.5.8 Non-Gaussian Noise Scenario

This scenario deals with the case when the measurement noise is modeled as Non-Gaussian, here we are assuming it has uniform distribution [82]. In the case of Gaussian input as depicted in Fig (2.8), the LMF family gives noticeably difference in the steady state and convergence, compared to the LMS family.

Figures (2.9) and (2.10) deals with the case when the input is correlated with sparsity rate equals 2. Figure (2.8) shows that the steady state floor of the TD-ZA-LMF is lower by $-5\ dB$ of TD-ZA-LMS. As for the convergence comparison, in Fig (2.10), we can clearly see that under this conditions, TD-ZA-LMF is superior as well. The trend holds when sparsity rank is increased to 4, as in figures (2.11)

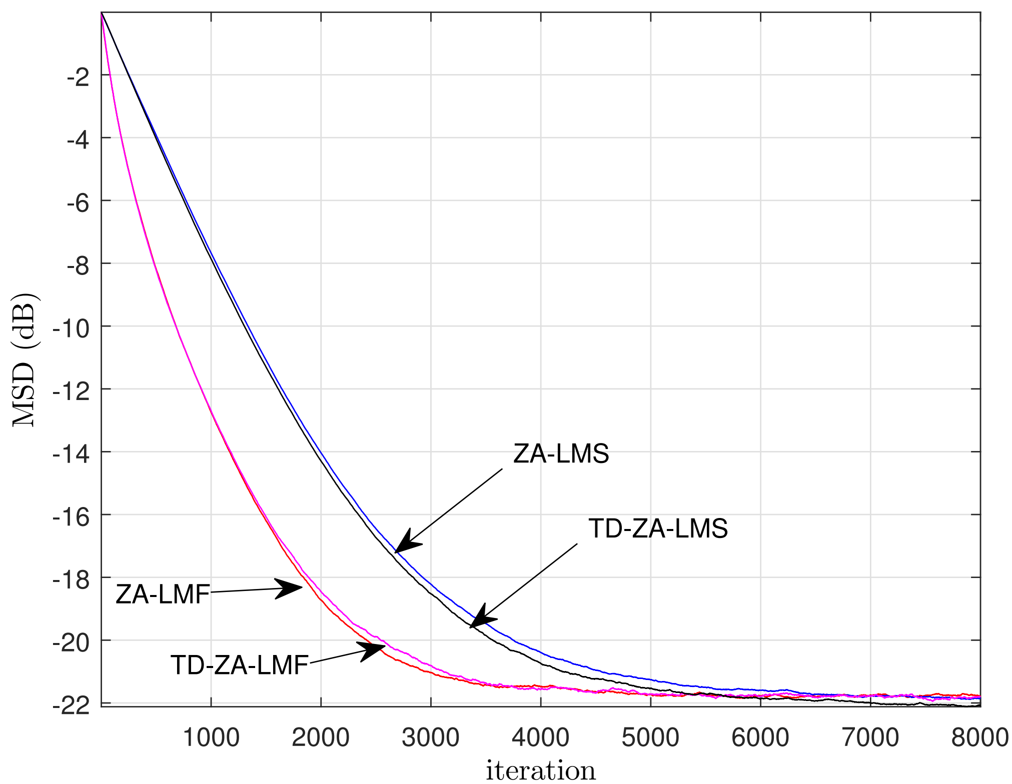

Figure 2.8: MSD curves for white Gaussian input with sparsity rate = 2.

and (2.12), however, we notice a degradation in the steady state floor of both algorithms, non-surprisingly as a result of increasing the non-zero elements.

2.6 Conclusion

By introducing the Zero Attractor to the TD-LMF, the resultant algorithm is able to explore sparse structure in correlated environment with low SNR. Compared to the existing methods, the proposed algorithm improved the convergence as confirmed by simulations. The proposed algorithm though, performed a bit slower in comparison with the ZA-LMF in case of white Gaussian input, because of the bias of the power normalization introduced in the zero-attractor.

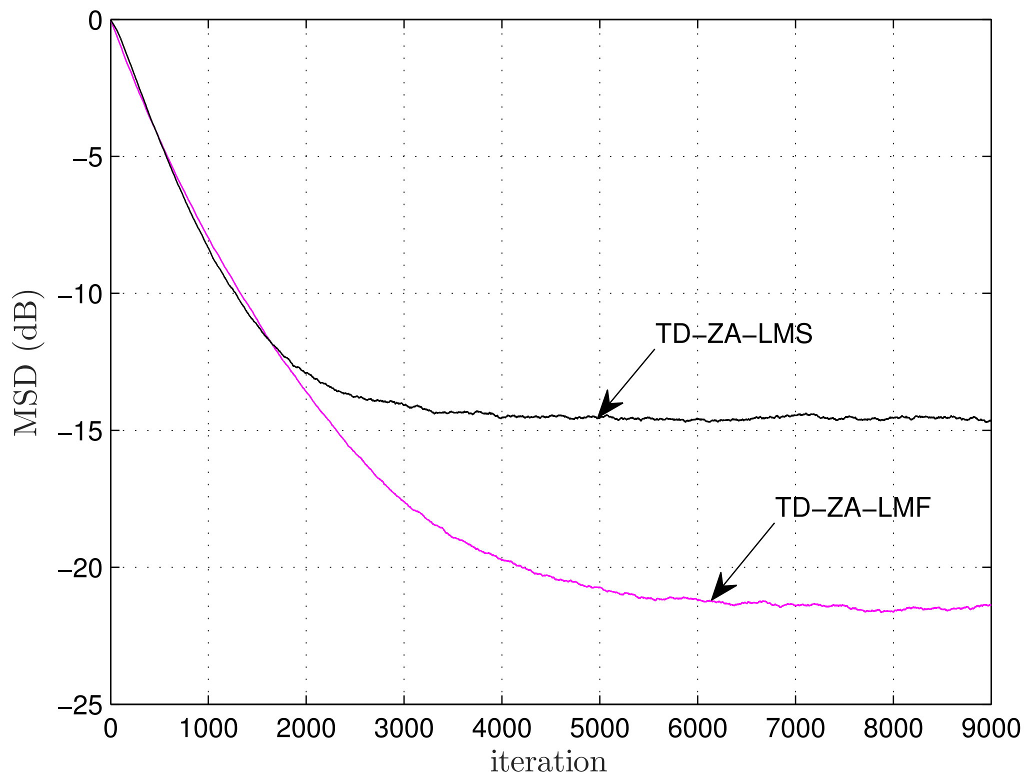

Figure 2.9: MSD curves for correlated input with sparsity rate = 2 and uniform measurement noise.

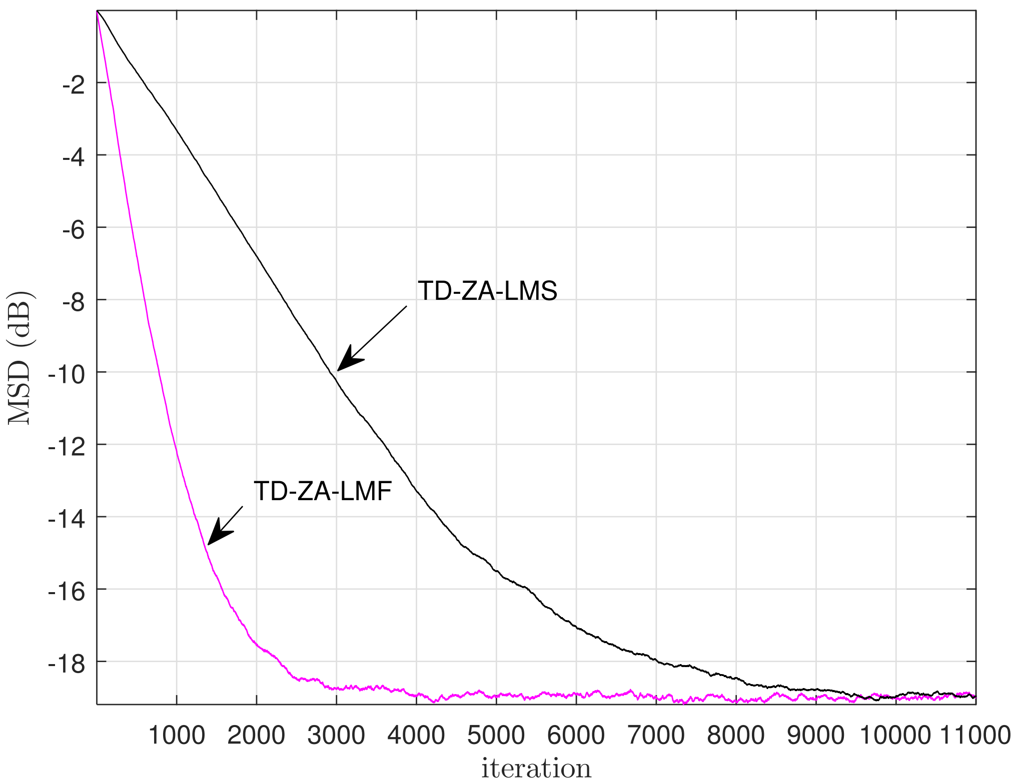

Figure 2.10: MSD curves for correlated input with sparsity rate = 2 and uniform measurement noise.

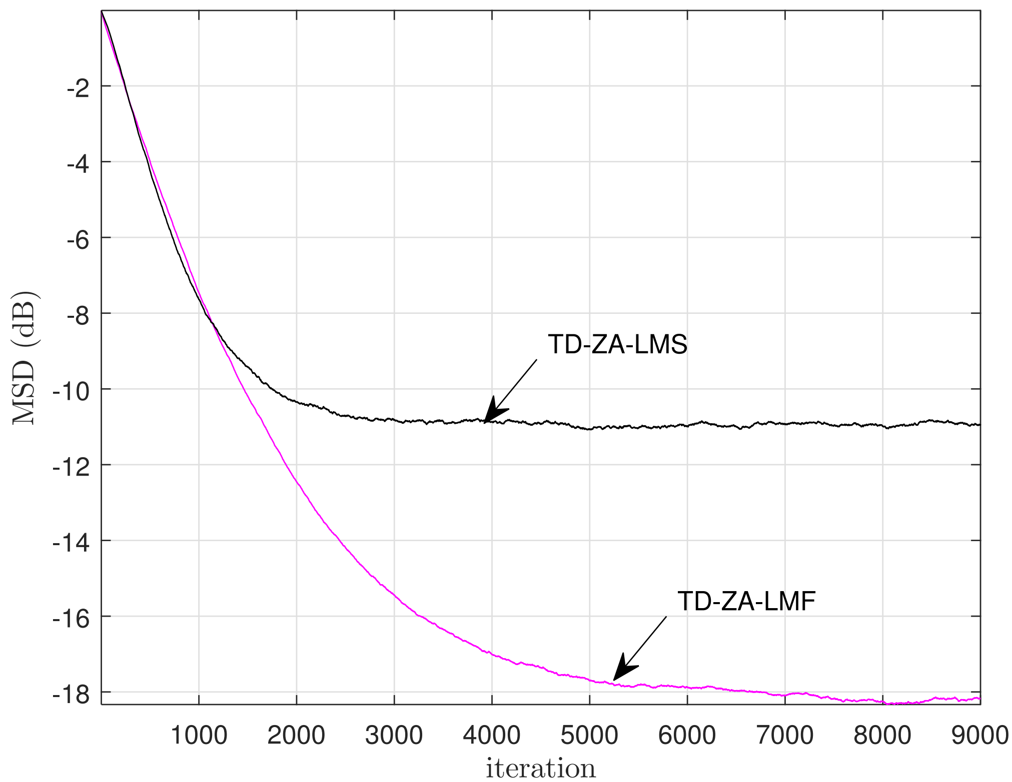

Figure 2.11: MSD curves for correlated input with sparsity rate = 4, with uniform measurement noise.

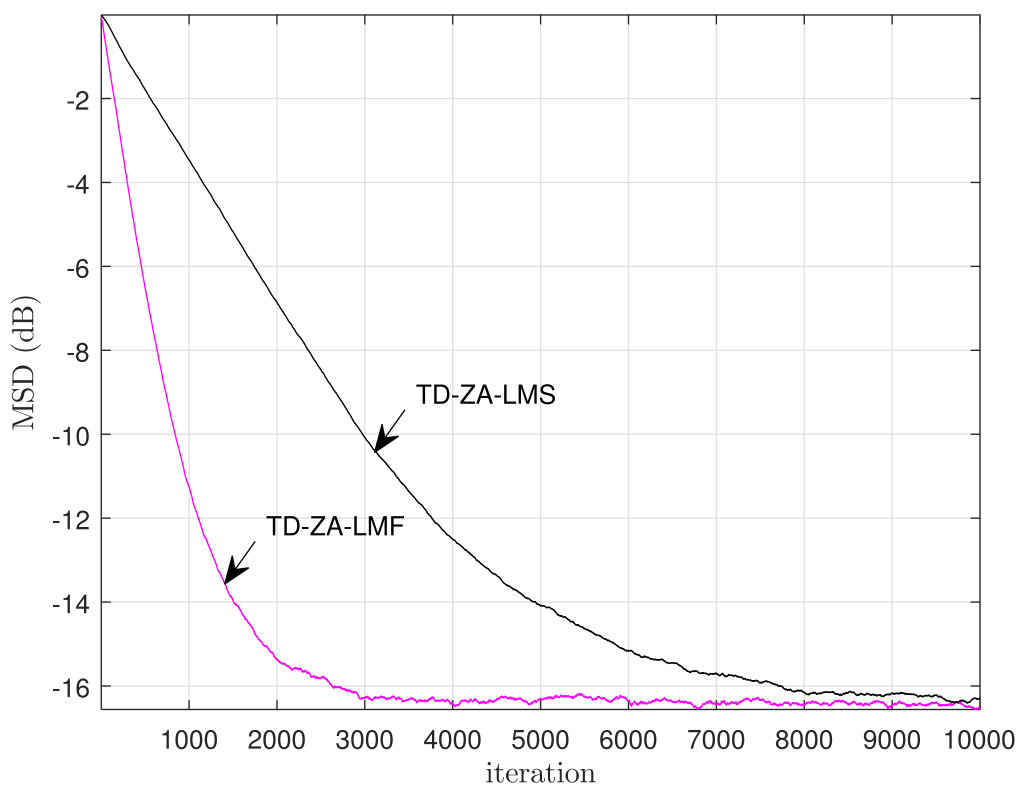

Figure 2.12: MSD curves for correlated input with sparsity rate = 4,with uniform measurement noise.

CHAPTER 3

CONVEX COMBINATION FILTERS FOR VARIABLE SPARSE SYSTEM IDENTIFICATION

3.1 Why Convex Combination Filters

To understand the motive ofr using this group of filters let us discuss the case of the LMS. One of the main characteristics of the LMS algorithm is relation between the step size (μ) and both convergence rate and steady state level. The relation between step size and convergence rate is proportional, that is, the higher the value of μ the faster the convergence to steady state. However, the relation between μ and the steady state is inversely proportional. The higher μ introduces worse steady state level. Hence, the choice of the step size μ can not offer both low steady state level *and* high convergence rate. The designer of the algorithm will have to make a trade off depending on the application on hand.

The convex combination filter is the answer to such following question: can we have LMS with *both* low steady state *and* high convergence rate?. For the case of LMS for example, we combine two filters in one filter. The first filter has high μ_1 and solves for high convergence. The second filter μ_2 solves for low steady state. The convex filter is the master filter that controls the component filters — and insures the low steady state and hihg convergence.

Generally, whenever there are conflicting demands of a filter, we can employ convex filter to reconcile. Here we will be employing it for the sparse adaptive algorithms. From (2.31), we see the MSD depends on the sparsity enforcing term $\mathbf{s}_i$. Assuming that we do not know the level of sparsity of the system we are employing, employing sparsity can be equally beneficial and deter mental. Hence, we combine two filters in one, the first is sparsity agnostic ($\mathbf{s}_i = 0$) and the second is sparsity aware. And the master convex filter will control both of them in a manner to always insures the lowest MSD regardless of sparsity level.

Before we introduce the algorithms in this chapter, let us first introduce the convex filter rigorously and shed some insightful results about its transient behavior.

