EFFECT OF INPUT CORRELATION ON (NORMALIZED) ADAPTIVE FILTERS

BY

Khaled Abdulaziz Goblan Al Hujaili

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

December 2013

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

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **KHALED ABDULAZIZ GOBLAN AL-HUJAILI** 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 DEPARTMENT**.

Thesis Committee

Dr. Tareq Y. Al-Naffouri (Adviser)

for

Dr. Muhammad Moinuddin (Co-adviser)

Dr. Ali A. Al-Shaikhi (Member)

Azzedine Zerguine

Dr. Azzedine Zerguine (Member)

Dr. Samir Al-Ghadban (Member)

Dr. Ali A. Al-Shaikhi
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

30|12|13
Date

DEANSHIP OF GRADUATE STUDIES
عمادة الدراسات العليا

إلى من أضاؤوا لي الطريق

والدي العزيزين

و إلى من ساندوني و تنازلوا عن حقوقهم

زوجتي و أبنائي رغد و رنيم و عبدالعزيز

و إلى إخوتي الأعزاء

To My Parents

To My Wife, My Daughters Raghad and Raneem and My son Abdulaziz

To My Brothers and Sisters

ACKNOWLEDGMENTS

First and foremost, I reserve all thanks and appreciation to Allah the Almighty for His countless blessings. Without the of guidance and help of Allah, this thesis would never have seen light.

I am grateful to King Fahd University of Petroleum and Minerals for providing me the opportunity to study and carry out my research at such a prestigious university under the guidance of highly qualified faculty.

I would like to express my gratitude and appreciation to my Advisor Dr. Tareq Al-Naffouri for his guidance and support throughout my work. I was fortunate to take two courses with him and learn from his immense knowledge. I would also like to express my deep gratitude to my co-advisor Dr. Mohammad Moinuddin for his support and guidance. I would also like to thank him for hosting me in his office in King Abdulaziz University in Jeddah.

I am also grateful to my thesis committee members Dr. Ali Al-Shaikhi, Dr. Azzedine Zerguine and Dr. Samir Al-Ghadban for their detailed review and excellent advices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

$[\cdot]^*$	Complex conjugate for scalers
$[\cdot]^H$	Hermitian transposition for matrices
$[\cdot]^T$	Matrix transpose
$\boldsymbol{\lambda}$	Vector of the eigenvalues of the matrix $R_{\boldsymbol{u}}$
$\boldsymbol{d}(i)$	The desired signal, zero-mean scaler valued random variable
$\boldsymbol{u}(i)$	The input regressor, zero-mean row vector random variable
$\boldsymbol{v}(i)$	Additive noise, zero-mean scaler valued random variable
$\boldsymbol{w}_i$	A guess of w^o at iteration i
Λ	Diagonal matrix with the eigenvalues of $R_{\boldsymbol{u}}$ as entries
λ_k	The k^{th} eigenvalue of the matrix $R_{\boldsymbol{u}}$
$\mathbb{R}$	Set of real numbers
μ	The step size along the p direction
∇	Gradient operator

$\sigma_{\boldsymbol{d}}^2$ Variance of $\boldsymbol{d}$

$diag\{\boldsymbol{A}\}$ Column vector with the diagonal entries of matrix $\boldsymbol{A}$

$diag\{\boldsymbol{a}\}$ Diagonal matrix with entries read from the column $\boldsymbol{a}$

$e(i)$ Estimation error

$E[\cdot]$ Expectation operator

$e_a(i)$ A priori estimation error

$e_p(i)$ A posteriori estimation error

$f[e(i)]$ A general function of the estimation error

$J(w)$ Cost function

J_{min} Minimum mean square error

M Length of the filter

p The update direction vector

$R_{\boldsymbol{du}}$ Cross covariance vector of $\{\boldsymbol{d}, \boldsymbol{u}\}$

$R_{\boldsymbol{u}}$ The covariance matrix of the input regressor

$Re[\cdot]$ Real part of $[\cdot]$

$Tr(\cdot)$ Trace operator

w^o Optimal weight vector

THESIS ABSTRACT

NAME:	Khaled Abdulaziz Goblan Al-Hujaili
TITLE OF STUDY:	Effect of input correlation on (normalized) adaptive filters
MAJOR FIELD:	Electrical Engineering Department
DATE OF DEGREE:	December 2013

The aim of this work is to give more insight about the performance of Adaptive Filters. Studying this performance will help researchers to understand the influences that will affect this. This work can be divided into two parts as follows : In the first part, we used the majorization theory as a mathematical tool to study the effect of the input correlation scenarios on the performance of adaptive filters. With this, we provide a mechanism to assess their performance. Also, with majorization theory, vector comparison is carried out and their order is preserved through Schur's functions. Each correlation scenario can be totally described by the eigenvalues of the covariance matrix $\boldsymbol{R_u}$. *Thus, a comparison between these scenarios can be done and a comparison between the responses of adaptive filters to these scenarios can also be done. In the second part, a new approach for studying the steady state performance of the Recursive Least Square (RLS) adaptive*

filter for a circularly correlated Gaussian input. The mean-square analysis of the RLS filter in the steady state relies on the moment of the random variable $\|\mathbf{u}_i\|^2_{\mathbf{P_i}}$*, where* $\mathbf{u}_i$ *is input to the RLS filter and* $\mathbf{P}_i$ *is the estimate of the inverse of input covariance matrix. Earlier approaches evaluate this moment by assuming that* $\mathbf{u}_i$ *and* $\mathbf{P}_i$ *are independent which could result in negative value of the steady-state Excess Mean Square Error (EMSE). In this work, we avoid this assumption and derive a closed from expression for this moment. This derivation is based on finding the cumulative distribution function (CDF) of the random variable of the form* $\frac{1}{\gamma+||\mathbf{u}||^2_{\mathbf{D}}}$*, where* $\mathbf{u}$ *is correlated circular Gaussian input and* $\mathbf{D}$ *is a diagonal matrix. As a result, we obtain more accurate estimate of the EMSE of the RLS algorithm.*

الخلاصة

الإسم : خالد عبدالعزيز قبلان الحجيلي

العنوان : تأثير الارتباط بين عناصر اشارة الدخل على المرشحات المتكيفة

الدرجة : ماجستير علوم

التخصص : هندسة كهربائية

تاريخ التخرج : ديسمبر ٢٠١٣

الهدف من هذه الرسالة هو دراسة اداء المرشحات المتكيفة. لأن دراسة هذا الأداء يساعد الباحثين في تحسين هذه المرشحات و معرفة المؤثرات التي تؤدي الى تحسين اداءها و ايضاً تمكن من المقارنة بين مختلف انواع هذه المرشحات. يمكن تقسيم العمل في هذه الرسالة لى قسمين هما :

الأول : هو دراسة تأثيير التغير في صفات التداخل او الأرتباط بين عناصر الإشارة الداخلة و اداء المرشحات المتكيفة.

و لإنجاز هذا الهدف فلابد من توفر معادلات او كميات رياضية تصف اداء المرشحات المتكيفة سواء في المرحلة العابرة او في مرحلة الإستقرار و ايضاً توفر اداة رياضية تمكننا من الربط بين اي تغيير يحصل في مواصفات الارتباط في اشارة الدخل و اداء المرشحات المتكيفة. بإستخدام هذه الأداة الرياضية نستطيع الترتيب بين المتجهات التي تمثل ($eigenvalues$) لمصوفة الأرتباط و المقارنة بينها ايضأ و الفاظ على هذا الترتيب من خلال الدوال المتقعرة. تم انجاز هذا الهدف لأكثر من فلتر و مازالت الدراسة مستمرة للإستفادة اكثر من هذه الطريقة.

و القسم الثاني من هذه الرسالة كان عن اداء المرشح ($Recursive\ Least\ Squares\ RLS$) في مرحلة الأستقرار حيث اقتصرت اغلب الدراسات السابقة على دراسة مصفوفة الخطأ فقط و بشكل تقريبي. و لكن في هذا العمل تمت دراسة هذا الاداء بشكل ادق و ذلك

عن طريق دراسة العلاقات التي تحتوي على هذه المصفوفة. و اخيرا، فإن النتائج المستخرجة من هذا العمل افضل من النتائج المستخرجة من الدراسات السابقة و هذا يظهر في النتائج المستخرجة من المحاكاة بإستخدام برنامج ($MATLAB$) .

CHAPTER 1

INTRODUCTION

1.1 Adaptive Filters

Adaptive filters are playing an important role in modern communication systems. The adaptive filters comprise an important part in statistical signal processing. When the signals come from an unknown statistics of an environment, the use of the adaptive filters offers a good solution to this problem. Adaptive filters can be used to perform many tasks inside the telecommunication system such that equalization [1, 2], noise cancellation [3] and system identification [1] .

The most widely used adaptive filter algorithm is the Least Mean Squares (LMS) which is the stochastic approximation of the Steepest Descent (SD) algorithm. The SD algorithm provides a solution to the Wiener criteria by minimizing the mean square value of error using equation (1.1)

$$\min_{\boldsymbol{w}} E|d - \boldsymbol{u}\boldsymbol{w}|^2 \tag{1.1}$$

The target from the estimation problem in (1.1) is to find the column weight vector $\boldsymbol{w}$ that will make the quantity $\boldsymbol{uw}$ is the best estimate of the desired signal d in the linear least squares sense. Where $\boldsymbol{u}$ is a row vector. The LMS has low computational complexity with order of the filter length M, where the number of operations per iterations relatively low, but on the other hand it has a slow convergence, especially on highly correlated signals. Overcoming this problem can be done by using the Normalized LMS algorithm (NLMS) [4]. In this algorithm the input signal is normalized by its power.

There are many algorithms belong to the LMS family such as Sign-error LMS [5], in which the error signal is replaced by its signed version and Leaky LMS [6].

Another approach to improve the performance of the LMS algorithm is by using a time varying step size [7]. The idea of variable step size is to use large step size when the algorithm is far from the solution to speed up the convergence and use small step size when the algorithm approaches to the solution to achieve small error.

Another type of adaptive filters families is the Recursive Least Squares (RLS) algorithms [8]. RLS can be classified as a stochastic gradient approximation to a Steepest Descent algorithm, in RLS a sophisticated approximation for the covariance matrix of the input $\boldsymbol{R_u} = E[\boldsymbol{u}^*\boldsymbol{u}]$ is used. However, its significance is more obvious if it's considered as the exact solution to this estimation problem

$$\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{w}\|^2 \tag{1.2}$$

where $\boldsymbol{y}$ is a column vector which is consisting of the observed signals and the matrix $\boldsymbol{H}$ is consisting of the regressors $\boldsymbol{u}$ vectors. The RLS algorithm is more costly than LMS-family algorithms [1, 8], where an order of M^2 operations per iteration is needed. However, it converges faster than LMS [1].

1.2 Applications of Adaptive Filters

There are a great number of different applications for adaptive filters. They could be applied in different fields such as, telecommunications, Radar, sonar, video-audio signal processing and noise reduction. The difference between these applications depends on the method of generating the desired signal $d(i)$. Some applications of adaptive filters are briefed in this section to assert their diversity and necessity.

1.2.1 System Identification

System identification is a way to model an unknown system. In Figure 1.1 the unknown system and the adaptive filter are excited by the sequence $\boldsymbol{u}_i$. The output from the unknown system is described by (1.3)

$$d(i) = \boldsymbol{u}_i \boldsymbol{c} + v(i) \tag{1.3}$$

where the column vector $\boldsymbol{c}$ is the impulse response of the unknown system and $v(i)$ is a random noise.

At each time instant i the output $d(i)$ is compared with the output from the adaptive filter $\hat{d}(i) = \boldsymbol{u}_i \boldsymbol{w}_{i-1}$. The error signal $e(i) = d(i) - \hat{d}(i)$ is used to adjust the adaptive filter coefficients after each iteration. In steady states the error signal will be small (if the stability conditions are satisfied) or the output from the adaptive filter will be close to the output from the unknown system. This convergence assumes that the adaptive filter characteristics will be closed to the unknown system [1].

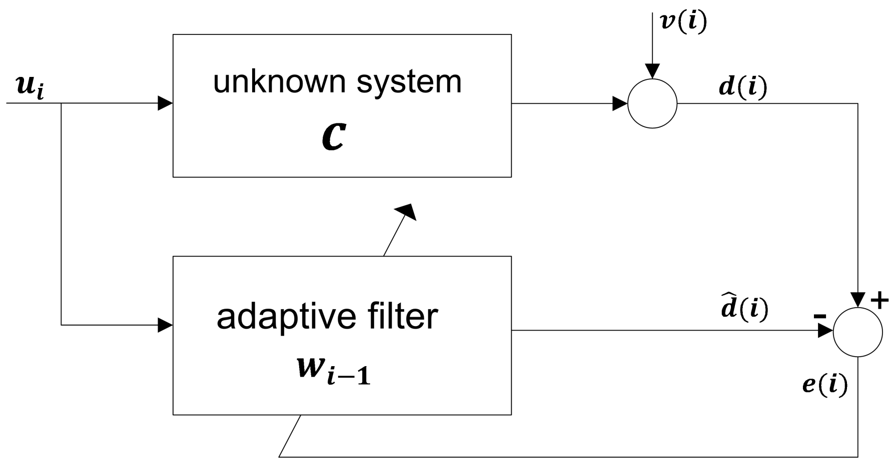

Figure 1.1: Adaptive filter for system identification

1.2.2 Linear Prediction

Linear prediction provides the best prediction of the signal at a future time. This application is used in speech processing applications such as speech coding in cellular telephony, speech enhancement and speech recognition. In this method the input of the adaptive filter is the delayed version of the desired signal. The error signal $e(i) = d(i) - \hat{d}(i)$ is used to adjust the coefficients of the adaptive filter after each iteration.

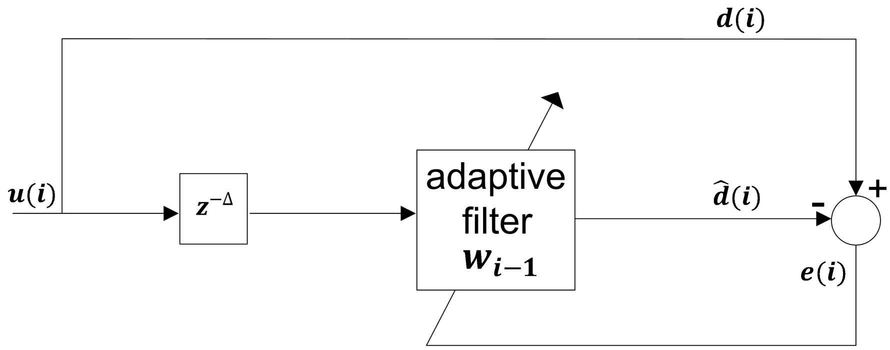

Figure 1.2: Adaptive filter for linear prediction

In steady states the error signal will be small (if the stability conditions are satisfied) or the output from the adaptive filter will be close to desired signal [9]. The linear prediction system is shown in Figure 1.2.