3.2 Convex Combination Filter — Review

3.2.1 Steady State Universality

The convex filter is an aggregation of two filters. The convex filter is defined by the output $y(i)$ and the error $e(i)$ defined by the following relations: The output

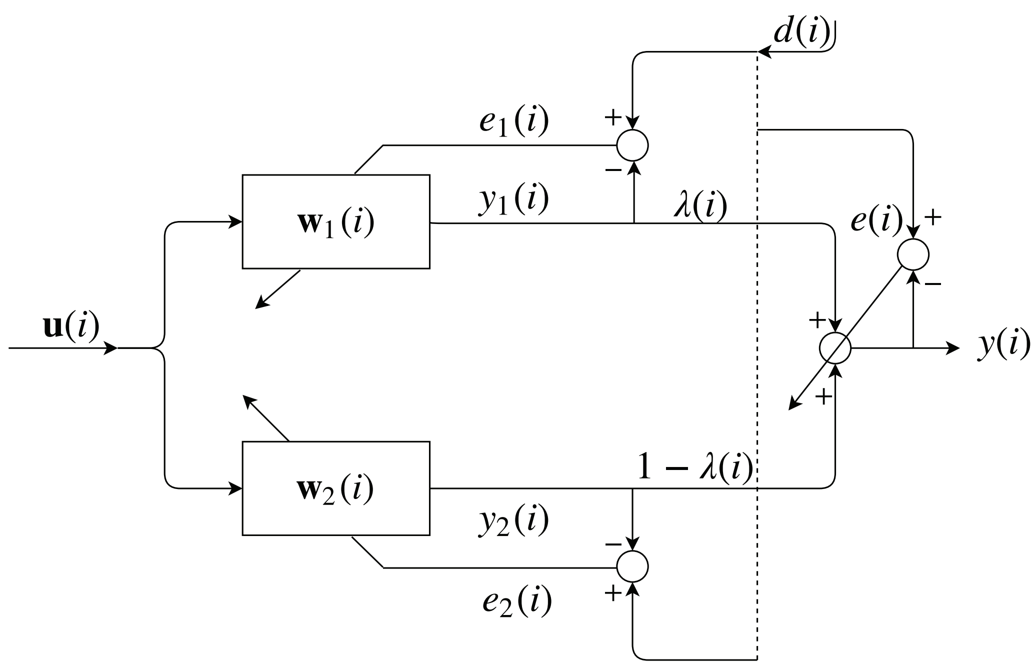

Figure 3.1: Diagram shows the Convex Combination filter proposed in [1].

of the convex filter is defined by the following equation:

$$y(i) = \lambda(i)y_1(i) + [1 - \lambda(i)]y_2(i) \tag{3.1}$$

$$e(i) = d(i) - e(i) \tag{3.2}$$

where $y_1(i)$ and $y_2(i)$ are the outputs from filters 1 and 2 respectively. The weight vector also follows similar convex relation:

$$\mathbf{w}_i = \mathbf{w}_{i,1}\lambda(i) + [1 - \lambda(i)]\mathbf{w}_{i,2} \tag{3.3}$$

The combination factor $\lambda(i)$ is given by the following equation:

$$\lambda(i) = \frac{1}{1 + e^{-a(i)}} \tag{3.4}$$

from this equation, it is obvious that the update recursion of the convex factor will be on $a(i)$ rather than $\lambda(i)$ directly, The recursion equation is given by:

$$\begin{aligned} a(i+1) &= a(i) - \frac{\mu_a}{2}\frac{\partial e^2(i)}{\partial a(i)} \\ &= a(i) + \mu_a e(i)[y_1(i) - y_2(i)]\lambda(i)[1-\lambda(i)] \end{aligned} \quad (3.5)$$

The universality, mathematically speaking, is to see where the combination goes as the nature of the system change, from sparse through semi-sparse to completely non-sparse. The combination should follow the change, and offer the lowest possible EMSE. In order to do this, we study the behavior of equation (1.47) at steady state. At steady state we assume that $\lambda(i)$ is independent of $e_{a,k}(i)$, we have the following equation describes the steady state behavior of $a(i)$:

$$E[a(i+1)] = E[a(i)] + \mu_a E[\lambda(i)[1-\lambda(i)]^2]\Delta J_2 - \mu_a E[\lambda(i)^2[1-\lambda(i)]]\Delta J_1 \quad (3.6)$$

where $\Delta J_1 = J_{ex,1}(\infty) - J_{ex,12}(\infty)$ and $\Delta J_2 = J_{ex,2}(\infty) - J_{ex,12}(\infty)$, equation(1.6) assumes that both algorithm has converged. And the equation describes the evolution of the $E[a(i)]$. W later we study equation(1.6) for the three cases of systems, considering that the two filers use the same step size.

3.2.2 Investigation of Transient Universality

In the previous section we explored the universality of the combination filter at the steady state and we found that combination will always switch to the component filter with lower MSD or MSE. In this section we would like to investigate the universality of the combination filter in the transient mode of operation.
By extending the universality definition of the steady state to the transient mode, we infer that, for a convex filter to be universal (non-asymptotically), the following condition is to be satisfied:

$$E[\mathcal{M}_\Sigma] < E\|\tilde{\mathbf{w}}_{i,k}\|^2_\Sigma \quad (3.7)$$

where $E[\mathcal{M}_\Sigma](i)$ is the performance metric (which depends on the choice of the matrix Σ, for example if $\Sigma = I$, then the metric reduces to the miss adjustment error of the combination filter). The condition in (2.1) says that, the transient universality requires the combination filter to insure a level of error always lower than its components, along the way to the steady state. Obviously, the inequality in (2.1) is very difficult to satisfy, and later on we will see how we can relax the condition to reach meaningful results.

Launching the investigation form the a posteriori error relations:

$$
\begin{aligned}
e_{a,1}(i) &= \mathbf{u}(i)\tilde{\mathbf{w}}_{i,1} &(3.8)\\
e_{a,2}(i) &= \mathbf{u}(i)\tilde{\mathbf{w}}_{i,2} &(3.9)\\
e_a(i) &= \lambda(i)e_{a,1}(i) + [1-\lambda(n)]e_{a,2}(i) &(3.10)\\
\tilde{\mathbf{w}}_i &= \lambda(i)\tilde{\mathbf{w}}i,1 + [1-\lambda(i)]\tilde{\mathbf{w}}i,2 &(3.11)
\end{aligned}
$$

and if we took the square of (2.4), the squared error evolution is governed by:

$$
\begin{aligned}
\|e_a(i)\|^2 &= \lambda^2(i)\|\tilde{\mathbf{w}}_{i,1}\|^2_{\mathbf{u}^T(i)\mathbf{u}(i)} + [1-\lambda(i)]\lambda(i)\tilde{\mathbf{w}}^T_{i,1}\mathbf{u}^T(i)\mathbf{u}(i)\tilde{\mathbf{w}}^2_{i,2}\\
&+ [1-\lambda(i)]^2\|\tilde{\mathbf{w}}_{i,2}\|_{\mathbf{u}^T(i)\mathbf{u}(i)} \qquad (3.12)
\end{aligned}
$$

now, we apply the statistical expectation with respect to $\lambda(i)$ [1], and applying the independence assumption (the error $\tilde{w}_i$ is independent of the input $\mathbf{u}(i)$), (2.6) reads:

$$
\begin{aligned}
E[\|e_a(i)\|^2|\lambda(i)] &= \lambda(i)^2E\|\tilde{w}_{i,1}\|^2_{\mathbf{R}_u} + [1-\lambda(i)]^2E\|\tilde{w}_{i,2}\|^2_{\mathbf{R}_u}\\
&+ \lambda(i)[1-\lambda(i)]\tilde{\mathbf{w}}^T_{i,1}\mathbf{R}_u\tilde{\mathbf{w}}_{i,2} \qquad (3.13)
\end{aligned}
$$

following the same procedure with (2.5), we reach:

$$
\begin{aligned}
E[\|\tilde{\mathbf{w}}(i)\|^2|\lambda(i)] &= \lambda^2(i)E\|\tilde{\mathbf{w}}\|^2_{1,i} + [1-\lambda(i)]^2E\|\tilde{\mathbf{w}}\|^2_{2,i}\\
&+ \lambda(i)[1-\lambda(i)]\tilde{\mathbf{w}}^T_{i,1}\tilde{\mathbf{w}}_{i,2} \qquad (3.14)
\end{aligned}
$$

both (2.7) and (2.8) can be seen as an instances of the following generalized recursion:

$$
\begin{aligned}
E[\mathcal{M}_\Sigma(i)|\lambda(i)] &= \lambda(i)^2E\|\tilde{w}_{i,1}\|^2_\Sigma + [1-\lambda(i)]^2E\|\tilde{w}_{i,2}\|^2_\Sigma\\
&+ \lambda(i)[1-\lambda(i)]E\{\tilde{\mathbf{w}}^T_{i,1}\Sigma\tilde{\mathbf{w}}_{i,2}\} \qquad (3.15)
\end{aligned}
$$

for example, if we set $\Sigma = \mathbf{R}_u$, the metric $\mathcal{M} = E[\|e_a(i)\|^2|\lambda(i)]$ and by defining the cross error by:

$$M_{12} = E\{\tilde{\mathbf{w}}^T_{i,1}\Sigma\tilde{\mathbf{w}}_{i,2}\} \qquad (3.16)$$

(2.9) reads:

$$E[\mathcal{M}_\Sigma(i)|\lambda(i)] = \lambda^2(i)\{M_1(i) + M_2(i)\} - \lambda(i)M_{12}(i) \qquad (3.17)$$

[1]We approach the problem by this setting for the following reason. Firstly, it is very difficilut to apply any kind of seperation principles betwenn the $\lambda(i)$ and $\tilde{w}_{i,k}$ and the second reason will be clear when we apply the Markov inequality, in which the conditional expectation is more becoming.

where:

$$M_k(i) = E\|\tilde{w}_{i,1}\|^2_\Sigma - M_{12}(i) \tag{3.18}$$

and as we said earlier, the condition in (2.1) is strict. We re-approach the problem by asking the question with a small twist; under what conditions, the probability of (2.1) to be satisfied is high (close to one), algebraically:

$$P[E[\mathcal{M}_\Sigma] < E\|\tilde{\mathbf{w}}_i\|^2_\Sigma] \approx 1 \tag{3.19}$$

by invoking the Markov's inequality, we have:

$$\begin{aligned}
P[E[\mathcal{M}_\Sigma] < E\|\tilde{\mathbf{w}}_i\|^2_\Sigma] &= 1 - P[E[\mathcal{M}_\Sigma] \geq E\|\tilde{\mathbf{w}}_i\|^2_\Sigma] \\
&= 1 - \frac{E_\lambda[E[\mathcal{M}_\Sigma]|\lambda(i)]}{E\|\tilde{\mathbf{w}}_{i,k}\|^2_\Sigma} \\
&= 1 - \frac{M_1(i) + M_2(i)}{M_k(i)} E\lambda^2(i) + \frac{2M_{12}(i)}{M_k(i)} E\lambda(i) \qquad (3.20) \\
&= 1 - f\{M_1(i), M_2(i).M_{12}(i), \lambda(i)\} \qquad (3.21)
\end{aligned}$$

to minimize the probability of exceedance —$f\{M_1(i), M_2(i).M_{12}(i), \lambda(i)\}$ — by taking the derivative with respect to $E\{\lambda(i)\}$ [2], and the optimum is:

$$E\lambda_o(i) = \frac{2M_{12}(i)}{M_1(i) + M_2(i)} \tag{3.22}$$

3.3 Convex Combination of LMF and ZA-LMF for Sparse System Identification

The Weiner filtering problem depends heavily on the assumption of the Gaussianty of the adaption error, which explains why we ar using the second moment of the error in the minimization problem that result in the famous LMS algorithm [83]. The LMS performs exceptionally well when both the input regressors and the measurement noise are Gaussian under high SNR environments. These conditions, insures from the first iterations that the adaption error is Gaussain process. The distribution of the error however is a function of the signal to noise ratio. The following figure depicts the relationship between the excess kurtosis and the SNR , for the adaption error at its first iteration in the LMS algorithms. We sue the excess kurtosis because it measure how much the distribution in our hand is similar to normal. If the excess kurtosis is zero, we have a perfect normal distribution, where negative and positive distributions correspond to platykurtic and leptokurtic distributions respectively. As it can be seen from the figure, as the SNR decrease the distribution of the error is rendered into platykurtic

[2] Note that we used the Jensen's inequality to modify (2.14) into a function of $E\lambda(i)$.

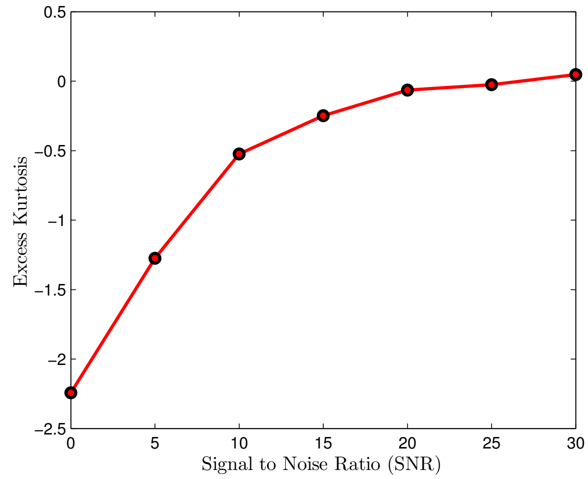

Figure 3.2: Excess Kurtosis versus SNR. Both the input and noise measurement follow Gaussian Distributions. The number of filter taps $M = 10$, with step size $\mu = 0.001$. The experiments are run 10000 times and the estimated Kurtosis is averaged over 50 experiments. We assume ergodicity of the error signal.

distribution, which is known for their large variations of observations. Hence, we infer from this experiment that the LMS at low SNR, it suffers convergence because it ignores to minimize the excess kurtosis. Where the LMF succeed in minimizing both the variance and the excess kurtosis simultaneously. We conclude from this observation that the LMF acknowledge the platykurtic distribution of the error in low SNR more than LMS, which explains the faster convergence and better MSE level described in [82].

3.3.1 Problem Formulation

The objective of this work is to introduce a solution to the problem of variable sparse system identification in low signal to noise ratio environments. The LMF is a sparse agnostic algorithm which for sparse systems it has higher excess MSE compared to its sparse aware companion called ZA-LMF given by:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i^T e^3(i) - \rho sign(\mathbf{w}_i) \tag{3.23}$$

which introduces lower EMSE in case of sparse optimum filter, however the EMSE starts to increase as the sparsity rate increases. Hence, both of the algorithms is not suitable (by itself) for the variable sparse system identification problem.
By combining the two algorithms, with the convex filter technique introduced in [1], the resultant algorithm is universal compared to its components. The universality here means that the convex filter performs as better as its best component elements (ZA-LMF and LMF in this case), in the steady state. The universality can be extended into the transient state, but this will not be covered for this case.

3.3.2 Convex Combination Filter — Review

The solution offered by the filter is given by the following equation:

$$\mathbf{w}(i) = \lambda(i)\mathbf{w}_1(i) + [1 - \lambda(i)]\mathbf{w}_2(i) \tag{3.24}$$

In order to insure that the combination is convex, we have to find a function of $\lambda(i)$ such that it is bounded from zero to one. Sigmoid f

3.3.3 Introduction

Since the introduction od the theory of Weiner filtering, the full density of the optimum solution (in a system identification setting)was taken fro granted as a fact. However, through time, it is found that the sparse system have quite noticeable frequent occurrence in the nature and practice. For example, in acoustic echo cancellation, it is found that the impulse response has small significant elements where the rest is negligibly small. The explanation of this phenomena is due to the large delays introduced in these type of systems [84]. Another interesting prominence is in wireless multi-path channels in communication systems, when the environment contains few signal reflectors implying that the impulse response of the channel is sparse— similar to the acoustic echo cancellation system [85]. Large reflectors are as well one of the features of the under sea communication channels, where the channel is found to be sparse as well [86]. Other types of systems has been found to have special structure as well, for example the envelope of the impulse response is exponential in special cases of acoustic echo cancellation, an additional fact that help the designer of adaptive algorithm to achieve better design . The sparsity of the optimal solution is considered even stronger as a feature and exploiting this information is a must for the algorithms designer [87]. This remarkable structure has triggered the research the sparse system identification, which is — nevertheless its novelty — has already produced a plethora of algorithms.

Before conceiving an abstract overview (before understanding the problem from sparsity structure vantage point) the solution is customized to the problem under consideration. This lead to introducing of family of algorithms based on proportionate update (PN) concept. PN algorithms assign adaption rate (step size) for each elements that is proportionate to tis value. PN-LMS is introduced firstly for solving the echo cancellation application [88], and its variants like μ-law MPNLMS [89] and Improved PNLMS (PNLMS) [90]. The PN-LMF algorithm has been recently proposed and analyzed in [91]. The PN technique rendered the sparse-agnostic algorithms (like LMS [4] and LMF [82]) to sparse aware. However there performance begins to deteriorate as the number of the significant elements increase.

Another line of research emerged after the advent of the compressive sensing and the LASSO operator techniques [92] [93]. The CS proposed solutions to the problem by using sparsity-recognizing norm l_0. However because of the difficulty on

optimizing this norm and its mathematical non-convexity properties, l_1 is ascribed instead to yield a real time solution to the problem. The LASSO technique is implemented with the MSE function of the LMS, and the resultant algorithm is called Zero Attractor LMS (LMS) [94]. The ZA-LMF algorithm is introduced in [79]. This LASSO technique resulted in lower steady state error and reduced the computational complexity at one shot. And to overcome the deterioration of EMSE of the ZA-LMS when sparsity increases, Re- weighted ZA-LMS, which assign weights for elements before taking the l_1 norm, i.e. which elements to be included in applying the attraction-to-zero force. The RZA-LMS (and RZA-LMF) are found to be very sensitive to the choice of there parameters, which implies that they can bot be considered as final remedy to the variable sparsity problem.
The convex combination filter — a filter that chooses between two component filters — is sought as a solution to the trade-off between tow algorithms [1]. For example, the tracking ability of the LMS is superior to the RLS under specific condition, and the converse is true for the RLS. by using a convex combination of the RLS and LMS [95], the trade-off is removed and the convex filter will always insure the best tracking ability, and the convex is called universal. And by looking at the variable sparsity identification as a trade-off between sparse-aware algorithm and sparse-agnostic,a convex combination is proposed with LMS and ZA-LMS as component filters [**?**].
The LMF is know to be superior to the LMS algorithm when the SNR is low and the measurement noise is non-Gaussian. In this work we propse a convex combination of LMF and ZA-LMF algorithm to solve the problem of variable sparse system identification under these conditions. The first section gives an insight about the mechanism of the LMF and why it is superior to the LMS in non-Gaussian environments with low SNR. In the second section we study the universality of the convex combination under three conditions, when the system is sparse, semi-sparse and completely dense. We conclude by the computer simulations, in which we found that the LMF convex combination offers lower steady state error compared to its counterpart, the LMS convex filter.