1.2.3 Inverse modeling or Equalization

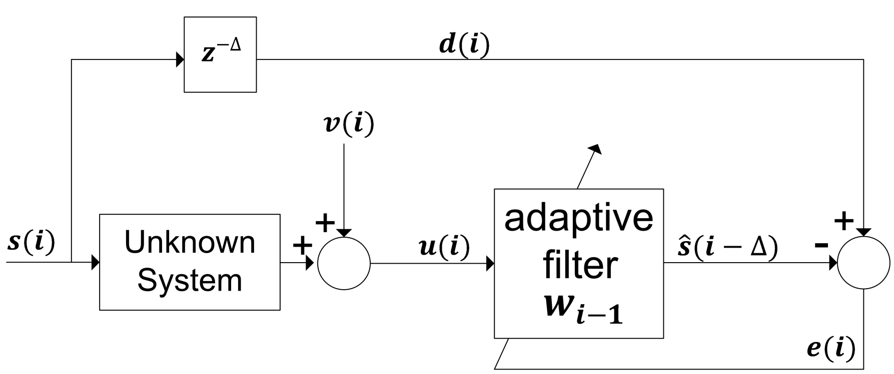

Figure 1.3: Adaptive channel equalizer

In Figure 1.3 at each time instant i the signal $d(i) = s(i-\triangle)$ is compared with the output from the adaptive filter $\hat{s}(i-\triangle)$ and the error signal $e(i) = d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}$ is generated. This signal error will be used in adjusting the adaptive filter coefficients after each iteration. In steady states the error signal will be small (if the stability conditions are satisfied) or the output from the adaptive filter $\hat{s}(i - \triangle)$ will be close to the output from the delay system $s(i-\triangle)$. This convergence assumes that characteristics of the adaptive filter will be close to the inverse of the unknown system [1].

1.2.4 Line Echo Cancellation

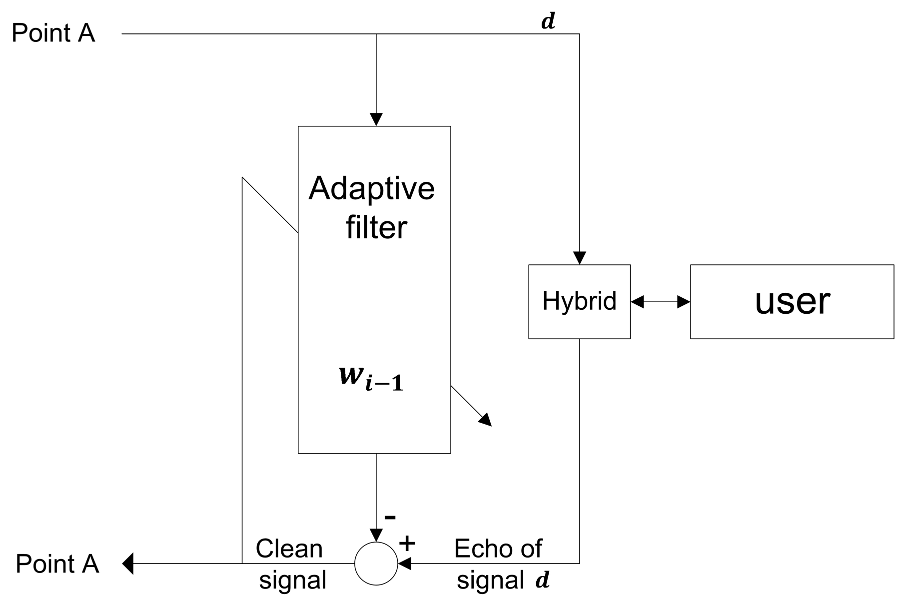

Figure 1.4: Adaptive Line Echo Canaller (LEC)

The signal d in Figure 1.3 travels back to point A as an echo plus the signal from the user. This echo results from the mismatch in circuitry. To overcome this problem an adaptive line echo canaller (LEC) is employed. At the user end the input to the adaptive LEC is the signal coming from A while the reference signal is its reflected version. The adaptive LEC generates a signal similar to d. Thus, it cancels its own echo and a clean signal is transmitted back to A [1].

1.3 Effect of Input Correlation on the Performance of Adaptive Filters

Studying the performance of adaptive filters is important; this study either in the transient or in the steady state will help researchers to understand the influences that will affect the performance of adaptive filters, to improve them and to compare between them.

Many researchers tried to study the behavior of adaptive filters from several aspects. However, due to the nature of these filters, they are time-variant and nonlinear systems, these studies often face some challenges. To overcome these challenges, they must rely on some assumptions that will facilitate this task. Such assumptions are small step size, separation principle, Gaussian assumption and long filters.

The performance studies of the adaptive filters in literature are classified under two regions; steady state and transient behavior.

The steady state analysis relies on **the Energy conservation relation** which was originally derived in [10] and **the variance relation** derived from the energy relation [11, 12].

There are a lot of studies done before these relations, and using these relations will give the same results. For LMS filter, the same result of the work in [13] was achieved by employing theses relations and the small step size assumption. Also, by applying the separation principle, the same result that obtained in [14] was derived by using theses relations.

For the NLMS filter the result that obtained in [15] for the EMSE was achieved by using this energy relation.

After that, new weighted versions of the Energy conservation relation and the variance relation were derived in [16], this weighted relations helped many researchers to investigate the transient part as well as the steady state part in the performance analysis. These relations are described in equations (1.4) and (1.5).

$$E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{\sigma}} = E\|\tilde{\boldsymbol{w}}_i\|^2_{diag\{\boldsymbol{F}\boldsymbol{\sigma}\}} + \mu\sigma_v^2 E\left[\frac{\|\boldsymbol{u}_i\|^2_{\boldsymbol{\sigma}}}{g^2[\boldsymbol{u}_i]}\right] \tag{1.4}$$

$$\boldsymbol{F} = \boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\boldsymbol{B} \tag{1.5}$$

where $\boldsymbol{F}$ is an $M \times M$ matrix, $\boldsymbol{\sigma}$ is a row vector $M \times 1$, $\boldsymbol{A} = 2E\left[\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right]$, $\boldsymbol{B} = E\left[\frac{\|\boldsymbol{u}_i\|^2_{\boldsymbol{\sigma}}\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right]$, $\|\tilde{\boldsymbol{w}}_i\|^2_{diag\{\boldsymbol{F}\boldsymbol{\sigma}\}} = \tilde{\boldsymbol{w}}_i^*\big(diag\{\boldsymbol{F}\boldsymbol{\sigma}\}\big)\tilde{\boldsymbol{w}}_i$, $g[\boldsymbol{u}_i]$ is a nonlinear function of the regressor $\boldsymbol{u}_i$, μ is the step size and σ_v^2 is the nose variance.

From the weighted energy relation, we can derive the EMSE and MSD learning curves as well as the steady state values. For example, in order to compute the EMSE leaning curve we chose $\boldsymbol{\sigma} = \boldsymbol{q}$ where $\boldsymbol{q} = [1\ ,\ 1\ ,\ \ldots\ ,\ 1]^T$ and for the steady state EMSE, it can be obtained by choosing this vector as $\boldsymbol{\sigma} = (\boldsymbol{I} - \boldsymbol{F})^{-1}\boldsymbol{q}$. From this, we can see that the performance of any adaptive filter depends on the matrices of moments $\boldsymbol{A}$ and $\boldsymbol{B}$ which depend on the statistics of the input. So, it is worthwhile to investigate effect of input correlation on the performance of the adaptive filters.

1.4 Thesis Objectives and Organization

The main objective of this work is to study the behavior of adaptive filters. This study is divided into two parts; the first part is to find a mathematical link between the input correlation scenarios and the performance of the adaptive filters. The second part of this thesis is deriving and studying the steady state performance of the RLS filter. According to these objectives the organization of this thesis will be as follows:

1.4.1 Correlation Effects and Majorization

The first part of this thesis (chapter 2 - chapter 4) is organized to fulfill this objective. Chapter 2 presents a description of the procedure of developing the stochastic gradient approximations or algorithms from the steepest descent methods. The steepest descent itself is also introduced as an iterative solution of the Wiener solution. These stochastic gradient algorithms results from the steepest descent methods by replacing the exact gradient vector and Hessian matrices with instantaneous approximations. Moreover, this chapter also describes the performance of the adaptive filters by deriving the mathematical expressions for the performance measures [1]. In chapter 3, a description of the majorization theory is presented by introducing the basic concepts and the main features that will help in this study [17]. While in chapter 4 the utilizing of the majorization theory and its techniques is introduced. In this chapter, the performance of some adaptive algorithms is studied using majorization theory.

1.4.2 Performance Analysis of the RLS Filter

In this part from the thesis (chapter 5 and chapter 6), a new approach for studying the steady state performance of the Recursive Least Square (RLS) adaptive filter for a circularly[1] correlated Gaussian input is presented. The mean-square analysis of the RLS relies on the moment of the random variable $\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}$, where $\boldsymbol{P}_i$ is the estimate of the inverse of input covariance matrix. Earlier approaches evaluate this moment by assuming that the $\boldsymbol{u}_i$ and $\boldsymbol{P}_i$ are independent which could result in negative value of the steady state Excess Mean Square Error (EMSE). In this work, this assumption is avoided and a closed form expression for this moment is derived. This derivation is based on finding the cumulative distribution function (CDF) of the random variable of the form $\frac{1}{\gamma+||\boldsymbol{u}||^2_D}$, where $\boldsymbol{u}$ is circular correlated Gaussian input and $\mathbf{D}$ is a diagonal matrix. As a result, more accurate estimation of the EMSE of the RLS filter is obtained. Simulation results corroborate the analytical findings.

[1]The following complex random variable $\boldsymbol{z} = \boldsymbol{x} + j\boldsymbol{y}$ is circular Gaussian random variable if $\boldsymbol{x}$ and $\boldsymbol{y}$ are real-valued Gaussian random variables with zero means, $\boldsymbol{R_x} = \boldsymbol{R_y}$ and $\boldsymbol{R_{xy}} = -\boldsymbol{R_{yx}}$. Where $\boldsymbol{R_x} = E[\boldsymbol{x}\boldsymbol{x}^*]$, $\boldsymbol{R_y} = E[\boldsymbol{y}\boldsymbol{y}^*]$ and $\boldsymbol{R_{xy}} = E[\boldsymbol{x}\boldsymbol{y}^*]$. This type of random variables is used in any two dimensions case, e.g., in Quadrature amplitude modulation (QAM) modulation and in Image Processing.

CHAPTER 2

ADAPTIVE FILTERING ALGORITHMS

An adaptive filter is an iterative algorithm. In each iteration i the adaptive filter updates its coefficients vector $\boldsymbol{w}_i$ $(M \times 1)$ to reach the optimal solution $\boldsymbol{w}^o$ as $i \to \infty$.

A general form for the adaptive filters to update $\boldsymbol{w}_i$ is given by

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + F(\boldsymbol{u}_i, e(i), \mu) \tag{2.1}$$

$$e(i) = d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1} \tag{2.2}$$

Here, $\boldsymbol{u}_i$ is a $(1 \times M)$ zero mean random input sequence, $d(i)$ is the desired signal, the parameter μ is the step size and $F(\cdot)$ is a function of all these quantities. The quantity $e(i)$ is the estimation error between the desired value $d(i)$ and the filter output at time i $(\boldsymbol{u}_i \boldsymbol{w}_{i-1})$.

In general, adaptive algorithms can be classified according to the estimation problem that will be solved by either of the two objective functions (cost functions) listed in Table 2.1.

Table 2.1: Classification of the adaptive algorithms

Algorithms	Objective function for estimation problem	Examples
Least mean algorithms	$\min_{w} E\|d - \boldsymbol{uw}\|^2$	LMS and variations of LMS
Least squares algorithms	$\min_{w} \|\boldsymbol{y} - \boldsymbol{Hw}\|^2$	RLS and Exponentially Weighted RLS

In the next section, we present the procedure to evaluate the adaptive filter's weights of some of the algorithms based on objective functions mentioned in Table 2.1. Moreover, we also investigate the different performance measures of these adaptive algorithms.

2.1 Least Mean Algorithms

The purpose of this section is to introduce the family of Least Mean Squares algorithms. The Wiener filter provides an optimum solution to the least mean objective function. The Steepest Decent method (SD) is an iterative solution for Wiener Solution. This iterative procedure will start from an initial guess for the solution and will give a better approximation as time progresses. Then, the Least Mean Squares is obtained from the Steepest Decent method by replacing the exact gradient vectors and Hessian matrices by some instantaneous approximations. These three classes of algorithms will be described in the following subsections.

2.1.1 The Wiener Solution

Let d be zero-mean scalar-valued random variable with variance σ_d^2, and let $\boldsymbol{u}$ be $1 \times M$ zero-mean random row vector with a positive definite covariance matrix $\boldsymbol{R_u}$. The target is to estimate d from $\boldsymbol{u}$ in the linear least mean squares sense by implementing the following optimization problem as mentioned in Table 2.1

$$\min_{\boldsymbol{w}} J(\boldsymbol{w}) = \min_{\boldsymbol{w}} E|d - \boldsymbol{u}\boldsymbol{w}|^2 \tag{2.3}$$

$$= \min_{\boldsymbol{w}} E\big(|d|^2 + \boldsymbol{w}^*\boldsymbol{u}^*\boldsymbol{u}\boldsymbol{w} - d^*\boldsymbol{u}\boldsymbol{w} - \boldsymbol{w}^*\boldsymbol{u}^*d\big) \tag{2.4}$$

$$= \min_{\boldsymbol{w}} \big(\sigma_d^2 + \boldsymbol{w}^*\boldsymbol{R_u}\boldsymbol{w} - \boldsymbol{R_{ud}}\boldsymbol{w} - \boldsymbol{w}^*\boldsymbol{R_{du}}\big) \tag{2.5}$$

where $\boldsymbol{w}$ is an $M \times 1$ defined as the weight vector which gives an estimate of optimum weights given ahead in equation (2.7), $\boldsymbol{R_u} = E\boldsymbol{u}^*\boldsymbol{u}$ and $\boldsymbol{R_{du}} = Ed\boldsymbol{u}^*$. The complex gradient vector of $J(\boldsymbol{w})$ with respect to $\boldsymbol{w}$ is

$$\nabla_{\boldsymbol{w}} J(\boldsymbol{w}) = \boldsymbol{w}^*\boldsymbol{R_u} - Ed^*\boldsymbol{u} \tag{2.6}$$