3.3.4 The universality of the combination filter

The universality in the excess mean square sense means that the convex filter always performs as good as its best component filter. Let us denote the EMSE of the LMF filter $J_{ex,1}(i)$ and the ZA-LMF filter with $J_{ex,2}(i)$. and to decide the universality we use the cross error between the two filters which is defined by:

$$J_{ex,12}(i) = E\{e_{a,1}(i)e_{a,2}(i)\} \tag{3.25}$$

We use this cross error as a metric to deiced what the filter choose in the specific condition. The conditions we have here are full sparse, semi-sparse or completely non-sparse system (dense). Before we study the universality we find the excess mean square errors, $J_{ex,1}(i), J_{ex,1}(i)$ and $J_{ex,cs}(i)$.

EMSE of LMF and ZA-LMF algorithms

We use the following general LMF recursion to launch the analysis of the EMSE.

$$\mathbf{w}_i = \mathbf{w}_{i-1} - \mathbf{s}_i + \mu e^3(i)\mathbf{u}_i^T \tag{3.26}$$

When the sparsity aware term $\mathbf{s}_i = 0$, we have the ordinary LMF and $\mathbf{s}_i = \rho sign(\mathbf{w}_i)$ results in ZA-LMF algorithm. We customize the analysis for the White Gaussian input, which dictates a linear relation between the EMSe and MSD, reads as:

$$MSD(\infty) = \sigma_u^2 EMSE(\infty) \tag{3.27}$$

[3].

by defining the misalignment error as $\mathbf{z}_i = \mathbf{w}_o - \mathbf{w}_i$The energy conservation is then derived as follows.

$$\begin{aligned} \|\mathbf{z}_i\|_2^2 &= \{\mathbf{z}_{i-1} + \mathbf{s}_i - \mu e^3(i)\mathbf{u}_i^T\}^T\{\mathbf{z}_{i-1} + \mathbf{s}_i - \mu e^3(i)\mathbf{u}_i^T\} \\ &= \|\mathbf{z}_{i-1}\|_2^2 - \mathbf{z}_{i-1}^T\mathbf{s}_i + \mu e^3(i)\mathbf{z}_{i-1}^T\mathbf{u}_i^T \qquad (3.28) \\ &+ \mathbf{s}_i^T\mathbf{z}_{i-1} + \mathbf{s}_i^2 - \mu e^3(i)\mathbf{s}_i\mathbf{u}_i \qquad (3.29) \\ &+ \mu e^3(i)\mathbf{u}_i\mathbf{z}_{i-1} - \mu e^3(i)\mathbf{u}_i\mathbf{s}_i + \mu^2 e^6(i)\|\mathbf{u}_i\|_2^2 \qquad (3.30) \end{aligned}$$

we assume here that all the signals are real [4]. The Energy relation then summarizes to:

$$\|\mathbf{z}_i\|_2^2 = \|\mathbf{z}_{i-1}\|_2^2 + \|\mathbf{s}_i\|_2^2 - 2\mathbf{z}_{i-1}^T\mathbf{s}_i + 2\mu e^3(i)\mathbf{z}_{i-1}\mathbf{u}_i^T - 2\mu e^3(i)\mathbf{u}_i\mathbf{s}_i + \mu^2 e^6(i)\|\mathbf{u}_i\|_2^2 \tag{3.31}$$

The adaption error is assumed to have a symmetrical PDF, and invoking the following assumptions:

1. The input is white Gaussian with Autocorrelation matrix $\sigma_u^2\mathbf{I}$.
2. $E\|\mathbf{z}_i\|_2^2 = E\|\mathbf{z}_{i-1}\|_2^2$ as $i|\infty$. (steady state definition).
3. The sparse enforcing term $\mathbf{s}_i$ is independent from the input regressor and adaption error. (this assumptions follow directly from the independent assumption).
4. The weight vector $\mathbf{w}_i$ is Gaussian at the steady state.
5. The measurement noise is white Gaussian with power σ_v^2.

[3]And if we choose $\sigma_u^2 = 1$, then we need one measure to evaluate the performance of the algorithm under consideration

[4]For sake of mathematical tractability, but it can be extended to the complex case.

$$\underbrace{2\mu Ee^3(i)\mathbf{z}_{i-1}\mathbf{u}_i^T}_{A} = \underbrace{\mu^2 Ee^6(i)E\|\mathbf{u}_i\|_2^2}_{B} + \underbrace{E\|\mathbf{s}_i\|_2^2}_{C} + \underbrace{2E\mathbf{z}_{i-1}^T\mathbf{s}_i}_{D} \tag{3.32}$$

We tackle the terms as following:
Term A

$$\begin{aligned} A &= 2\mu Ee^3(i)\mathbf{z}_{i-1}\mathbf{u}_i^T = 2\mu Ee^3(i)e_a(i) \\ &= 2\mu Ee_a^4(i) + 6\mu\sigma_v^2 Ee_a^2(i) \end{aligned} \tag{3.33}$$

Term B

$$\begin{aligned} B &= \mu^2 M\sigma_u^2 Ee^6(i) \\ &= \mu^2 M\sigma_u^2\{Ee_a^6(i) + 15\sigma_v^2 Ee_a^4(i) + 15\eta_v^4 Ee_a^2(i) + \eta_v^6\} \end{aligned} \tag{3.34}$$

Where $\eta_v^j = E|v(i)|^j$
Term C

$$C = E\|\mathbf{s}_i\|_2^2 = M\rho^2 \tag{3.35}$$

And $\rho = 0$ for LMF.
Term D
Clearly this term vanishes is the algorithm is sparse agnostic (LMF). For the case of the ZA-LMF algorithm, we segment the elements of the sparse optimum filter into two groups, the zero and the non-zero elements (i.e.significant).For the non-zeros elements (k $\in NZ$), the sparsity acknowledging term $\mathbf{s}_i$ will become independent of the filter update vector $\mathbf{w}_i$,while this fact does not hold for the zeros elements (k $\in$ Z).Then term D than can be expanded into its component as following:

$$\begin{aligned} D &= E\sum_{k\in Z} z(i-1,k)s(i,k) + \sum_{k\in NZ} E\{z(i-1,k)\}s(i,k) \\ &= -\rho E\sum_{k\in Z} w(i-1,k)sgn\{w(i-1,k)\} = \eta_z \end{aligned} \tag{3.36}$$

By invoking assumption 3 and the Price's theorem, we deduce the following:

$$Ew(i-1,k)sgn(w(i-1,k)) = aEw(i-1,k)^2 > 0 \tag{3.37}$$

$$a = \sqrt{\{(\frac{2}{\pi\sigma_{w,k}^2})\}} \tag{3.38}$$

By applying (1.11)-(1.14) into (1.10), we reach EMSE expression as following:

$$Ee_a(i)^2 = \underbrace{\frac{\mu M\sigma_u^2\eta_v^6}{6\sigma_v^2 - 15\mu M\sigma_u^2\eta_v^4}}_{\zeta_{LMF}} + \underbrace{\frac{M\rho^2 + \eta_z}{6\mu\sigma_v^2 - 15\mu^2 M\sigma_u^2\eta_v^4}}_{\zeta_{ZA-LMF}} \tag{3.39}$$

For the case of $\rho = 0$, (1.16) falls back into the EMSE expression of the LMF described in [3].Note that $M\rho^2 + \eta_z = \rho\{M\rho - \sqrt{\frac{2}{\pi}}(M-N)\sigma_w\} < 0$ [5],when the number of Non-zeros (N) is very small. From (1.17), we can see that it is always beneficial to apply a zero attractor to the SSI problem, since it will always insures lower EMSE — given that we have chosen the suitable attraction factor value ρ.What remains is to find cross excess mean square error for the combination filter defined in (1.3).

Cross Excess Mean Square Error, CEMSE

Starting with the following two recursions of the LMF and ZA-LMF respectively,

$$\mathbf{w}_1(i) = \mathbf{w}_1(i-1) + \mu e_1^3(i)\mathbf{u}_i^T \tag{3.40}$$

$$\mathbf{w}_2(i) = \mathbf{w}_2(i-1) + \mu e_2^3(i)\mathbf{u}_i^T - \mathbf{s}(i) \tag{3.41}$$

[6] and the relation between the CEMSE and the CMSD, for Gaussian input defined by:

$$CEMSE(i) = \sigma_u^4 CMSD(i) \tag{3.42}$$

we can follow an energy conservation argument similar to (1.6) as follows:

$$\begin{aligned} \mathbf{z}_1(i)^T\mathbf{z}_2(i) &= \mathbf{z}_1(i-1)^T\mathbf{z}_2(i-1) - \mu e_2^3(i)\mathbf{u}_i\mathbf{z}_1(i-1) \\ &+ \mathbf{z}_1^T(i-1)\mathbf{s}_i - \mu e_1^3(i)\mathbf{u}_i\mathbf{z}_2(i-1) \\ &+ \mu^2 e_1^3(i)e_2^3(i)\|\mathbf{u}_i\|_2^2 - \mu e_1^3(i)\mathbf{u}_i\mathbf{s}_i \end{aligned} \tag{3.43}$$

where $\mathbf{z}_i = \mathbf{w}_o - \mathbf{w}_i$ similar to the component filter, at the steady state the following condition is satisfied [1]:

$$E\{\mathbf{z}_1(i)^T\mathbf{z}_2(i)\} = E\{\mathbf{z}_1(i-1)^T\mathbf{z}_2(i-1)\} \tag{3.44}$$

then (1.20) reduces to:

$$\mu E e_1^3(i)e_{a,1}(i) + \mu E e_2^3(i)e_{a,2}(i) = \mu^2 M\sigma_u^2 E e_1^3(i)e_2^3(i) + E\mathbf{z}_1(i-1)^T\mathbf{s}(i) \tag{3.45}$$

By ignoring the higher power errors and recalling assumption 4, (1.22) further reduced to:

$$6\mu\sigma_v^2 J_{ex,12}(i) = 9\mu^2 M\sigma_u^2\eta_v^4 J_{ex,12}(i) + 9\mu^2 M\sigma_u^2\eta_v^6 + E\underbrace{\mathbf{z}_1(i-1)^T\mathbf{s}(i)}_{\zeta_z} \tag{3.46}$$

$$J_{ex,12}(\infty) = \frac{\mu M\sigma_u^2\eta_v^6}{6\sigma_v^2 - 9\mu M\sigma_u^2\eta_v^4} + \frac{\zeta_s}{6\mu\sigma_v^2 - 9\mu^2 M\sigma_u^2\eta_v^6} \tag{3.47}$$

[5]Practically, $\sqrt{\frac{2}{\pi}} >> \rho$

[6]we assume equal step sizes for the two recursions.

following the same argument in Term D (that leads (1.14)), we occlude:

$$
\begin{aligned}
\zeta_z &= -\rho \sum_{k \in Z} w_1(\infty,k) sgn(w_2(\infty,k)) = -a\rho \sum_{k \in Z} E\{w_1(\infty,k) w_2(\infty,k)\} \\
&= -a\rho \sum_{k \in Z} E\{z_1(\infty,k) z_2(\infty,k) \qquad (3.48)
\end{aligned}
$$

Where $a = \sqrt{\{(\frac{2}{\pi\sigma^2_{w,k}})\}}$, In to investigate the sign of ζ_z, we study the behavior of the cross miss-adjustment as $i \to \infty$. The next step is to prove that the term $E\{w_1(\infty,k)w_2(\infty,k)\}$ is positive in the steady state.

Sign of $E\{w_1(\infty,k)w_2(\infty,k)\}$ in the steady state:

Starting with the matrix version of (1.20):

$$
\begin{aligned}
\mathbf{z}_1(i+1) &= \mathbf{z}_1(i) - \mu e_1^3(i)\mathbf{x}_i^T \qquad (3.49) \\
\mathbf{z}_2(i+1) &= \mathbf{z}_2(i) - \mu e_2^3(i)\mathbf{x}_i^T + \rho sgn\{\mathbf{w}_2(i)\} \qquad (3.50) \\
E\{\mathbf{z}_2(i+1)\mathbf{z}_1^T(i+1)\} &= E\{\mathbf{z}_2(i) - \mu e_2^3(i)\mathbf{x}_i^T + \rho sgn\{\mathbf{w}_2(i)\}\}\{\mathbf{z}_1(i) - \mu e_1^3(i)\mathbf{x}_i^T\}^T \\
&= E\{\mathbf{z}_2(i)\mathbf{z}_1^T(i)\} - \mu E\{\mathbf{z}_2(i)\mathbf{x}_i e_1^3(i)\} - \mu E\{\mathbf{x}_i^T \mathbf{z}_1(i) e_2^3(i)\} \\
&+ \mu^2 E\{e_1^3(i) e_2^3(i)\mathbf{x}_i\mathbf{x}_i^T\} + \rho E\{sgn\{\mathbf{w}_2(i)\}\mathbf{z}_1(i-1)\} \\
&- \mu\rho E\{sgn\{\mathbf{w}_2(i)\}\mathbf{x}_i e_1^3(i)\} \qquad (3.51)
\end{aligned}
$$

by ignoring the higher power errors, and evoking assumption (5), and the famous independence assumptions, recursion(1.28) becomes:

$$
\begin{aligned}
\mathbf{c}(i+1) &= \mathbf{c}(i) - 3\mu\sigma_v^2\mathbf{c}(i)\mathbf{R}_x - 3\mu\sigma_v^2\mathbf{R}_x\mathbf{c}(i) + \rho\mathbf{b}(i) - \mu\rho\sigma_v^2\mathbf{b}(i)\mathbf{R}_x + 9\mu^2\eta_v^6\mathbf{R}_x \\
&+ 18\mu^2\eta_v^4\sigma_x^4\mathbf{c}(i) + 9\mu^2\eta_v^4\sigma_x^4 E\{\mathbf{z}_1^T(i)\mathbf{z}_2(i)\}\mathbf{1} \qquad (3.52)
\end{aligned}
$$

Recalling that the input correlation matrix of the input is $\sigma_x^2\mathbf{I}$,, we finally have:

$$
\begin{aligned}
\mathbf{c}(i+1) &= \mathbf{c}(i)[1 - 6\mu\sigma_v^2\sigma_x^2 + 18\mu^2\eta_v^4\sigma_x^4] + \rho(1 - \mu\sigma_v^2\sigma_x^2)\mathbf{b}(\mathbf{i}) \\
&+ 9\mu^2\sigma_x^2\eta_v^6\mathbf{1} + 9\mu^2\eta_v^4\sigma_x^4 E\{\mathbf{z}_1^T(i)\mathbf{z}_2(i)\}\mathbf{1} \qquad (3.53)
\end{aligned}
$$

Where $\mathbf{c}(i) = diag[E\{\mathbf{z}_2(i)\mathbf{z}_1^T(i)\}]$ and $\mathbf{b}(\mathbf{i}) = diag[E\{sgn\{\mathbf{w}_2(i)\}\mathbf{z}_1(i)\}]$. We pursue (1.30) more by grouping the elements of $\mathbf{c}(i)$ and $\mathbf{b}(i)$ into the zero $[k \in Z]$ and non zero $[k \in NZ]$ elements.
For the $\mathbf{b}(i) = [\mathbf{b}_{NZ}(i), \mathbf{b}_Z(i)]^T$, we notice that — for the non-zero (significant) elements — as $i \to \infty$, $E[sign\{w_{k,2}(i)\}z_{k,1}] = sign\{w^o_{k,2}\}E[z_{k,1}](i) = 0$. And for the zero elements, $E[sign\{w_{k,2}(i)\}z_{k,1}] = -aE[w_{k,2}(i)w_{k,1}] = -aE[z_{k,2}(i)z_{k,1}]$, where a is given in (1.15). Hence, $\mathbf{b}(i) = -a\mathbf{c}(i)$, when $i \to \infty$. Next, the

evolution of the cross weight error vectors, $E[\mathbf{z_1^T(i)z_2(i)}]\mathbf{1} = \lambda(i)$is given by:

$$\begin{aligned}\lambda(i+1) &= (1-6\mu\sigma_v^2\sigma_x^2+18\mu^2\eta_v^4\sigma_x^4)\mathbf{1}^T\mathbf{c}_{NZ}(i)\\ &+ (1-6\mu\sigma_v^2\sigma_x^2+18\mu^2\eta_v^4\sigma_x^4 - a\rho(1-\mu\sigma_v^2\sigma_x^2))\mathbf{1}^T\mathbf{c}_z(i)\\ &+ 9\mu^2 M\sigma_x^2\eta_v^6 + 9M\mu^2\eta_v^4\sigma_x^4\lambda(i) \end{aligned} \tag{3.54}$$

(1.31) convergence depends almost on $(1-6\mu\sigma_v^2\sigma_x^2+18\mu^2\eta_v^4\sigma_x^4) = \alpha$,since $\rho \approx 0$. The attraction factor is normally chosen to be small since it introduces high EMSE for high ρ, for the non-zero elements. For the convergence condition given in [82],that is, $0 < \mu < \frac{1}{6\sigma_v^2 Tr(\mathbf{R}_x)} = \frac{1}{6M\sigma_v^2\sigma_x^2}$. $\alpha < 1$. (1.31) is interesting (and descriptive) because it gives the convergence of the cross weight errors in terms of the cross weight error of the zero and non-zeros elements, $\mathbf{c}_Z(i)$ and $\mathbf{c}_{NZ}(i)$ respectively. By neglecting $\rho(.)$, we have:

$$\lambda(i+1) = \lambda(i)[1-6\mu\sigma_v^2\sigma_x^2+(9M+18)\mu^2\eta_v^4\sigma_x^4] + 9M\mu^2\sigma_x^2\eta_v^6 \tag{3.55}$$

$$\lambda(\infty) = \frac{9M\mu^2\sigma_x^2\eta_v^6}{\mu\sigma_x^2(6\sigma_v^2-(9M+18)\mu\eta_v^4\sigma_x^2)} > 0 \tag{3.56}$$

For large M [7] and Gaussian noise.
For $k \in Z$ (1.30) becomes:

$$c_k(i+1) = c_k(i)[1-6\mu\sigma_v^2\sigma_x^2+18\mu^2\eta_v^4\sigma_x^4] + 9\mu^2\sigma_x^2\eta_v^6 + 9\mu^2\eta_v^4\sigma_x^4\lambda(i) \tag{3.57}$$

$$c_k(\infty) = \frac{9M\mu^2\sigma_x^2(\eta_v^4+\sigma_x^2\eta_v^6\lambda(\infty))}{\mu\sigma_x^2(6\sigma_v^2-(9M+18)\mu\eta_v^4\sigma_x^2)} > 0 \tag{3.58}$$

from (1.35) if follows that $\zeta_z < 0$ in (1.26). Lemma 1, summarizes the EMSE analysis of the three filter. Interestingly, from (1.41) and (1.43), we conclude that for a combination of two identical LMF filters, the EMSE is always less than the LMF. Which implies full universality of the combination — unlike the combination of tow identical LMS, which is near optimal [1].