By equating (2.6) to zero and solving for $\boldsymbol{w}$, the minimizer or the desired solution for this optimization problem (denoted as $\boldsymbol{w}^o$) is given by

$$\boldsymbol{w}^o = \boldsymbol{R_u}^{-1}\ \boldsymbol{R_{du}} \tag{2.7}$$

The resulting minimum mean square error or J_{min} will be [1]

$$m.m.s.e = J_{min} = \sigma_d^2 - \boldsymbol{R_{ud}}\boldsymbol{R_u}^{-1}\boldsymbol{R_{du}} \tag{2.8}$$

2.1.2 Implementation of the Wiener Solution

The solution $\boldsymbol{w}^o$ in equation (2.7) is given in closed form. Sometimes it is not possible to find this solution in closed form for different performance criteria, i.e., performance criteria different from that of (2.3), other than the mean square error criterion [1]. Then, an approximation for the solution $\boldsymbol{w}^o$ can be found in an iterative way.

This iterative procedure can be in the form:

$$(\text{new guess of } \boldsymbol{w}^o) = (\text{old guess of } \boldsymbol{w}^o) + (\text{a correction term})$$

or

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu\boldsymbol{p}\ ,\ i \geq 0 \tag{2.9}$$

where $\boldsymbol{w}_i$ is a guess for the solution $\boldsymbol{w}^o$ at iteration i, $\boldsymbol{p}$ is the update direction vector and the positive scalar μ is called the step size. The success of this iterative method depends on effective choices of $\boldsymbol{p}$ and the step size μ.

The target from this iterative algorithm is to guarantee that the cost function $J(\boldsymbol{w}_i)$ is reduced along the direction $\boldsymbol{p}$ with each iteration, i.e.,

$$J(\boldsymbol{w}_i) < J(\boldsymbol{w}_{i-1}) \tag{2.10}$$

Starting from the objective function $J(\boldsymbol{w}_i)$ in (2.5) and by replacing $\boldsymbol{w}$ by its value in (2.9) to relate $J(\boldsymbol{w}_i)$ and $J(\boldsymbol{w}_{i-1})$ as

$$J(\boldsymbol{w}_i) = \sigma_d^2 + \boldsymbol{w}_i^* \boldsymbol{R_u} \boldsymbol{w}_i - \boldsymbol{R_{ud}} \boldsymbol{w}_i - \boldsymbol{w}^* \boldsymbol{R_{du}} \tag{2.11}$$

$$= \sigma_d^2 + (\boldsymbol{w}_{i-1} + \mu \boldsymbol{p})^* \boldsymbol{R_u} (\boldsymbol{w}_{i-1} + \mu \boldsymbol{p}) - \boldsymbol{R_{ud}} (\boldsymbol{w}_{i-1} + \mu p) - (\boldsymbol{w}_{i-1} + \mu \boldsymbol{p})^* \boldsymbol{R_{du}} \tag{2.12}$$

$$= J(\boldsymbol{w}_{i-1}) + \mu(\boldsymbol{w}_{i-1}^* \boldsymbol{R_u} - \boldsymbol{R_{du}^*}) \boldsymbol{p} + \mu \boldsymbol{p}^* (\boldsymbol{R_u} \boldsymbol{w}_{i-1} - \boldsymbol{R_{du}}) + \mu^2 \boldsymbol{p}^* \boldsymbol{R_u} \boldsymbol{p} \tag{2.13}$$

$$= J(\boldsymbol{w}_{i-1}) + 2\mu \mathbf{Re}[\nabla_{\boldsymbol{w}} J(\boldsymbol{w}_{i-1}) \boldsymbol{p}] + \mu^2 \boldsymbol{p}^* \boldsymbol{R_u} \boldsymbol{p} \tag{2.14}$$

where $\nabla_{\boldsymbol{w}} J(\boldsymbol{w}_{i-1}) = \boldsymbol{w}_{i-1}^* \boldsymbol{R_u} - \boldsymbol{R_{du}^*}$.

From (2.14) the term $(\mu^2 \boldsymbol{p}^* \boldsymbol{R_u} \boldsymbol{p})$ is positive for all nonzero $\boldsymbol{p}$ since $\boldsymbol{R_u} > 0$, then the condition that we need for the case in (2.10) to be satisfied is

$$Re[\nabla_{\boldsymbol{w}} J(\boldsymbol{w}_{i-1}) \boldsymbol{p}] < 0 \tag{2.15}$$

There are many choices of vector $\boldsymbol{p}$ that satisfy (2.15). For example, any $\boldsymbol{p}$ of the form [1]:

$$\boldsymbol{p} = -\boldsymbol{B} [\nabla_{\boldsymbol{w}} J(\boldsymbol{w}_{i-1})]^* \tag{2.16}$$

for any Hermitian positive definite matrix $\boldsymbol{B}$ will satisfy (2.15).

In the **Steepest Descent (SD)** method, $\boldsymbol{B}$ is simply the identity matrix I [1,18]. Then, the update direction vector $\boldsymbol{p}$ will be

$$\begin{aligned}\boldsymbol{p} &= -[\nabla_{\boldsymbol{w}} J(\boldsymbol{w}_{i-1})]^* \\ &= \boldsymbol{R_{du}} - \boldsymbol{R_u}\boldsymbol{w}_{i-1}\end{aligned} \tag{2.17}$$

with this vector $\boldsymbol{p}$ the recursion in (2.9) will be

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu[\boldsymbol{R_{du}} - \boldsymbol{R_u}\boldsymbol{w}_{i-1}] \ , \quad \boldsymbol{w}_{-1} = 0 \tag{2.18}$$

Introducing the weight error vector $\tilde{\boldsymbol{w}}_i = \boldsymbol{w}^o - \boldsymbol{w}_i$. Subtracting both sides of (2.18) from $\boldsymbol{w}^o$ yields

$$\tilde{\boldsymbol{w}}_i = [\boldsymbol{I} - \mu\boldsymbol{R_u}]\tilde{\boldsymbol{w}}_{i-1} \tag{2.19}$$

Analyzing this recursion will give the following condition on the step size μ to ensure the convergence ($\tilde{\boldsymbol{w}}_i \to 0$ or $\boldsymbol{w}_i \to \boldsymbol{w}^o$ as $i \to \infty$) if and only if

$$0 < \mu < \frac{2}{\lambda_{max}} \tag{2.20}$$

and the optimal step size represented by μ^o [1], is given by

$$\mu^o < \frac{2}{\lambda_{max} + \lambda_{min}} \tag{2.21}$$

where λ_{max} and λ_{min} are the largest and the smallest eigenvalues of $\boldsymbol{R_u}$ respectively.

In the **Newton's Method** the matrix $\boldsymbol{B}$ in (2.16) is the inverse of the exact Hessian $[\nabla_w^2 J(w_{i-1})]^{-1}$, then the recursion (2.9) will be

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{R_u}^{-1}[\boldsymbol{R_{du}} - \boldsymbol{R_u}\boldsymbol{w}_{i-1}] \ , \quad \boldsymbol{w}_{-1} = 0 \tag{2.22}$$

The weight error recursion for the Newton's Method is

$$\tilde{\boldsymbol{w}}_i = (1-\mu)\tilde{\boldsymbol{w}}_{i-1} \tag{2.23}$$

The condition on the step size to ensure convergence will be

$$|1-\mu| < 1 \Leftrightarrow 0 < \mu < 2 \tag{2.24}$$

which is independent of $\boldsymbol{R_u}$. Furthermore, the convergence will be guaranteed for a single iteration if the step size is chosen as $\mu = 1$ [1, 18].

Sometimes the Hessian matrix $[\nabla_{\boldsymbol{w}}^2 J(\boldsymbol{w}_{i-1})]$ is close to a singular. To avoid this scenario we can use regularization. Thus, in **Regularized Newton's** method the matrix $\boldsymbol{B}$ in (2.16) is set to

$$\boldsymbol{B} = [\epsilon \boldsymbol{I} + \nabla_{\boldsymbol{w}}^2 J(\boldsymbol{w}_{i-1})]^{-1}$$

for $\epsilon > 0$, the recursion (2.1) becomes

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu[\epsilon I + \boldsymbol{R_u}]^{-1}[\boldsymbol{R_{du}} - \boldsymbol{R_u}\boldsymbol{w}_{i-1}] \ , \quad \boldsymbol{w}_{-1} = 0 \tag{2.25}$$

2.2 Stochastic Gradient Approximations of Steepest Descent

Stochastic gradient algorithms are obtained from the previous algorithms by replacing gradient vectors and the Hessian matrix by some approximations. We need this to avoid the need of the exact signal statistics (covariance and cross-covariance e.g. $\boldsymbol{R_{du}}$ and $\boldsymbol{R_u}$) which are in general not available practically, and we can't implement these algorithms without these statistics.

One of the simplest approximations for these moments is to drop the expectation operator and instead employ the instantaneous values as follows

$$\boldsymbol{R_{du}} \approx d(i)\boldsymbol{u}_i^*, \qquad \boldsymbol{R_u} \approx \boldsymbol{u}_i^*\boldsymbol{u}_i \tag{2.26}$$

where $d(i) \in \{d(0)\ ,\ d(1)\ ,\ d(2)\ ,\ \ldots\}$ and $\boldsymbol{u}_i \in \{\boldsymbol{u}_0\ ,\ \boldsymbol{u}_1\ ,\ \boldsymbol{u}_2\ ,\ \ldots\}$.

2.2.1 Least-Mean-Square algorithm (LMS)

Using these approximations of (2.26) in the Steepest Descent recursion in (2.18) yields the Least Mean Squares (LMS) algorithms.

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu\boldsymbol{u}_i^* e(i)\ , \quad \boldsymbol{w}_{-1} = 0 \tag{2.27}$$

where $e(i) = d(i) - \boldsymbol{u}_i\boldsymbol{w}_{i-1}$

2.2.2 Normalized Least-Mean-Square algorithm (NLMS)

By using these approximations, recursion (2.25) becomes

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu[\epsilon \boldsymbol{I} + \boldsymbol{u}_i^* \boldsymbol{u}_i]^{-1} \boldsymbol{u}_i^* [d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}] \tag{2.28}$$

Using the matrix inversion lemma to simplify the term $[\epsilon I + \boldsymbol{u}_i^* \boldsymbol{u}_i]^{-1}$ yields

$$[\epsilon \boldsymbol{I} + \boldsymbol{u}_i^* \boldsymbol{u}_i]^{-1} = \epsilon^{-1} \boldsymbol{I} - \frac{\epsilon^{-2}}{1 + \epsilon^{-1} \|\boldsymbol{u}_i\|^2} \boldsymbol{u}_i^* \boldsymbol{u}_i \tag{2.29}$$

multiplying both sides of (2.29) by $\boldsymbol{u}_i^*$ from the right yields

$$[\epsilon \boldsymbol{I} + \boldsymbol{u}_i^* \boldsymbol{u}_i]^{-1} \boldsymbol{u}_i^* = \epsilon^{-1} \boldsymbol{u}_i^* - \frac{\epsilon^{-2}}{1 + \epsilon^{-1} \|\boldsymbol{u}_i\|^2} \boldsymbol{u}_i^* \|\boldsymbol{u}_i\|^2 \tag{2.30}$$

$$= \frac{\boldsymbol{u}_i^*}{\epsilon + \|\boldsymbol{u}_i\|^2} \tag{2.31}$$

Using this result in equation (2.28), we obtain

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \frac{\mu}{\epsilon + \|\boldsymbol{u}_i\|^2} \boldsymbol{u}_i^* [d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}] \quad , i \geq 0 \ , \boldsymbol{w}_{-1} = 0 \tag{2.32}$$

This stochastic gradient approximation is known as the ϵ-Normalized LMS (ϵ-NLMS) algorithm [1]. If the parameter ϵ is set to zero the resulting algorithm is known as NLMS algorithm.

<u>Remark:</u>

The update recursions in (2.27) and (2.32) are both special cases of the following general form

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \frac{\boldsymbol{u}_i^*}{g[\boldsymbol{u}_i]} e(i) \tag{2.33}$$

$$e(i) = d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1} \tag{2.34}$$

Several adaptive filters algorithms can be extracted from this general form by changing the generic function $g[\boldsymbol{u}_i]$ of the regression vector $\boldsymbol{u}_i$. Examples of these filters are listed in Table 2.2.

Table 2.2: Different variations of Least Mean Squares algorithms

Algorithm	$g[\boldsymbol{u}_i]$
LMS	1
NLMS	$\|\boldsymbol{u}_i\|^2$
ϵ-NLMS	$\epsilon + \|\boldsymbol{u}_i\|^2$

Other adaptive filters with error nonlinearity can be derived from the following general form

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* f[e(i)] \tag{2.35}$$

Examples of these filters [1, 12, 19–21] are listed in Table 2.3.

Table 2.3: Different variations of Least Mean Squares algorithms

Algorithm	$f[e(i)]$
LMS [22]	$e(i)$
LMF [20]	$e(i)\|e(i)\|^2$
NLMF [23]	$\frac{e(i)\|e(i)\|^2}{\\|\boldsymbol{u}_i\\|^2}$
LMMN [24]	$e(i)(\alpha + (1-\alpha)\|e(i)\|^2)$
Sign-error [25]	$csgn[e(i)]$

2.3 Performance Measures of Adaptive Filters

It is important to test the performance of these adaptive filters algorithms in order to classify or compare them. In this work a common performance measure will be used. Four quantities for measuring the performance will be considered. These quantities can be either functions of i (iteration time to generate the learning curves) or in terms of the steady state values (as $i \to \infty$).

The performance measures are:

1. The Mean Square Error (MSE)

$$MSE(i) = E|d(i) - \boldsymbol{u}_i\boldsymbol{w}_{i-1}|^2 = J(\boldsymbol{w}_i) \tag{2.36}$$

where $J_{min} = \sigma_d^2 - \boldsymbol{R_{ud}}\boldsymbol{R_u}^{-1}\boldsymbol{R_{du}}$.