[7]Sparse filter are naturally large.

Summary of EMSE:
To sum up, the EMSE of the LMF is given by:

$$\zeta_{LMF} = \frac{\mu M \sigma_u^2 \eta_v^6}{6\sigma_v^2 - 15\mu M \sigma_u^2 \eta_v^4} = J_{ex,2}(\infty) \tag{3.59}$$

And fro the ZA-LMF, EMSE is defined by:

$$\zeta_{ZA-LMF} = \zeta_{LMF} + S(\infty) = J_{ex,1}(\infty) \tag{3.60}$$

$$S(\infty) = \frac{\rho\{M\rho - \sqrt{\frac{2}{\pi}}(N-M)\sigma_w\}}{6\mu\sigma_v^2 - 15\mu^2 M \sigma_u^2 \eta_v^4} \tag{3.61}$$

Where $S(\infty) < 0$ for sparse systems (i.e. $N << M$), and $S(\infty) > 0$ for non-sparse systems where $n \approx M$. and the Cross-EMSE of the two filters is given by:

$$\zeta_{12} = \frac{\mu M \sigma_u^2 \eta_v^6}{6\sigma_v^2 - 9\mu M \sigma_u^2 \eta_v^4} + S_{12}(\infty) = J_{ex,12}(\infty) \tag{3.62}$$

$S_{12}(\infty) < 0$, regardless of the sparsity level.

Next , we study the combination behavior in the three state under consideration.

Non-sparse system — dense: In this case the number of significant elements in the optimal vector solution is very large. Hence, $S_{12}(\infty) \approx 0$ and $S(\infty) > 0$. which means $J_{ex,1} > J_{ex,2}$ for this case. The error difference quantities in (1.6) summarizes as follows:

$$\begin{aligned}\Delta J_2 &= J_{ex,2}(\infty) - J_{ex,12}(\infty) \\ &= \frac{6M^2\mu^2\sigma_u^4\eta_v^4\eta_v^6}{(+)} \approx 0 \end{aligned} \tag{3.63}$$

$$\Delta J_1 \approx \frac{\mu M \sigma_u^2 \eta_v^6}{6\sigma_v^2 - 15\mu M \sigma_u^2 \eta_v^4} - \frac{M\rho^2}{6\sigma_v^2 - 9\mu M \sigma_u^2 \eta_v^4} > 0 \tag{3.64}$$

Note that we have used the small step size approximation, which is apropos for the LMF filters, since we use very small step size to annihilate the probability of divergence[8]. This case corresponds to the second case introduced in [1]. And the evolution of $Ea(i)$ is described by:

$$E\{a(i+1)\} \leq [E\{a(i)\} - C]_{a^-}^{a^+} \tag{3.65}$$

where $C = \lambda^+(1-\lambda^+)^2(\Delta J_1 - \Delta J_2)$, indicating that the asymptotic value of $Ea(i) = -a^+$. In another words, the combination chooses the LMF over the ZA-LMF.

[8]The LMF can diverge even if we use step size that satisfies that insures the convergence conditions, for more information, please check [96] and the reference therein.

Semi-Sparse Systems

In this case the following condition is satisfied:

$$\Delta J_1 > 0, \Delta J_2 > 0 \tag{3.66}$$

since $\Delta J_{ex,1}(\infty) > \Delta J_{ex,2}(\infty) > \Delta J_{ex,1}(\infty)$ but we also have $\Delta J_{ex,2}(\infty) > \Delta J_{ex,12}(\infty)$ because $S_{12}(\infty) < 0$ — because the number of zero elements increased dramatically for this case compared to the first case (dense systems). The semi-sparse case then corresponds to the third case in [1], where the asymptotic combination factor is given by:

$$E[\lambda(\infty)] = \frac{\Delta J_2}{\Delta J_1 + \Delta J_2} > 0.5 \tag{3.67}$$

In this case the combination filter is completely optimal and generates EMSE lower than the EMSE of both components filter, i.e.:

$$J_{ex}(\infty) \leq min\{J_{ex,1}(\infty), J_{ex,2}(\infty)\} \tag{3.68}$$

Sparse Systems

For this case, $J_{ex,2}(\infty) > J_{ex,1}(\infty)$, since the $S(\infty) < 0$ because of the dominance of the zero terms. $J_{ex,2}(\infty) > J_{ex,1}(\infty)$ is also satisfied because $S_{12}(\infty) < 0$. Hence, $\Delta J_2 > 0$. What remains is ΔJ_2 which is given by:

$$\begin{aligned}
\Delta J_1 &= J_{ex,1}(\infty) - J_{ex,12}(\infty) &(3.69)\\
&\approx \frac{M\rho}{6\sigma_v^2}\left[\frac{\rho}{\mu} - \sqrt{\frac{2}{\pi}}(1-\frac{N}{M})\frac{\sigma_w}{\mu} + \sqrt{\frac{2}{\pi}}\frac{9}{2}\frac{\sigma_v^2}{\sigma_w}\right] &(3.70)\\
&= k\rho^2 + l\rho &(3.71)
\end{aligned}$$

and $k = \frac{M}{6\sigma_v^2\mu} > 0$ and $l = \sqrt{\frac{2}{\pi}}[\frac{\sigma_w}{\sigma_v^2} - \frac{9\sigma_v^2}{2\sigma_w}]$. Assuming that A.5 is satisfied. The behavior of ΔJ_1 is a function of the sign of l. we consider the two cases of sign as following:

$l > 0$

In this case $\Delta J_1 > 0$, and the sparse case matches again the third case introduced in [1], i.e., the combination filter introduces lower EMSE compared to the LMF and ZA-LMF.

$l < 0$

For this case $\Delta J_1 < 0$, for the range of $0 < \rho < \frac{l}{k}$. This case corresponds to the first case in [1]. In which, the evolution of the combination factor $a(i)$ is described by:

$$E\{a(i+1)\} \geq [E\{a(i)\} + C]_{a^-}^{a^+} \tag{3.72}$$

which leads eventually to the asymptotic value of $\lambda(i) = \lambda^+$, which implies that the combination filter switches to the ZA-LMF completely.

3.3.5 Computer Simulation

The convex combination is tested under low SNR (= 10) environment with measurement noise modeled with uniform process. The optimum system with number of taps $M = 80$, is firstly set to be very sparse with fixed value (= 1) at support $S = \{5\}$, then changed to semi sparse system with $S = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$, and completely non-sparse (dense) system for the last stage. The step sizes of the algorithms is 0.002, the zero-attractor factors are 2×10^{-6},1×10^{-4} for LMF and LMS based algorithms respectively. The step size of the convex combination filters (i.e. μ_a) has the value of 50,10 for LMF and LMS respectively [9].

It should be noted that the MSD as a measure of performance is equivalent to the EMSE, especially we are using Gaussian with unity input power, we can clearly see this fact from (1.9). The results of the first experiment is depicted in Figures [1.2 — 1.3]. For the first stage, the sparse system, the LMS convex combination converged faster to steady state, but with much higher miss-adjustment compared to the convex combination of the LMF. This trend continued over the three stages of the experiment, proving that the LMF is superior to the LMS under low SNR with uniform noise process.

The LMF convex filter followed the ZA-LMF completely— which reached steady lower state level and slightly faster.However, the MSD level of the ZA-LMF deteriorates as the density of the significant elements in the optimal vector increased. We conclude from the experiment that not only the LMF convex filter insured lower MSd level compared to the LMS combination, but it also reduced the dynamic range of the MSD compared to LMS.

[9]Unlike the conventional very high values of μ_a (in terms of 1000,10000), to avoid instability under the conditions of the experiment— this values are found by experiment

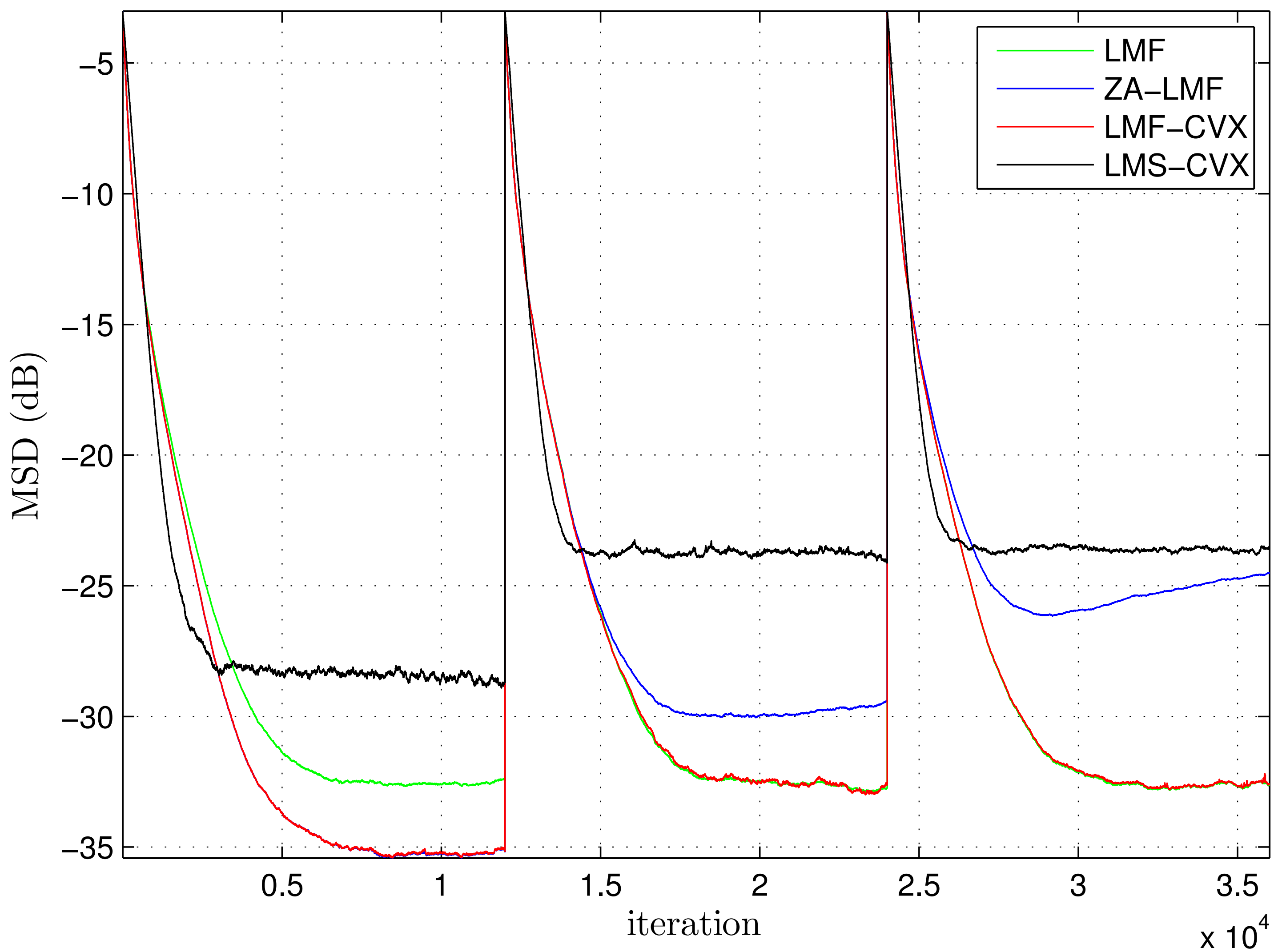

Figure 3.3: MSD curves for LMF, ZA-LMF algorithms and their convex combination. SNR level is 10 dB with uniform measurement noise.

3.4 Convex Combination of NLMF and ZA-NLMF algorithms

One of the problems of the LMF + ZA—LMF convex combination is stability, which is inherited from the instability of LMF. To resolve this issue we construct a convex combination of the stable NLMF [97] and ZA-NLMF algorithms. In addition to stability, an additional benefit of improved convergence under correlated environment is harvested. The stabilized NLMF algorithm has the form:

$$\mathbf{w}_{1,i} = \mathbf{w}_{1,i-1} + \frac{\mu_1 e_1^3(i)\mathbf{u}_{i,1}^T}{\|\mathbf{u}_{i,1}\|^2(\mathbf{u}_{i,1} + e_1(i))^2} \tag{3.73}$$

where we assume here the input is real. Similarly, ZA-NLMF [98]:

$$\mathbf{w}_{2,i} = \mathbf{w}_{2,i-1} + \frac{\mu_2 e_2^3(i)\mathbf{u}_{i,2}^T}{\|\mathbf{u}_{i,2}\|^2(\mathbf{u}_{i,2} + e_2(i))^2} - \rho sign\{\mathbf{w}_{2,i-1}\} \tag{3.74}$$

3.5 Convex Combination of TD-LMS and TD-AZ-LMS

3.6 abstract

The objective of this work is to introduce a convex combination of two filters to solve the problem of variable sparsity rate under highly correlated input environments. The two filters chosen are Transform Domain LMS (TD-LMS) algorithm and its sparse-aware L_1 version known as the Zero- Attractor Transform Domain LMS (TD-ZA-LMS) algorithm. This combination have the ability to converge to the sparse and non-sparse solutions in the case that the system is sparse or dense, respectively. The transform domain algorithms are known also for their ability to reach the steady state condition faster than the LMS algorithm , when the input is highly correlated. The analysis of the mentioned combination proved that the aggregation is universal; it will perform, at least, as the best of the two algorithms. Simulation results is performed to verify the universality of the combination.