2. The Excess Mean Square Error (EMSE)

$$EMSE(i) = MSE(i) - J_{min} \tag{2.37}$$

3. The Missadjustment $\mathcal{M}(i)$

$$\mathcal{M}(i) = EMSE(i)/J_{min} \tag{2.38}$$

4. The Mean Square Deviation (MSD)

$$MSD(i) = E\|\boldsymbol{w}^o - \boldsymbol{w}_i\|^2 \tag{2.39}$$

2.3.1 Performance Analysis of Steepest Descent algorithm and Newton's Method

The original problem is in the case of Steepest Descent and Newton's Method is to minimize the following objective function

$$J(\boldsymbol{w}) = E|d - \boldsymbol{u}\boldsymbol{w}|^2$$

Dealing with $J(\boldsymbol{w})$ as a function i and using $\boldsymbol{w}_i$ as an estimate of $\boldsymbol{w}$ will give a useful information about the behavior of these two filters.

The learning curve (it is also known as the mean square error (MSE) curve) and denoted by $J(\boldsymbol{w}_i)$ will be used to represent the behavior of theses algorithms.

Starting from the definition of $J(\boldsymbol{w}_i)$ as

$$J(\boldsymbol{w}_i) = E|d - \boldsymbol{u}\boldsymbol{w}_i|^2 \tag{2.40}$$

$$= \sigma_d^2 + \boldsymbol{w}^*\boldsymbol{R_u}\boldsymbol{w} - \boldsymbol{R_{ud}}\boldsymbol{w} - \boldsymbol{w}^*\boldsymbol{R_{du}} \tag{2.41}$$

$$= \begin{bmatrix} 1 & \boldsymbol{w}_i^* \end{bmatrix} \begin{bmatrix} \sigma_d^2 & -\boldsymbol{R_{ud}} \\ -\boldsymbol{R_{du}} & \boldsymbol{R_u} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{w}_i \end{bmatrix} \tag{2.42}$$

from [1] the center matrix can be factorized as

$$\begin{bmatrix} \sigma_d^2 & -\boldsymbol{R_{ud}} \\ -\boldsymbol{R_{du}} & \boldsymbol{R_u} \end{bmatrix} = \begin{bmatrix} 1 & -\boldsymbol{R_{ud}}\boldsymbol{R_u}^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_d^2 - \boldsymbol{R_{ud}}\boldsymbol{R_u}^{-1}\boldsymbol{R_{du}} & 0 \\ 0 & \boldsymbol{R_u} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\boldsymbol{R_{ud}}\boldsymbol{R_u}^{-1} & 1 \end{bmatrix} \tag{2.43}$$

substituting this in (2.42) will give

$$J(\boldsymbol{w}_i) = (\sigma_d^2 - \boldsymbol{R_{ud}}\boldsymbol{R_u}^{-1}\boldsymbol{R_{du}}) + (\boldsymbol{w}_i - \boldsymbol{R_u}^{-1}\boldsymbol{R_{du}})^*\boldsymbol{R_u}^{-1}(\boldsymbol{w}_i - \boldsymbol{R_u}\boldsymbol{R_{du}}) \tag{2.44}$$

$$= J_{min} + \tilde{\boldsymbol{w}}_i^*\boldsymbol{R_u}\tilde{\boldsymbol{w}}_i \tag{2.45}$$

Introducing the Eigen-decomposition of $\boldsymbol{R_u} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ and replacing $\tilde{\boldsymbol{w}}_i$ by $\boldsymbol{U}x_i$, equation (2.45) will be

$$J(\boldsymbol{w}_i) = J_{min} + \tilde{\boldsymbol{w}}_i^*\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*\tilde{\boldsymbol{w}}_i \tag{2.46}$$

$$= J_{min} + x_i^*\boldsymbol{U}^*\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*\boldsymbol{U}x_i \tag{2.47}$$

$$= J_{min} + \sum_{k=1}^{M} \lambda_k |\boldsymbol{x}_k(i)|^2 \tag{2.48}$$

where $\boldsymbol{x}_k(i)$ is the kth entry of $\boldsymbol{x}_i$ and $\boldsymbol{U}\boldsymbol{U}^* = \boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I}$.

The vector $\boldsymbol{x}_i$ can be derived from (2.19) by multiplying both sides by $\boldsymbol{U}^*$ from the right and by replacing $\boldsymbol{R_u}$ by its decomposition to get

$$\boldsymbol{x}_i = [\boldsymbol{I} - \mu\boldsymbol{\Lambda}]\boldsymbol{x}_{i-1} \tag{2.49}$$

From (2.49), the k^{th} entry of $\boldsymbol{x}_i$ will be

$$x_k(i) = (1 - \mu\lambda_k)x_k(i-1) \tag{2.50}$$

$$= (1 - \mu\lambda_k)^{i+1}x_k(-1) \tag{2.51}$$

Substituting this in (2.48) yields

$$J(\boldsymbol{w}_i) = J_{min} + \sum_{k=1}^{M} \lambda_k(1 - \mu\lambda_k)^{2(i+1)}|\boldsymbol{x}_k(-1)|^2 \tag{2.52}$$

Using the same procedure above, the learning curve for Newton's algorithm can be shown to be

$$J(\boldsymbol{w}_i) = E|d - \boldsymbol{u}\boldsymbol{w}_i|^2 \tag{2.53}$$

$$= J_{min} + \sum_{k=1}^{M} \lambda_k|x_k(i)|^2 \tag{2.54}$$

$$= J_{min} + (1 - \mu)^{2(i+1)} \sum_{k=1}^{M} \lambda_k|x_k(-1)|^2 \tag{2.55}$$

where the k^{th} entry of $\boldsymbol{x}_i$ will be in this case

$$x_k(i) = (1-\mu)^{i+1} x_k(-1) \tag{2.56}$$

Taking the limit as $i \to \infty$ for the two learning curves in (2.52) and (2.55) will give the same result for both SD and Newton's algorithms as

$$\lim_{i\to\infty} J(\boldsymbol{w}_i) = J_{min} \tag{2.57}$$

this conclusion is constrained by choosing the step size μ according to the conditions that mentioned in (2.20) and (2.24) for SD and Newton's method respectively.

2.3.2 Performance Analysis of Least Mean Squares (LMS)

As mentioned in Chapter 1, the analysis of the stochastic gradient algorithms relies on the energy relation and its weighted version. The steady state analysis of the LMS filter can be done by utilizing the energy relation, but going in this way will not give exact expressions. Using the weighted version of the energy relation can be used to find the transient and the steady state behaviors. Moreover, the steady state analysis will be accurate and without relying on any assumptions, such as small step size and separation assumption.