3.7 Introduction

Least **M**ean **S**quare (LMS) algorithm proposed by [4] is well known in the adaptive filtering applications because of its simplicity and robustness, however, and since its advent, the highly correlated input proved to deteriorate the convergence rate of LMS [3]. Several approaches are used to remove this short coming. For example, in Normalized Least Mean Square (NLMS) instantaneous power normalization is employed to reduce eigenvalue spread and the approach is extended to Affine projection algorithms [99]. Transform Domain LMS (TD-LMS) introduced

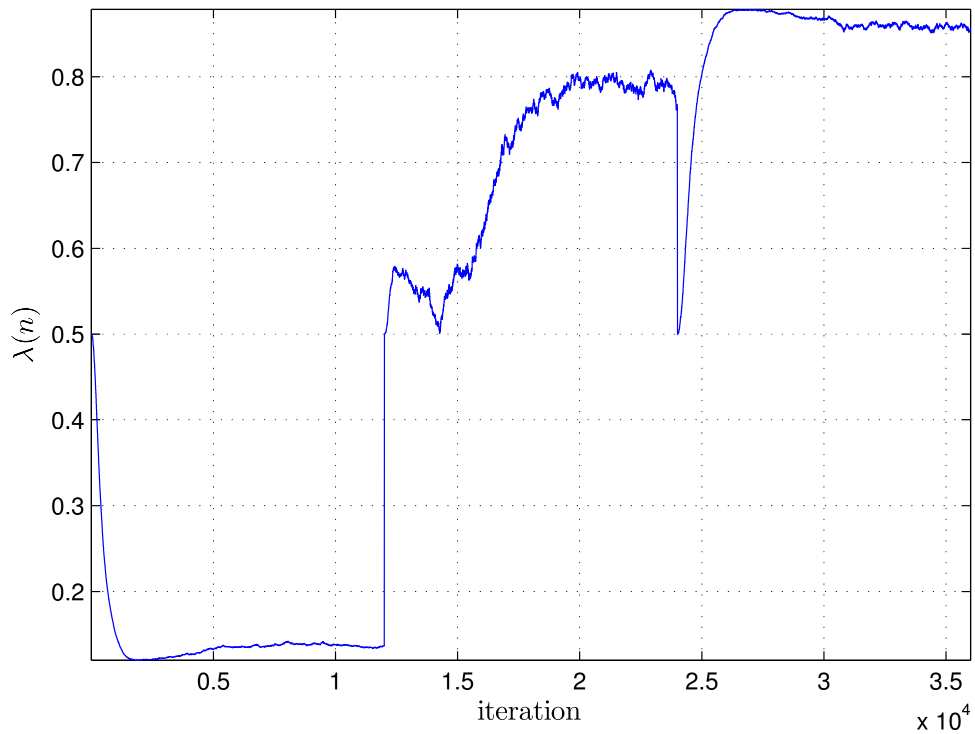

Figure 3.4: Combination factor evolution for the first experiment

by [6] employed unitary transformation like DFT and DCT accompanied with power normalization, which results in whitening the input. The whitening process geometrically results in rendering the contours of the error surface into circular contours, which significantly enhanced the convergence [100]. Both algorithms are extensively applied in wide range of applications from channel identification to Acoustic Echo Cancellation (AEC). It is also noted as well that the performance of the algorithms become worse when the impulse response under identification tends to be sparse [99]. Sparse-aware LMS [94] is developed by adding a zero attractor to the LMS recursion (ZA-LMS). The ZA-LMS proved to recognize sparse systems and offered lower error compared to LMS. The Transform Domain version of ZA-LMS is proposed by [101].

The level of sparsity of impulse is found to change with time in many applications. For example in AEC, the sparsity changes with the distance between the microphone and the speakers [84]. The same phenomena appears in wireless channels, where the environmental obstacles reallocation affects the sparsity rate. In these scenarios, we clearly can't use solely LMS neither ZA-LMS, because of the variable nature of the structure of the impulse response. The convex combination of ZA-LMS and LMS is prosed by [102], removed the dilemma of choice, and gave always the lower EMSE compared to its components regardless of the level of sparsity. The convex combination filter is used, whenever, we have a trade-off to make between two conflicting interests, as stated in [1].

The convex filter is proved to be universal, i.e, it always performs at least as better as the best of its components. TD-LMS with DFT and DCT transforms, are used as component filter in [103] and [104] respectively, with distinct step size for each component filter. In order to resolve the problem of slow convergence in the case of highly correlated input, in this work we introduce the transform domain version of [102].

In section II we review the convex filter and propose the algorithms that are used in our case. In the section III we venture into the steady state study of the component filter and follow that by universality study of the convex filter proposed in section IV. The performance of the filter is the tested in section V.

3.8 Convex Combination Filter

The convex combination filter as described in Fig.1, is an aggregation of two component filters, in our case one TD-LMS and second is TD-ZA-LMS filter. The

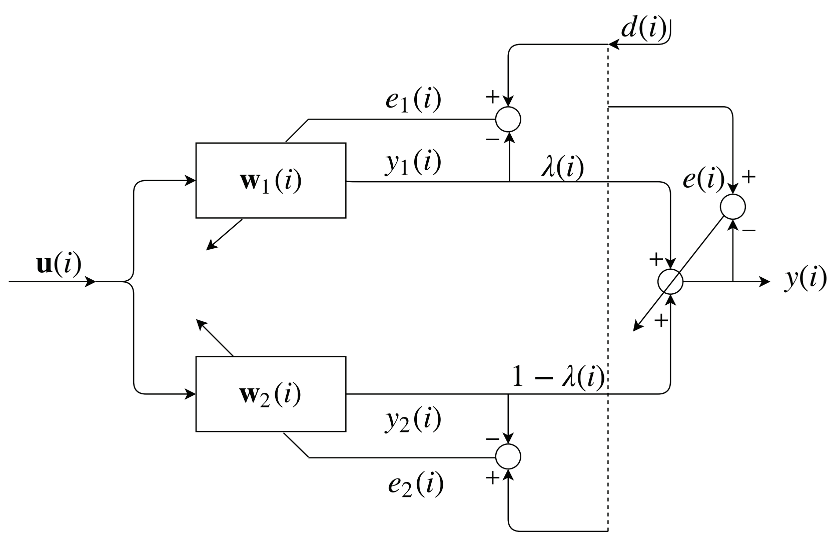

Figure 3.5: Diagram shows the Convex Combination filter proposed in [1].

output of the convex filter is defined by the following equation:

$$y(i) = \lambda(i)y_1(i) + [1 - \lambda(i)]y_2(i) \tag{3.75}$$

Where $y_1(i)$ is the TD-LMS filter and $y_2(i)$ is the ZA-TD-LMS. The weight vector also follows similar convex relation as (1) and is given by:

$$\mathbf{w}_i = \mathbf{w}_{i,1}\lambda(i) + [1 - \lambda(i)]\mathbf{w}_{i,2} \tag{3.76}$$

where $\mathbf{w}_{1,i}$,$\mathbf{w}_{2,i}$ are the estimates of TD-LMS and TD-ZA-LMS, respectively. The error of the convex filter is defined by:

$$e(i) = d(i) - y(i) \tag{3.77}$$

The error is minimized by updating the convex parameter $\lambda(i)$, indirectly, through the middle variable $a(i)$, the two variables $\lambda(i)$ and $a(i)$ are linked with sigmoid function of the following format:

$$\lambda(i) = \frac{1}{1 + e^{a(i)}} \tag{3.78}$$

and $a(i)$ is updated in a gradient descent setting with:

$$a(i+1) = a(i) + \mu_a e(i)[y_1(i) - y_2(i)]\lambda(i)[1 - \lambda(i)] \tag{3.79}$$

To insure the universality of the convex combination filter, $a(i)$ is set to the interval of $[a^+, a^-]$, which reflects back into $\lambda(i)$ in $[\lambda^+, \lambda^-]$. Limiting the range of the update coefficients endows the filter with continuous ability to learn, an essential requirement for universality.

3.9 Excess Mean Square Error for Component Filters

The cost function of the sparse aware algorithm is of the form [94]:

$$\min_{w} \quad \underbrace{f(e(i), \mathbf{w}_i, \mathbf{u}_i)}_{\text{the main estimation part}} \quad + \quad \underbrace{\bar{\mathbf{S}}_{\mathbf{i}}}_{\text{Sparsity enforcing condition}} \tag{3.80}$$

the sparsity enforcing term can take several forms, for example $\|\mathbf{w}_i\|_0$ or $\|\mathbf{w}_i\|_p$ and $\|\mathbf{w}\|_1$. we employ the general term in (6) since it will help to shed light on the mechanism of sparse algorithm. Applying stochastic descent approach to solve (6), the following general recurrence manifest itself:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{v}_i^H e(i) - \mathbf{s}_i \tag{3.81}$$

and $\mathbf{v}_i$ is defined by:

$$\mathbf{v}_i^H = \mathbf{\Lambda}_i^{-1} \mathbf{x}_i^H \tag{3.82}$$

where $\mathbf{x}_i$ and $\mathbf{w}_i$ represent the transformed orignial regressors and filter coefficients , respectively, and define:

$$\mathbf{x}_i = \mathbf{u}_i \mathbf{T} \tag{3.83}$$

$$\mathbf{w}_i = \mathbf{T}^H \bar{\mathbf{w}}_i \tag{3.84}$$

The transformation $\mathbf{T}$ is unitary transformation, which conserves the eigenvalue spread of the input $\mathbf{u}_i$, and the power normalization matrix $\boldsymbol{\Lambda}_i^{-1}$ shrinks the eigenvalue spread. When $\mathbf{s_i}$ is set to zero, the resultant algorithm is TD-LMS, otherwise, the algorithm is sparse-aware TD-LMS algorithm. and when the $\mathbf{s_i}$ has the following description:

$$\mathbf{s}_i = \rho \boldsymbol{\Lambda}_i^{-1} \mathbf{T}^H sgn[\mathbf{T}\mathbf{w}_i] \tag{3.85}$$

then recurrence (7) is called as the TD-ZA-LMS algorithm. Note that we are exposing the sparsity in the time domain. The transformation generally needs not to be real, but in this analysis we will assume this without loss of generality, since the analysis can be extended easily to the complex case.
The relation between a-priori and posteriori errors:

$$e_p(i) = e_a(i)(1 - \mu\|\mathbf{v}_i\|_2^2) - \mu\|\mathbf{v}_i\|_2^2 n(i) + \mathbf{v}_i\mathbf{s}_i \tag{3.86}$$

We use (12) to deduce the energy conservation relationship at the steady state, given by the following equation:

$$\begin{aligned} E\|e_a(i)\|^2 &= \frac{\sigma_n^2 Tr(\mathbf{R}_v)}{2 - \mu Tr(\mathbf{R}_v)} \\ &+ \frac{1}{2\mu - \mu^2 Tr(\mathbf{R}_v)} \Bigg\{ 2E\frac{e_a(i)\mathbf{v}_i\mathbf{s}_i}{\|\mathbf{v}_i\|_2^2} \\ &+ E\frac{\|\mathbf{s}_i\|_{\mathbf{v}^H\mathbf{v}}}{\|\mathbf{v}_i\|_2^2} - 2\mu E e_a(i)\mathbf{v}_i\mathbf{s}_i \Bigg\} \end{aligned} \tag{3.87}$$

Where we assumed that the measurement noise is independent of the input and the weights vector. Applying the separation principle, (13) reads:

$$\begin{aligned} EMSE &= \underbrace{\frac{\sigma_n^2 Tr(\mathbf{R}_v)}{2 - \mu Tr(\mathbf{R}_v)}}_{ns} \\ &+ \frac{1}{2\mu - \mu^2 Tr(\mathbf{R}_v)} \Bigg\{ E\frac{\|\mathbf{s}_i\|_{\mathbf{R}_v}}{Tr(\mathbf{R}_v)} \\ &+ 2E\widetilde{\mathbf{w}}_i^H \mathbf{R}_v \left(\frac{1}{Tr(\mathbf{R}_v)} - \mu \right) \mathbf{s}_i \Bigg\} \end{aligned} \tag{3.88}$$

The first term in (14) represents the EMSE introduced due to the non-sparse part of the algorithm, and the second term is due to introducing the sparse aware term to the algorithm. Note that the second term vanishes, when $\mathbf{s}_i = \mathbf{0}$, where equation (7) becomes an ordinary LMS algorithm, the EMSE is given by:

$$EMSE_{TD-LMS} = \frac{\sigma_n^2 \mu Tr(\mathbf{R}_v)}{2 - \mu Tr(\mathbf{R}_v)} \tag{3.89}$$

and for the ordinary LMS $\mathbf{R}_v = \mathbf{R}_u$. Which matches exactly the EMSE of the LMS given in [3] .For the case of sparse-aware algorithm (ZA-LMS , where $\mathbf{T} = \mathbf{I}$ and $\mathbf{s}_i = \rho sgn(\bar{\mathbf{w}})$), the EMSE behavior depends on the elements of the weight vector. For the non-zero elements in the sparse optimum vector, the EMSE equation reduces to:

$$\begin{aligned} EMSE_{k\in NZ} &= \frac{\sigma_n^2 \mu Tr(\mathbf{R}_v)}{2-\mu Tr(\mathbf{R}_v)} \\ &+ \frac{1}{2\mu - \mu^2 Tr(\mathbf{R}_v)} E \frac{\|\mathbf{s}_i\|^2_{\mathbf{R}_v}}{Tr(\mathbf{R}_v)} \end{aligned} \tag{3.90}$$

where in this case, we can assume that $\mathbf{s}_i$ is independent from $\widetilde{\mathbf{w}}$. For the zero-elements in the sparse vector, the expected value of $\mathbf{s}_i$ is zero (assuming the algorithms converges) and the EMSE is then:

$$\begin{aligned} EMSE_{k\in Z} = & \frac{\sigma_n^2 \mu Tr(\mathbf{R}_v)}{2-\mu Tr(\mathbf{R}_v)} \\ & - \frac{2}{2\mu - \mu^2 Tr(\mathbf{R}_v)} E\widetilde{\mathbf{w}_i}^H \mathbf{R}_v \left(\mu - \frac{1}{Tr(\mathbf{R}_v)}\right) \mathbf{s}_i \end{aligned} \tag{3.91}$$

Here the sparse-aware term affects the wights error directly, and we independence assumption cannot be taken.

From (3.90) and (3.91), it is clear that the LMS-sparse aware algorithm increases the EMSE in the non-zero elements and reduces it for zero-elements. When the number of non-zeros is very small (very sparse), the overall EMSE is significantly less than the LMS. The EMSE of the TD-ZA-LMS:

$$\begin{aligned} EMSE_{TD-ZA-LMS} = & EMSE_{TD-LMS} \\ & + \frac{1}{2-\mu Tr(\mathbf{R_V})} \Bigg[2\rho E\widetilde{\mathbf{w}_\infty}^H \\ & \times \left(\frac{1}{Tr(\mathbf{R}_v)} - \mu \right) sgn(\mathbf{w}_\infty) \\ & + \rho^2 \frac{E\|sgn(\mathbf{w}_\infty)\|^2_{\mathbf{R}_v \mathbf{\Lambda}^{-2}}}{Tr(\mathbf{R}_v)} \Bigg] \\ = & \zeta^{TD-LMS} + J(\infty) \end{aligned} \tag{3.92}$$

3.10 Universality of the Convex Filter

To evaluate the steady state performance of the proposed convex filter,Cross-Excess Mean Square Error (CMSE) exerted by the combination filter is studied.

The CMSE is defined as:

$$CMSE = E\{e_{a,ns}(i)e_{a,s}(i)\} \tag{3.93}$$

The energy conservation relation for the combination filter at the steady state:

$$E\{\frac{e_{a,ns}(i)e_{a,s}(i)}{\|\mathbf{v}_i\|_2^2}\} = E\{\frac{e_{p,ns}(i)e_{p,s}(i)}{\|\mathbf{v}_i\|_2^2}\} \tag{3.94}$$

Starting from equation (11), the CMSE is given by:

$$\begin{aligned} CMSE &= \frac{\sigma_n^2 \mu_s \mu_{ns} Tr(\mathbf{R}_v)}{\eta} \\ &+ \frac{1}{\eta} E\widetilde{\mathbf{w}}_{ns}\mathbf{R}_v\left(\frac{1}{Tr(\mathbf{R}_v)} - \mu_{ns}\right)\mathbf{s}_i \end{aligned} \tag{3.95}$$

where $\eta = \mu_s + \mu_{ns} - \mu_s\mu_{ns}Tr(\mathbf{R}_v)$. Comparing equation (19) with (13), we can see that the convex filter is already introducing less bias to the non-zero elements. To study the steady state behavior of the filter, we study the evolution of the combination factor $\lambda(i)$ as i tends to infinity. We do so by studying $a(i)$:

$$\begin{aligned} E[a(i+1)] &= E[a(i)] + \mu_a E[\lambda(i)[1-\lambda(i)]^2]\Delta J_2 \\ &- \mu_a E[\lambda(i)^2[1-\lambda(i)]]\Delta J_1 \end{aligned} \tag{3.96}$$

where:

$$\Delta J_1 = J_{ex,1}(\infty) - J_{ex,12}(\infty)$$
$$\Delta J_2 = J_{ex,2}(\infty) - J_{ex,12}(\infty)$$

For the case under consideration,, $J_{ex,2} = EMSE_{TD-LMS}$ while $J_{ex,1} = EMSE_{TD-ZA-LMS}$. The cross error of the convex in (3.95) represents $J_{ex,12}$. To study the universality of the proposed filter, three scenarios are considered. The first case when the systems under identification is completely dense, and as the number of active taps reduces from the dense setting, we move to the semi-sparse setting and the final case is only spare.

3.10.1 Non-Sparse System — $J_{ex,2} \leq J_{ex,12} \leq J_{ex,1}$:

For this case, and from equations (17) and (21), we can see that $J_{ex,12} \approx J_{ex,2}$ since ρ^2 is very small. equivalently, $\Delta J_2 \approx 0$ and $\Delta J_2 > 0$ or $J_{ex,1} > J_{ex,2}$, which is logical since for non sparse systems, TD-LMS is better than TD-ZA-LMS. equation (21) translates to:

$$E[a(i+1)] = E[a(i)] - \mu_a E[\lambda(i)^2[1-\lambda(i)]]\Delta J_1 \tag{3.97}$$

limiting $a(i)$ to be in the range $a^+ < a(i) < -a^+$ which corresponds to $\lambda^+ < \lambda(i) < \lambda^-$, then the term $f(\lambda(i)) = E[\lambda(i)^2[1-\lambda(i)]] \geq f(1-\lambda^+) = C > 0$, equation (22) becomes:

$$E[a(i+1)] \leq E[a(i)] - \mu_a C \Delta J_1 \tag{3.98}$$

(24) says that, the combination factor evolves to reach the limit of $-a^+$ at ∞. That means the combination chooses the second filer i.e, TD-LMS filter, since it performs better.

3.10.2 Semi-Sparse System:

In this case, both $\Delta J_1 > 0$ and $\Delta J_2 > 0$, solving (21) at $i = \infty$, the stationary point is given by:

$$E[\lambda(\infty)] = \frac{\Delta J_2}{\Delta J_1 + \Delta J_2} \tag{3.99}$$

This condition is optimal , since it leads to:

$$J_{ex,12} \leq min[J_{ex,1}, J_{ex,2}] \tag{3.100}$$

which means that the combination is performing better than both of them.

3.10.3 Sparse System — $J_{ex,1} \leq J_{ex,12} \leq J_{ex,2}$:

In this case, TD-ZA-LMS performs better, which leads to $\Delta J_1 \approx 0$ and $\Delta J_2 > 0$, then (22) summarize to:

$$\begin{aligned} E[a(i+1)] &= E[a(i)] + \mu_a E[\lambda(i)^2[1-\lambda(i)]]\Delta J_2 \\ &\geq \mu_a C \Delta J_2 \end{aligned} \tag{3.101}$$

which means that as $a(i)$ goes to a^+ (equivalently λ^+), selects the first filter i.e, the TD-ZA-LMS.Hence the combination filter, as expected, inherits the same universality properties as its time domain version. Next, we explore the performance of the filter when the input is highly correlated.

3.11 Computer Simulation

In this section the performance of the Transformed convex combination is investigated. Two experiments are performed and, in each, three stages are introduced to measure the tracking ability. The three phases are sparse, semi-sparse and dense.

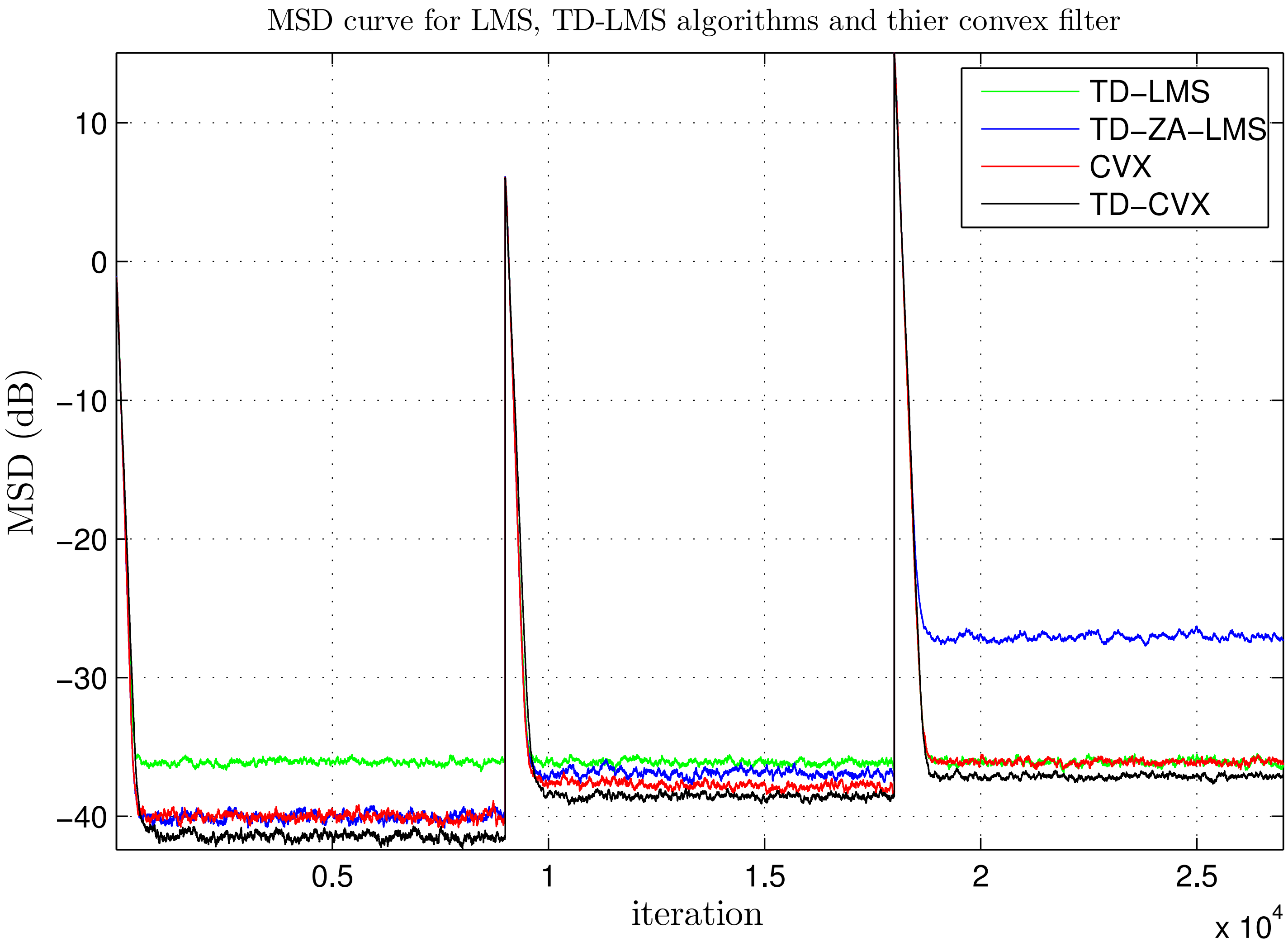

Figure 3.6: MSD curves for LMF, ZA-LMF algorithms and their convex combination. SNR level is 10 dB with uniform measurement noise.

In the first experiment, the input to the convex filter is white Gaussian. DFT is used as a unitary transfromation and the power normalization is estimated using exponential windowing described by:

$$\sigma_k^2(i+1) = \beta\sigma_k^2(i) + (1-\beta)\|v_k(i)\|^2 \tag{3.102}$$

where $\beta = 0.95$ is good choice for stationary inputs. The step sizes of the component algorithms equal 0.01, and $\mu_a = 1000, \mu_{td} = 100$ for the combination filters.The step sizes are chosen to insure that the algorithms reach the same Mean Square Deviation (MSD) floor, which allows us to compare the convergence speed more efficiently. Both filters has the same zero attraction power $\rho = 5 \times 10^{-5}$. The SNR is set to 30 dB. For first 1000 iterations, the system under identification is considered to be a sparse system with sparsity rate of $\frac{M_{active}}{M} = 1/16$, where M_{active} is the cardinality of active taps , and M is the number of taps of the filter. At iteration number 1000, the system is switched to semi sparse case with rate $\frac{5}{16}$, and after 2000 system is considered to be dense.. The same procedure of university testing is duplicated for the second experiment.

From Fig. 2, we see that both types of filters offer the same convergence, since the DFT and power normalization can slightly wighten the already White Gaussian input. The difference in convergence in Fig. 3 though is dramatic. The transformed convex filter offer much higher rate of convergence compared to its time domain version. It can also be noted that, the transformed convex grantees optimal universality,in the case of semi-sparse case. In which the EMSE is strictly less than the component filters.

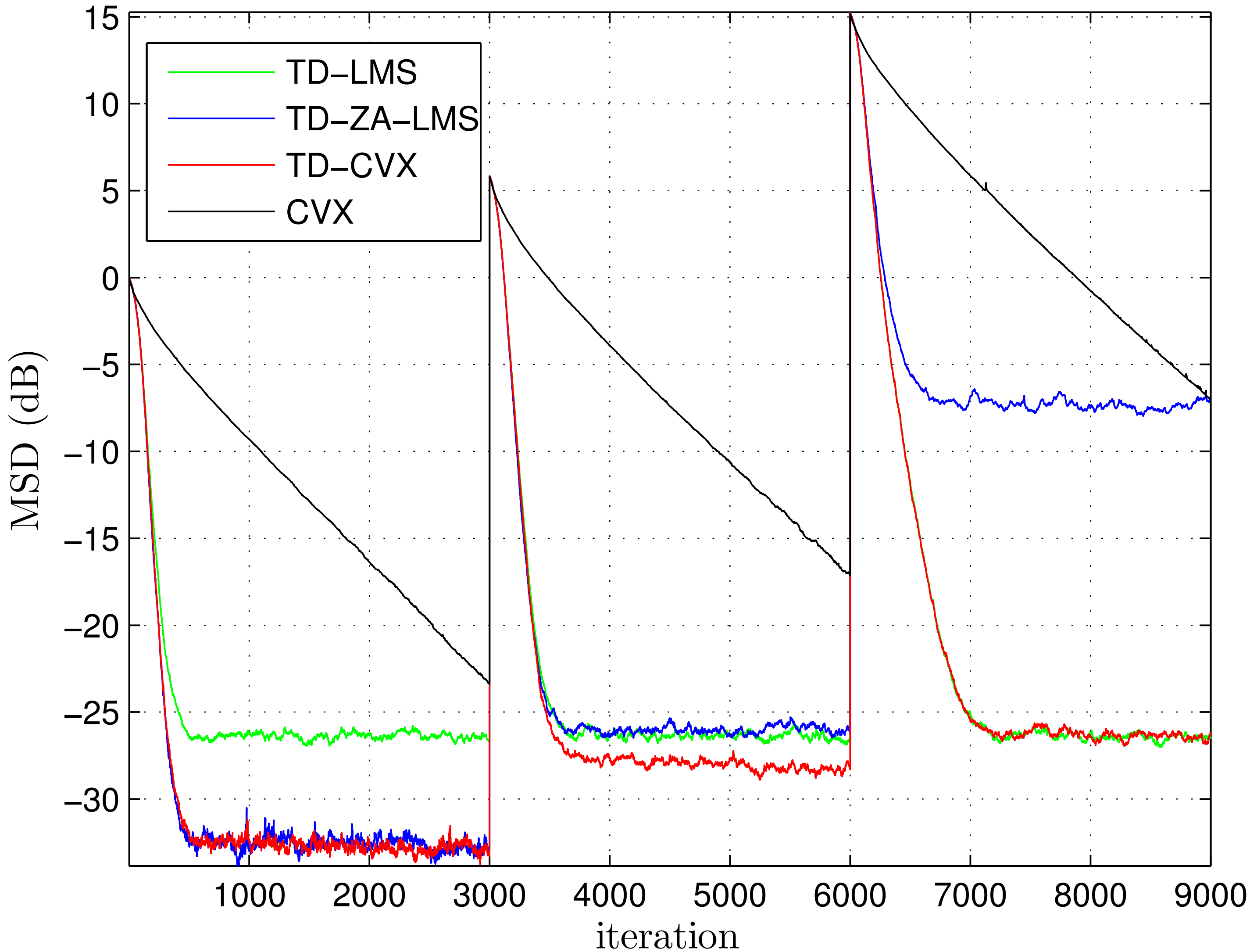

Figure 3.7: Combination factor evolution for the second experiment — Correlated Input

3.12 Conclusion

In this work, a convex combination filter for variable sparse system identification problem is proposed, under highly correlated environment. The proposed filter merge two algorithms, namely the transform domain LMS and its sparse aware version, TD-ZA-LMS. It is verified that, compared to the time domain convex proposed in [102], the transformed domain convex offers much higher speed of convergence.

CHAPTER 4

NON-NEGATIVE SPARSE SYSTEM IDENTIFICATION

4.1 Sparse Non-negative NLMF algorithms

4.1.1 Introduction

The non-negativity constraint is another structure can be employed to refine the optimum solution. The non-negative LMS algorithm introduced in [] solves employ this condition with the square error function of the LMS, and applying the KKT conditions to reach an equation of the form $f(x) = 0$, where it is solved using fixed point recursion to find the root which happens to be the optimum solution, since the $f(x)$ is strictly convex. HHLMS algorithm, though superior in Gaussian noise environments, its performance deteriorates in non-Gaussian environments. NN-LMF algorithm is proposed as a fit solution to these environments, because the non-negative algorithm will inherit the properties of the LMF — which serves lower EMSE compared to the LMS in these environments. In this section we propose the sparse NN-LMF algorithm, which employs in addition to the non-negativity constraints, the l_1 norm of the solution we seek.

4.1.2 Zero Attractor Non-Negative LMF Algorithm

The cost function we seek to minimize for the optimum solution $\mathbf{w}_o$ is given by:

$$\begin{aligned} \mathbf{w}_o = \underset{\mathbf{w}}{\operatorname{argmin}}\ [\mathbf{u}_i\mathbf{w}_{i-1} - d(i)]^4 + \lambda\|\mathbf{w}_{i-1}\|_1 = f(\mathbf{w}_{i-1}) \\ s.t w_o(i) > 0 \forall i \end{aligned} \tag{4.1}$$

Introducing the non-negative vectors $\mathbf{w}_i^+$ and $\mathbf{w}_i^-$, the l_1 regularized problem renders to [105]:

$$\begin{aligned} \tilde{\mathbf{w}}_o = \underset{\mathbf{w}}{\operatorname{argmin}}\ [\begin{bmatrix} \mathbf{u}_i \\ -\mathbf{u}_i \end{bmatrix}^T \tilde{\mathbf{w}}_{i-1} - d(i)]^4 + \lambda \mathbf{1}_{2M}\tilde{\mathbf{w}}_{i-1} \\ s.t \tilde{\mathbf{w}}_{i-1} \succeq \mathbf{0} \end{aligned} \tag{4.2}$$

Clearly, the problem in (1.118) is convex. and λ is the Lagrange multiplier, and the Karush-Kuhn-Tucker must be satisfied at the optimal solution $\tilde{\mathbf{w}}_o$. Let $\mathbf{Q}(\tilde{\mathbf{w}}_o, \lambda) = \mathbf{J}(\mathbf{w}) + \lambda^T \mathbf{w}$, Then the KKT conditions are given by:

$$\nabla_{\mathbf{w}} \mathbf{Q}(\mathbf{w}_o, \lambda_o) = 0 \tag{4.3}$$

$$w_i^o \lambda_i^o = 0 \tag{4.4}$$

And the two conditions can be concatenated in one, given by:

$$\mathbf{D}_{\mathbf{w}} \mathbf{w}_o [\nabla J(\mathbf{w}_o)] = \mathbf{\Phi}(\mathbf{w}_o) = 0 \tag{4.5}$$

which translates the problem from minimization of (1.118) to finding the root of (1.121). In order to solve for the root iteratively, we resort an interior point iterative method, namely, the fixed point iteration scheme. The fixed point scheme generally is used to solve problems in the form $f(x) = 0$, by a mediate function $g(x) = x$, and iterating using $x_{n+1} = g(x_n)$, hoping that the sequence $x_0, x_1, x_2, x_3...$ will converge to the solution x.
By employing the fixed point iteration (FPI) to the point in hand, and introducing μ_n to act as a step size to endow (FPI) with convergence control, the optimum solution $\mathbf{w}_o$ can be sought using:

$$\tilde{\mathbf{w}}_{n+1} = \tilde{\mathbf{w}}_n + \mu_n \mathbf{D}_{\tilde{w}} [\tilde{\mathbf{u}}_i e^3(i) - \mathbf{1}^T \lambda] \tag{4.6}$$

$$\tilde{\mathbf{w}}_{n+1} = (1 - \mu\lambda)\tilde{\mathbf{w}}_n + f(\tilde{\mathbf{w}}_{i-1}) \mathbf{D}_{\tilde{w}} \tilde{\mathbf{u}}_i e^3(i) \tag{4.7}$$

where $f(\tilde{\mathbf{w}}_{i-1}) = \frac{1}{4}$ is a positive function on $\tilde{\mathbf{w}}_{i-1}$. $\tilde{\mathbf{u}}_i \begin{bmatrix} \mathbf{u}_i & -\mathbf{u}_i \end{bmatrix}$. where:

$$e(i) = d(i) - \tilde{\mathbf{u}}_i \tilde{\mathbf{w}} = d(i) - \mathbf{u}_i \mathbf{w} \tag{4.8}$$

and t$\mathbf{w}_i$ is calculated by:

$$\mathbf{w}_i = \mathbf{w}_i^+ - \mathbf{w}_i^- \tag{4.9}$$

Note that because of the use of the positive variables difference method, and the fixed point procedure, the algorithm seems to be sparse agnostic. However, by comparing to the algorithm in (1.123) with the ZA-LMF algorithm:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mathbf{u}_i e^3(i) - \lambda sign(\mathbf{w}_n) \tag{4.10}$$

we see that, by introducing $\mathbf{D}_{\tilde{\mathbf{w}}}$ which prevents the algorithm for looking of negative values of the optimum solution, which renders the zero attractor term $-\lambda sign(\mathbf{w}_{i-1})$ into $-\lambda \mathbf{w}_{i-1}$, hence we can see that (1.123) is sparse aware, when the optimum sought is non-negative.

4.1.3 Mean Behavior of the l_1 NNLMF algorithm

We study now the mean behavior of the l_1 NNLMF. We assume that the measurement noise $z(n) = d(n) - \mathbf{u}_n\mathbf{w}_o$ is white Gaussian with variance σ_z^2. As for the input the input and desired signals are assumed to be stationary. We customize the analysis for the proposed algorithm with White Gaussian input with power σ_x^2. The error vector is defined as following:

$$\mathbf{c}(n) = \tilde{\mathbf{w}}(n) - \tilde{\mathbf{w}}_o \tag{4.11}$$

and the error vector evolution recurssion is then:

$$\mathbf{c}(n+1) = (1-\mu\lambda)\mathbf{c}(n) + \mu e^3(n)\mathbf{D}_{\tilde{x}}\tilde{\mathbf{w}}(n) + \mu\lambda\mathbf{w}_o \tag{4.12}$$

and by redefining the error in term of error miss-adjustment vector $\mathbf{c}(n)$, that is , $e(n) = z(n) - \tilde{\mathbf{w}}(n)\mathbf{c}(n)$ is rendered to(and applying the statistical expectation E.):

$$E\mathbf{c}(n+1) = E\mathbf{c}(n) + \mu\{-3z^2(n)\tilde{\mathbf{x}}(n)\tilde{\mathbf{w}}(n) + [\tilde{\mathbf{x}}(n)\tilde{\mathbf{w}}(n)]^3\}\mathbf{D}_{\tilde{x}}[\mathbf{c}(n) + \tilde{\mathbf{w}}_o] + \mu\lambda\mathbf{w}_o \tag{4.13}$$

At this stage we resort to the independence assumption in order to simplify the analysis. the i^{th} entry of the vector $\mathbf{c}(n)$ is then reads:

$$\begin{aligned} c_i(n+1) &= (1-\mu\lambda)c_i(n) - 3\mu\sigma_z^2 E[x_i(n)\tilde{\mathbf{x}}(n)]E[c_i(n)\tilde{\mathbf{c}}(n)] \\ &+ \mu E[x_i(n)\tilde{\mathbf{x}}(n)]^3 E[c_i(n)\tilde{\mathbf{c}}(n)]^4 \\ &- 3\mu\sigma_z^2 w_{o,i} E[x_i(n)\tilde{\mathbf{x}}(n)]E[c_i(n)] \\ &+ \mu w_{o,i} E[x_i(n)\tilde{\mathbf{x}}(n)]^3 E[c_i(n)\tilde{\mathbf{c}}(n)]^3 \end{aligned} \tag{4.14}$$

And by extending the independence assumption, and assuming the error taps are statistically independent, we have:

$$\begin{aligned} Ec_i(n+1) &= (1-\mu\lambda - 3\mu\sigma_z^2\sigma_x^2 w_o)Ec_i(n) - 3\mu\sigma_z^2\sigma_x^2 Ec_i^2(n) \\ &+ 3\mu w_o\sigma_x^4 Ec_i^3(n) + 3\mu\sigma_x^4 w_o Ec_i^4(n) + \mu\lambda w_o \end{aligned} \tag{4.15}$$

at the steady state, $c_i(n+1) = c_i(n)$ when $w_o > 0$, and the higher moments of error (third and fourth) are expected to be very small, with this in hand, (1.131) becomes:

$$E[c_i(\infty)^2] = \frac{\lambda w_o}{3\sigma_z^2\sigma_x^2} \tag{4.16}$$

In this case of $w_o = 0$,$E[c_i(\infty)] \neq 0$ is not valid since the error can only random walk around the solution from one direction only, hence the average location is above zero. Hence the steady state is then:

$$E[c_i(\infty)^2] = \frac{\lambda c_i(\infty)}{3\sigma_z^2\sigma_x^2} \tag{4.17}$$

and the total MSD is the sum of the incurred MSD for the non-zero and zero elements. As for the step size condition to insure the convergence, we have to insure $\mu \leq \frac{2}{3w_o\sigma_x^2\sigma_z^2}$. The overall seems not to depend on the step size μ, but note that $E[c_i(\infty)]$ is a function of the learning rate μ, as it can be seen from (1.131).

4.1.4 Simulation

To evaluate the performance of the algorithm, two scenarios are tested with system identification setting. The length of the unknown filter is assumed to be $M = 32$. In both of the scenarios, a uniform initialization vector is used, with $w(o)_i = \frac{1}{M}$. Firstly, simulating Low SNR ($= 5dB$) environment with both the input sequence and the measurement noise are assumed to be white Gaussian input. The performance of the algorithm compared to its sparse agnostic version is superior as depicted in figure (1.6), and offer nearly $10dB$ improvement in steady state excess MSD. And for the second scenario the measurement noise is assumed to be uniform, and the algorithm grantees even lower steady state. The convergence of both algorithms is slow, basically because of the diminishing step size required for the LMF algorithms generally, to reduce the probability of divergence.

4.1.5 Conclusion

In this work, a sparse aware non negative LMF algorithm is proposed. Where it is found that it offers much lower EMSE compared its agnostic version, In Gaussian and non Gaussian noise environments (where LMS version algorithms suffers performance).

4.2 Convex Combination of NNLMS and l_1 NNLMS for variable non-negative sparse system identification

4.2.1 Introduction

The problem of sparse system identification has been quite famous for now because it is recognized, in myriad of applications in physical phenomena, that the underlying system is sparse in it is nature. And with the fort of mathematical tools introduced by the LASSO techniques, the problem is tackled systematical rather than one by one cases. In addition to sparsity structure, that is, very few non-zero elements are active, other systems inherit more customized setting, where the non-zero are strictly non-negative — mainly, because of the physical

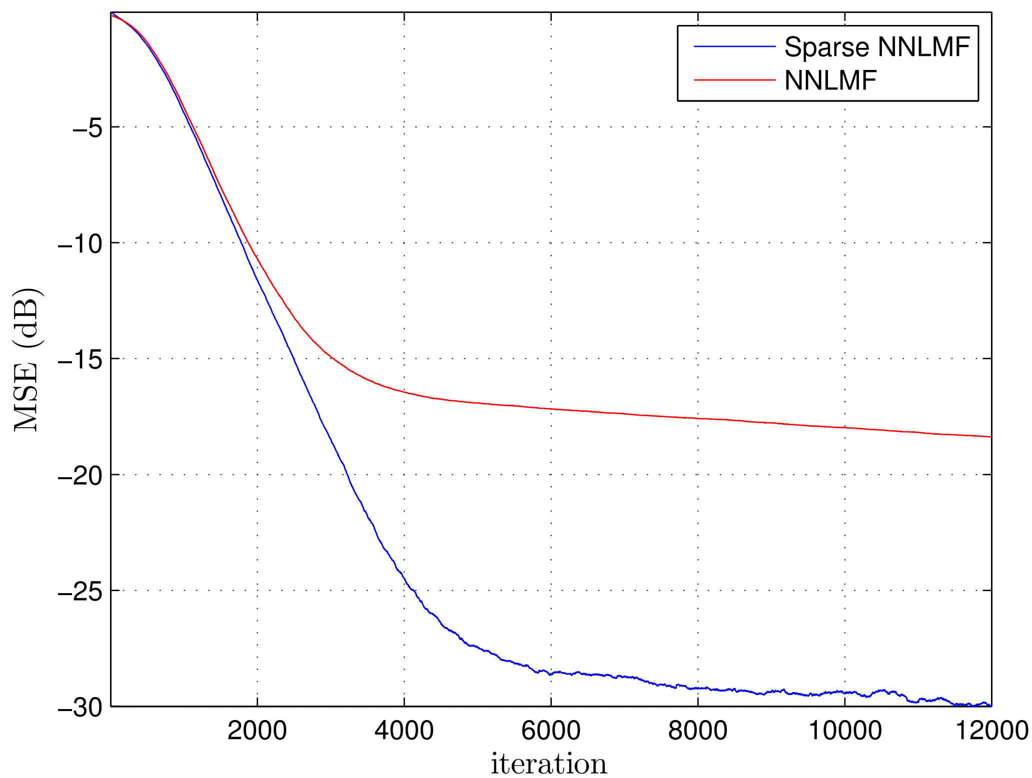

Figure 4.1: MSD curve for Sparse NNLMF, NNLMF algorithms,with White Gaussian Input and White Gaussian noise, and sparsity rate $\frac{1}{32}$ with random support. Simulation parameters are set to $\mu = 0.001$, $\lambda = 0.001$ and $M = 32$.

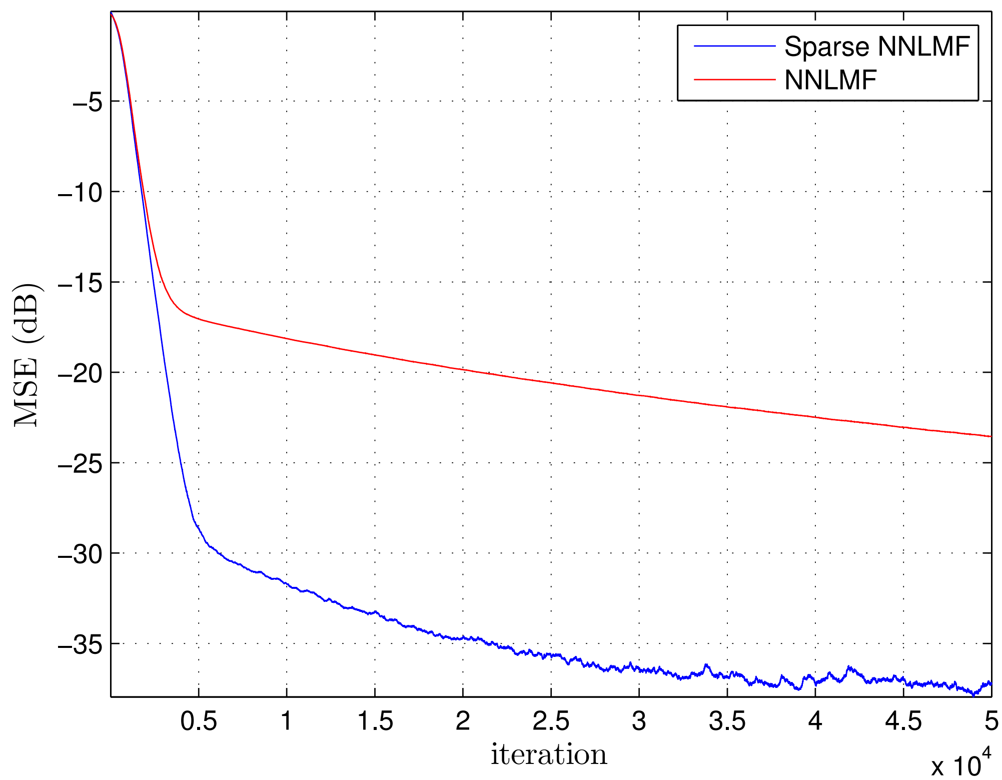

Figure 4.2: MSD curve for Sparse NNLMF, NNLMF algorithms,with White Gaussian Input and uniform noise, and sparsity rate $\frac{1}{32}$ with random support. Simulation parameters are set to $\mu = 0.001$, $\lambda = 0.001$ and $M = 32$.

meaning of the non-zero quantities. The l_1 NNLMS algorithm proved to recognize more these systems (non-negative sparse systems) efferently than their counterpart NNLMS. However, in many systems, the nature of the system can flow from sparse to dense system back and forth, an obstacle to select one algorithm over the other. However, the bewilderment of the choice can be remved completely by introducing a convex combination of the two algorithms, that is, merged filter of NNLMS and l_1 — NNLMS. The convex filter will automatically shifts from one filter to another depending on the system to be identified, in a system identification setting. Even more, when the system is semi-sparse, the convex offers a performance better than each of its components.
IN this work, we firstly introduce the convex combination of the two filters, and in the second part we study the universality of the convex filter: the EMSE of the convex compared to its components. In the third we propose the affine filter, which offers performance tantamount to the convex, with slightly less computational complexity reduction. Ee conclude the study by the computer simulations, to verify the theoretical analysis.

4.2.2 Problem Formulation

The first filter, F1, is the NNLMS algorithm, with recursion:

$$\mathbf{w}_1(n+1) = \mathbf{w}_1(n) + \mu e_1(n)\mathbf{D}_x(n)\mathbf{w}_1(n) \tag{4.18}$$

and for the l_1 non negative algorithm:

$$\begin{aligned}\alpha(n+1) &= \alpha(n)(1-\mu\lambda) + \mu e_2(n)\mathbf{D}_{\tilde{x}}\alpha_1(n) & (4.19)\\ \mathbf{w}_2(n) &= \alpha^+(n) - \alpha^-(n) & (4.20)\end{aligned}$$

where λ is the zero attractor power. clearly, $e_1(n) = d(n) - \mathbf{x}_n\mathbf{w}_1$ and $e_2(n) = d(n) - \tilde{\mathbf{x}}_n\alpha$. To study the universality of the convex filter, three parameters are sought, $J_{ex,1}(\infty) = E|e_{a,1}(\infty)|$, $J_{ex,2}(\infty) = E|e_{a,2}(\infty)|$ and $J_{ex,12}(\infty) = E|e_{a,1}(\infty)e_{a,2}(\infty)|$.

4.3 Computer Simulation

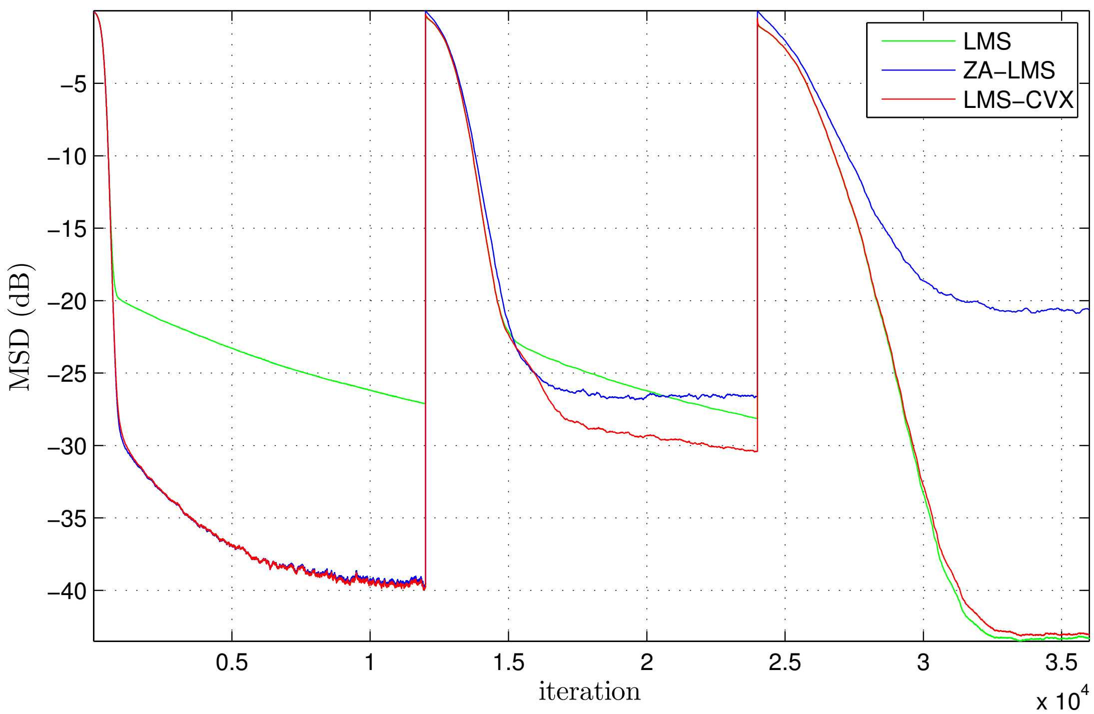

Figure 4.3: MSD curve for Sparse NNLMS, NNLMS and their convex filter,with White Gaussian Input

CHAPTER 5

CONCLUSION AND FUTURE WORK

The field of adaptive filters has been active for quite a while because of the simplicity of these systems . Recently, the advent of the compressive sensing theory has reinvented the whole systems to be sparse aware. In this work, the main theme is the sparse system identification problem where we tried to find solutions under special environments, namely, the non-Gaussian with correlated input. Moreover, we dealt with the problem of variable sparsity of the system under consideration, that is, the number of active coefficients is allowed to change, in addition to the support (locations). Endowing these algorithms in the communication systems for example, is expected to improve the Symbol error rate, because of good reconstruction of the communication channel for example.
As we have mentioned earlier, the sparsity is a very strong structure that is always beneficial to explore. Another equipotent structure is the block sparsity. Endowing block sparse aware algorithms with ability to recognize the variable block structure is mandatory in non-stationary environments. To do so, we will study the convex combination of block sparse aware algorithms, under Gaussian conditions. We resort ourself to the Gaussian conditions here, because of simplicity of analysis that is reflected in insightful results, that helps designer to efferently cook-up their algorithms. Moreover, the Gaussian assumption is quite reasonable because of central limit theorem.

REFERENCES

[1] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1078–1090, 2006.

[2] A. H. Sayed, *Adaptive Filters*. John Wiley & Sons, 2011.

[3] ——, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

[4] B. WIDROW, M. E. HOFF *et al.*, "Adaptive switching circuits." 1960.

[5] B. Widrow, J. McCool, and M. Ball, "The complex lms algorithm," in *IEEE Proceedings*, vol. 63, 1975, p. 719.

[6] S. Narayan, A. M. Peterson, and M. J. Narasimha, "Transform domain lms algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 3, pp. 609–615, 1983.

[7] A. H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003.

[8] A. Zerguine, "Convergence behavior of the normalized least mean fourth algorithm," *... of the Thirty-Fourth Asilomar Conference on*, pp. 275–278, 2000.

[9] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*.

[10] H. C. Andrews, "Multidimensional Rotations in Feature Selection," *IEEE Transactions on Computers*, vol. C-20, no. 9, pp. 1045–1051, Sep. 1971.

[11] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.

[12] K. S. Shanmugam, "Comments on "Discrete Cosine Transform"," *IEEE Transactions on Computers*, vol. C-24, no. 7, pp. 759–759, Jul. 1975.

[13] R. J. Clarke, "Performance of Karhunenâ€"LoeÌ€ve and discrete cosine transforms for data having widely varying values of intersample correlation coefficient," *Electronics Letters*, vol. 19, no. 7, p. 251, 1983.

[14] M. Hamidi and J. Pearl, "Comparison of the cosine and Fourier transforms of Markov-1 signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 5, pp. 428–429, Oct. 1976.

[15] R. J. Clarke, "Relation between the Karhunenâ€"Loeve and sine transforms," *Electronics Letters*, vol. 20, no. 1, p. 12, 1984.

[16] A. Jain, "A Fast Karhunen-Loeve Transform for a Class of Random Processes," *IEEE Transactions on Communications*, vol. 24, no. 9, pp. 1023–1029, Sep. 1976.

[17] R. R. Gallagher, "A new transform with symmetrical coding performance for Markov (1) signals," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2195–2198, 1995.

[18] S. S. Haykin, *Adaptive Filter Theory.*

[19] S. Narayan, "Transform domain LMS algorithm," *Acoustics, Speech and ...*, no. 3, pp. 609–615, 1983.

[20] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[21] C. L. JAE and K. U. , CHONG, "Performance of transform-domain LMS adaptive digital filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 3, pp. 499–510, Jun. 1986.

[22] G. Murmu and R. Nath, "Convergence Performance Comparison of Transform Domain LMS Adaptive Filters for Correlated Signal," *2011 International Conference on Devices and Communications (ICDeCom)*, vol. 1, no. 2, pp. 1–5, Feb. 2011.

[23] S. Zhao, Z. Man, S. Khoo, and H. Wu, "Stability and convergence analysis of transform-domain LMS adaptive filters with second-order autoregressive process," *Signal Processing, IEEE ...*, vol. 57, no. 1, pp. 119–130, Jan. 2009.

[24] M. D. Courville and P. Duhamel, "Adaptive filtering in subbands using a weighted criterion," *Signal Processing, IEEE ...*, vol. 46, no. 9, pp. 2359–2371, 1998.

[25] E. Lobato, O. Tobias, and R. Seara, "Stochastic modeling of the transform-domain ϵlms algorithm for correlated gaussian data," *Telecommunications ...*, pp. 912–917, 2006.

[26] ——, "Stochastic modeling of the transform-domain algorithm," *Signal Processing, IEEE ...*, vol. 56, no. 5, pp. 1840–1852, 2008.

[27] F. Boroujeny and S. Gazor, “Performance Analysis of Transform Domain Normalized LMS Algorithm,” *IEEE*, pp. 2133–2136, 1991.

[28] S. C. Chan and Y. Zhou, “On the convergence analysis of the transform domain normalized LMS and related M-estimate algorithms,” *APCCAS 2008 - 2008 IEEE Asia Pacific Conference on Circuits and Systems*, pp. 205–208, Nov. 2008.

[29] Yuexian Zou, Shing-Chow Chan, and Tung-Sang Ng, “Least mean M-estimate algorithms for robust adaptive filtering in impulse noise,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 12, pp. 1564–1569, 2000.

[30] T. Aboulnasr and K. Mayyas, “Selective coefficient update of gradient-based adaptive algorithms,” *Acoustics, Speech, and Signal . . .*, vol. 00, pp. 1929–1932, 1997.

[31] S. Attallah and S. W. Liaw, “Analysis of DCTLMS algorithm with a selective coefficient updating,” *Circuits and Systems II: Analog and Digital . . .*, vol. 48, no. 6, pp. 628–632, 2001.

[32] R. C. Bilcu, P. Kuosmanen, and K. Egiazarian, “A transform domain LMS adaptive filter with variable step-size,” *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 51–53, Feb. 2002.

[33] ——, “A new variable step LMS algorithm for transform domain,” *ICECS 2001. 8th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.01EX483)*, no. 2, pp. 1161–1164.

[34] S. C. Chan, Y. J. Chu, and Z. G. Zhang, “A new regularized transform-domain NLMS adaptive filtering algorithm,” *2010 IEEE Asia Pacific Conference on Circuits and Systems*, pp. 696–699, Dec. 2010.

[35] ——, “A New Variable Regularized Transform Domain NLMS Adaptive Filtering Algorithmâ€”Acoustic Applications and Performance Analysis,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 4, pp. 868–878, Apr. 2013.

[36] A. Tandon, M. N. S. Swamy, and M. O. Ahmad, “L-infinity-NORM BASED PARTIAL-UPDATE ADAPTIVE FILTERING ALGORITHM FOR ECHO CANCELLATION,” *Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4, pp. 353–356, 2005.

[37] A. W. H. Khong and P. a. Naylor, “Efficient use of sparse adaptive filters,” *Conference Record - Asilomar Conference on Signals, Systems and Computers*, pp. 1375–1379, 2006.

[38] H. Deng and M. Doroslovački, "Improving convergence of the PNLMS algorithm for sparse impulse response identification," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181–184, 2005.

[39] J. Arenas-Garcia and A. R. Figueiras-Vidal, "Adaptive combination of proportionate filters for sparse echo cancellation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1087–1098, 2009.

[40] C. Paleologu, J. Benesty, and S. CiochinÇŽ, "An improved proportionate NLMS algorithm based on the l0 norm," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 309–312, 2010.

[41] B. K. Das, L. a. Azpicueta-Ruiz, M. Chakraborty, and J. Arenas-Garcia, "A comparative study of two popular families of sparsity-aware adaptive filters," *4th International Workshop on Cognitive Information Processing - Proceedings of CIP 2014*, 2014.

[42] Y. Chen, Y. Gu, and A. O. H. Iii, "SPARSE LMS FOR SYSTEM IDENTIFICATION Department of EECS , University of Michigan , Ann Arbor , MI 48109-2122 , USA," no. 3, pp. 3125–3128, 2009.

[43] J. Yang and G. E. Sobelman, "Sparse LMS with segment zero attractors for adaptive estimation of sparse signals," *IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings, APCCAS*, pp. 422–425, 2010.

[44] K. Shi and P. Shi, "Convergence analysis of sparse LMS algorithms with l1-norm penalty based on white input signal," *Signal Processing*, vol. 90, no. 12, pp. 3289–3293, 2010.

[45] Y. Gu, J. Jin, and S. Mei, "Norm Constraint LMS Algorithm for Sparse System Identification," vol. 16, no. 9, pp. 774–777, 2009.

[46] P. S. Bradley and O. L. Mangasarjan, "Feature Selection via Concave Minimization and Support Vector Machines," pp. 82–90, 1998.

[47] G. Su, J. Jin, Y. Gu, and J. Wang, "Performance analysis of l0 norm constraint least mean square algorithm," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2223–2235, 2012.

[48] O. Taheri and S. a. Vorobyov, "Sparse channel estimation with lp-norm and reweighted l1-norm penalized least mean squares," *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2864–2867, 2011.

[49] E. M. S. Algorithm and F. I. R. Pr-qmp, "Signal/'Image and," no. 4, pp. 1–4.

[50] Y. Kopsinis, K. Slavakis, S. Theodoridis, and S. McLaughlin, "Thresholding-based online algorithms of complexity comparable to sparse LMS methods," *Proceedings - IEEE International Symposium on Circuits and Systems*, no. 2, pp. 513–516, 2013.

[51] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Signal Reconstruction Using Projections," vol. 59, no. 3, pp. 936–952, 2011.

[52] J. Jin, Y. Gu, and S. Mei, "A Stochastic Gradient Approach on Compressive Sensing Signal Reconstruction Based on Adaptive Filtering Framework," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 409–420, 2010.

[53] K. Shi and X. Ma, "Transform domain LMS algorithms for sparse system identification," *...Speech and Signal Processing (ICASSP), 2010 ...*, pp. 3714–3717, 2010.

[54] M. Aliyu, M. Alkasim, and M. Salman, "A zero-attracting Variable Step-Size LMS Algorithm for Sparse System Identification," *Signal, Image and Video Processing*, pp. 2–5, 2012.

[55] M. O. Bin Saeed and A. Zerguine, "A variable step size strategy for sparse system identification," *2013 10th International Multi-Conference on Systems, Signals and Devices, SSD 2013*, no. 5, pp. 18–21, 2013.

[56] R. C. De Lamare and R. Sampaio-Neto, "Sparsity-aware adaptive algorithms based on alternating optimization and shrinkage," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 225–229, 2014.

[57] Y. Chen, Y. Gu, and A. O. Hero, "Regularized Least-Mean-Square Algorithms," p. 9, 2010.

[58] B. K. Das, M. Chakraborty, and S. Banerjee, "Adaptive identification of sparse systems with variable sparsity," *Proceedings - IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 1267–1270, 2011.

[59] R. C. de Lamare and V. H. Nascimento, "Sparsity-aware affine projection adaptive algorithms for system identification," *Sensor Signal Processing for Defence (SSPD 2011)*, pp. 7–7, 2011.

[60] M. V. S. Lima, I. Sobron, W. a. Martins, and P. S. R. Diniz, "Stability and MSE analyses of affine projection algorithms for sparse system identification," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 6399–6403, 2014.

[61] M. Yukawa, "Sparsity-Aware Adaptive Filters Based on p -Norm Inspired Soft-Thresholding Technique," no. 2.

[62] Y. Li, W. Li, W. Yu, J. Wan, and Z. Li, "Sparse Adaptive Channel Estimation Based on l p -Norm-Penalized Affine Projection Algorithm," vol. 2014, pp. 1–15, 2014.

[63] S. Nunoo, U. A. K. Chude-okonkwo, R. Ngah, and Y. K. Zahedi, "Variable Step-Size l 0 -Norm NLMS Algorithm for Sparse Channel Estimation," pp. 88–91, 2014.

[64] G. Gui, S. Kumagai, A. Mehbodniya, and F. Adachi, "Variable is good: Adaptive sparse channel estimation using VSS-ZA-NLMS algorithm," *2013 International Conference on Wireless Communications and Signal Processing, WCSP 2013*, 2013.

[65] S. Zhang and J. Zhang, "Transient analysis of zero attracting NLMS algorithm without Gaussian inputs assumption," *Signal Processing*, vol. 97, pp. 100–109, 2014.

[66] L. Weruaga, S. Member, S. Jimaa, and S. Member, "Exact NLMS Algorithm with -Norm Constraint," vol. 22, no. 3, pp. 366–370, 2015.

[67] C. S. Lin and I. T. Lee, "Applying multiple description coding to enhance the streaming scalability on CDN-P2P network," *International Journal of Communication Systems*, vol. 23, no. April 2013, pp. 553–568, 2010.

[68] G. Gui and F. Adachi, "Sparse least mean forth filter with zero-attracting \$l_1\$ norm constraint," *Submitted for IEICE Electronics Express*, pp. 1–6, 2013.

[69] G. Gui, "Adaptive Spare Channel Estimation Using Re- Weighted Zero Attracting Normaalized Least Mean Fourth."

[70] G. Gui, L. Xu, X. mei Zhu, and Z. xin Chen, "Novel Realization of Adaptive Sparse Sensing with Sparse Least Mean Fourth Algorithm," pp. 3–7, 2014.

[71] G. Gui, L. Xu, and F. Adachi, "RZA-NLMF algorithm-based adaptive sparse sensing for realizing compressive sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, p. 125, 2014.

[72] G. Gui and F. Adachi, "Adaptive sparse channel estimation for time-variant MIMO-OFDM systems," *2013 9th International Wireless Communications and Mobile Computing Conference, IWCMC 2013*, pp. 878–883, 2013.

[73] B. K. Das and M. Chakraborty, "Sparse Distributed Learning via Heterogeneous Diffusion Adaptive Networks," pp. 0–3.

[74] W. Shi, Q. Ling, and G. Wu, "Sparsity-enhanced linear time-invariant MIMO system identification," *Proceedings of the 2011 Chinese Control and Decision Conference, CCDC 2011*, pp. 2026–2029, 2011.

[75] J. K. Pant, W. sheng Lu, L. Fellow, A. Antoniou, and L. Fellow, "New Improved Algorithms for Compressive Sensing Based on p Norm," vol. 61, no. 3, pp. 198–202, 2014.

[76] B. Dumitrescu, A. Onose, P. Helin, and I. TÇŽbuÅŸ, "Greedy sparse RLS," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2194–2207, 2012.

[77] Y. Chen and A. O. Hero, "Recursive $l_{1,\infty}$ Group lasso," vol. 60, no. 8, p. 8, 2011.

[78] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online Adaptive Estimation of Sparse Signals: Where RLS Meets the l1 -Norm," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, p. 3436, 2010.

[79] G. Gui and F. Adachi, "Sparse least mean fourth algorithm for adaptive channel estimation in low signal-to-noise ratio region," *International Journal of Communication Systems*, vol. 27, no. 11, pp. 3147–3157, 2014.

[80] S. Zhao, Z. Man, S. Khoo, and H. R. Wu, "Stability and convergence analysis of transform-domain lms adaptive filters with second-order autoregressive process," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 119–130, 2009.

[81] C. Samson and V. Reddy, "Fixed point error analysis of the normalized ladder algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 5, pp. 1177–1191, 1983.

[82] E. Walach and B. Widrow, "The least mean fourth (lmf) adaptive algorithm and its family," *Information Theory, IEEE Transactions on*, vol. 30, no. 2, pp. 275–283, 1984.

[83] J. C. Principe, D. Xu, and J. Fisher, "Information theoretic learning," *Unsupervised adaptive filtering*, vol. 1, pp. 265–319, 2000.

[84] E. Hänsler, "The hands-free telephone problem-an annotated bibliography," *Signal Processing*, vol. 27, no. 3, pp. 259–271, 1992.

[85] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on.* IEEE, 2008, pp. 5–10.

[86] V. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *OCEANS'95. MTS/IEEE. Challenges of Our Changing Global Environment. Conference Proceedings.*, vol. 3. IEEE, 1995, pp. 1417–1422.

[87] Y. Gu, K. Tang, H. Cui, and W. Du, "Convergence analysis of a deficient-length lms filter and optimal-length sequence to model exponential decay impulse response," *Signal Processing Letters, IEEE*, vol. 10, no. 1, pp. 4–7, 2003.

[88] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *Speech and Audio Processing, IEEE Transactions on*, vol. 8, no. 5, pp. 508–518, 2000.

[89] H. Deng and M. Doroslovacki, "Improving convergence of the pnlms algorithm for sparse impulse response identification," *Signal Processing Letters, IEEE*, vol. 12, no. 3, pp. 181–184, 2005.

[90] J. Benesty and S. L. Gay, "An improved pnlms algorithm," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2. IEEE, 2002, pp. II–1881.

[91] M. O. Sayin, Y. Yilmaz, A. Demir, and S. S. Kozat, "The krylov-proportionate normalized least mean fourth approach: Formulation and performance analysis," *Signal Processing*, vol. 109, no. 0, pp. 1–13, 2015.

[92] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[93] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[94] Y. Chen, Y. Gu, and A. O. Hero, "Sparse lms for system identification," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on.* IEEE, 2009, pp. 3125–3128.

[95] M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 3137–3149, 2008.

[96] N. J. Bershad and J. C. Bermudez, "Mean-square stability of the normalized least-mean fourth algorithm for white gaussian inputs," *Digital Signal Processing*, vol. 21, no. 6, pp. 694–700, 2011.

[97] E. Eweda and A. Zerguine, "A normalized least mean fourth algorithm with improved stability," in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers.* IEEE, Nov. 2010, pp. 1002–1005.

[98] C. Ye, G. Gui, L. Xu, and N. Shimoi, "Improved adaptive sparse channel estimation using re-weighted l1-norm normalized least mean fourth algorithm," *arXiv preprint arXiv:1501.07648*, 2015.

[99] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan (Part I: Communications)*, vol. 67, no. 5, pp. 19–27, 1984.

[100] F. Beaufays, "Transform-domain adaptive filters: an analytical approach," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 422–431, 1995.

[101] K. Shi and X. Ma, "Transform domain lms algorithms for sparse system identification," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3714–3717.

[102] B. K. Das and M. Chakraborty, "Sparse adaptive filtering by an adaptive convex combination of the LMS and the ZA-LMS Algorithms," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 5, pp. 1499–1507, 2014.

[103] L. A. Azpicueta-Ruiz, A. R. Figuieras-Vidal, and J. Arenas-Garcia, "Acoustic echo cancellation in discrete fourier transform domain based on adaptive combination of adaptive filters," in *Proceedings of Meetings on Acoustics*, vol. 19, no. 1. Acoustical Society of America, 2013, p. 055043.

[104] J. Arenas-Garcia, V. Gomez-Verdejo, and A. R. Figueiras-Vidal, "New algorithms for improved adaptive convex combination of lms transversal filters," *Instrumentation and Measurement, IEEE Transactions on*, vol. 54, no. 6, pp. 2239–2249, 2005.

[105] M. Schmidt, "Least squares optimization with l1-norm regularization," *CS542B Project Report*, 2005.

Vitae

- Name: Murwan Mohamed Elmahdi Bashir
- Nationality: Sudanese
- Date of Birth: 9^{th} November 1988
- Email: *murwanbashir@gmail.com*
- Permenant Address: Khartoum Sudan