The weighted energy conservation relation [10] is given by

$$\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{\Sigma}} + \frac{|e^{\boldsymbol{\Sigma}}_a(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{\Sigma}}} = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{\Sigma}} + \frac{|e^{\boldsymbol{\Sigma}}_p(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{\Sigma}}} \tag{2.58}$$

and the weighted variance relation [11, 12] is

$$E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{\Sigma}} = E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{\Sigma}'} + \mu^2\sigma_v^2 E\left(\frac{\|\boldsymbol{u}_i\|^2_{\boldsymbol{\Sigma}}}{g^2[\boldsymbol{u}_i]}\right) \tag{2.59}$$

$$\boldsymbol{\Sigma}' \quad = \boldsymbol{\Sigma} - \mu\boldsymbol{\Sigma}\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]} - \mu\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\boldsymbol{\Sigma} + \mu^2\frac{\|\boldsymbol{u}_i\|^2_{\boldsymbol{\Sigma}}}{g[\boldsymbol{u}_i]}\boldsymbol{u}_i^*\boldsymbol{u}_i \tag{2.60}$$

where the notation

$$\|\boldsymbol{x}\|^2_{\boldsymbol{\Sigma}} = \boldsymbol{x}^*\boldsymbol{\Sigma}\boldsymbol{x} \tag{2.61}$$

for some Hermitian positive definite weighting matrix $\boldsymbol{\Sigma}$, σ_v^2 is the noise variance and $g[\boldsymbol{u}_i]$ is a positive valued function of $\boldsymbol{u}_i$.

For the LMS filter where $g[\boldsymbol{u}_i] = 1$, the variance relation in (2.59) and (2.60) will be

$$E\|\bar{\boldsymbol{w}}_i\|^2_{\bar{\boldsymbol{\Sigma}}} = E\|\bar{\boldsymbol{w}}_i\|^2_{\bar{\boldsymbol{\Sigma}}'} + \mu^2\sigma_v^2 E\|\bar{\boldsymbol{u}}_i\|^2_{\bar{\boldsymbol{\Sigma}}} \tag{2.62}$$

$$\bar{\boldsymbol{\Sigma}}' \quad = \bar{\boldsymbol{\Sigma}} - \mu\bar{\boldsymbol{\Sigma}}\boldsymbol{u}_i^*\boldsymbol{u}_i - \mu\bar{\boldsymbol{u}}_i^*\bar{\boldsymbol{u}}_i\bar{\boldsymbol{\Sigma}} + \mu^2\|\bar{\boldsymbol{u}}_i\|^2_{\bar{\boldsymbol{\Sigma}}}\bar{\boldsymbol{u}}_i^*\bar{\boldsymbol{u}}_i \tag{2.63}$$

where in the last equations, the following transformation is used

$$\bar{\boldsymbol{w}}_i = \boldsymbol{U}^* \tilde{\boldsymbol{w}}_i, \quad \bar{\boldsymbol{u}}_i = \boldsymbol{w}_i \boldsymbol{U}, \quad \bar{\boldsymbol{\Sigma}} = \boldsymbol{U}^* \boldsymbol{\Sigma} \boldsymbol{U} \tag{2.64}$$

Assuming that $\bar{\boldsymbol{u}}_i$ is circular Gaussian with a diagonal covariance matrix $\boldsymbol{\Lambda}$, the variance relation in (2.62) and (2.63) can be rewritten as

$$E\|\bar{\boldsymbol{w}}_i\|^2_{\bar{\boldsymbol{\Sigma}}} = E\|\bar{\boldsymbol{w}}_i\|^2_{\bar{\boldsymbol{\Sigma}}'} + \mu^2 \sigma_v^2 Tr(\Lambda \bar{\boldsymbol{\Sigma}}) \tag{2.65}$$

$$\bar{\boldsymbol{\Sigma}}' \quad = \bar{\boldsymbol{\Sigma}} - \mu \bar{\boldsymbol{\Sigma}} \Lambda - \mu \Lambda \bar{\boldsymbol{\Sigma}} + \mu^2 [\Lambda Tr(\bar{\boldsymbol{\Sigma}} \Lambda) + \Lambda \bar{\boldsymbol{\Sigma}} \Lambda] \tag{2.66}$$

Now introduce the following $(M \times 1)$ vectors

$$\bar{\boldsymbol{\sigma}} = diag\{\bar{\boldsymbol{\Sigma}}\}, \quad \boldsymbol{\lambda} = diag\{\boldsymbol{\Lambda}\} \tag{2.67}$$

with these new vectors, the relations in (2.65) and (2.66) can be written as

$$E\|\bar{\boldsymbol{w}}_i\|^2_{diag\{\bar{\boldsymbol{\sigma}}\}} = E\|\bar{\boldsymbol{w}}_{i-1}\|^2_{diag\{\bar{\boldsymbol{F}}\bar{\boldsymbol{\sigma}}\}} + \mu^2 \sigma_v^2 (\boldsymbol{\lambda}^T \bar{\boldsymbol{\sigma}}) \tag{2.68}$$

$$E\|\bar{\boldsymbol{w}}_i\|^2_{\bar{\boldsymbol{\sigma}}} = E\|\bar{\boldsymbol{w}}_{i-1}\|^2_{\bar{\boldsymbol{F}}\bar{\boldsymbol{\sigma}}} + \mu^2 \sigma_v^2 (\boldsymbol{\lambda}^T \bar{\boldsymbol{\sigma}}) \tag{2.69}$$

where $\bar{\boldsymbol{\sigma}}' = \bar{\boldsymbol{F}}\bar{\boldsymbol{\sigma}}$, $\bar{\boldsymbol{F}} = (\boldsymbol{I} - 2\mu\boldsymbol{\Lambda} + \mu^2\boldsymbol{\Lambda}^2) + \mu^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^T$ and the $diag\{\}$ is dropped for compactness of notation.

The MSD learning curve which is defined in (2.39) is

$$MSD(i) = E\|\boldsymbol{w}^o - \boldsymbol{w}_i\|^2 = E\|\bar{\boldsymbol{w}}_i\|^2_{\boldsymbol{q}} \tag{2.70}$$

by setting the vector $\bar{\boldsymbol{\sigma}}$ in (2.68) to

$$\bar{\boldsymbol{\sigma}} = diag\{\boldsymbol{I}\} = [1\ ,\ \ldots\ ,\ 1]^T = \boldsymbol{q} \tag{2.71}$$

and iterating from $i = 0$ with $\bar{\boldsymbol{w}}_{-1} = \boldsymbol{w}^o$, we obtain the following expression for the MSD learning curve of the LMS filter

$$\begin{aligned} MSD(i) &= \|\boldsymbol{w}^o\|^2_{diag\{\bar{\boldsymbol{F}}^{i+1}\boldsymbol{q}\}} + \mu^2\sigma_v^2 \sum_{k=0}^{i} \boldsymbol{\lambda}^T \bar{\boldsymbol{F}}^k \boldsymbol{q} \\ &= \|\boldsymbol{w}^o\|^2_{\bar{\boldsymbol{F}}^{i+1}\boldsymbol{q}} + \mu^2\sigma_v^2 \sum_{k=0}^{i} \boldsymbol{\lambda}^T \bar{\boldsymbol{F}}^k \boldsymbol{q} \quad , i \geq 0 \end{aligned} \tag{2.72}$$

Figure 2.1 shows a comparison between the MSD learning curves from (2.72) and from simulation. In this simulation, these curves are generated by using the following $\boldsymbol{\lambda} = [1.2090, 1.0910, 1.0000, 0.9000, 0.8000]^T$, $\boldsymbol{w}^o = [0.1348, 0.2697, 0.4045, 0.5394, 0.6742]^T$, $\mu = 0.05$ and $\sigma_v^2 = 0.001$.

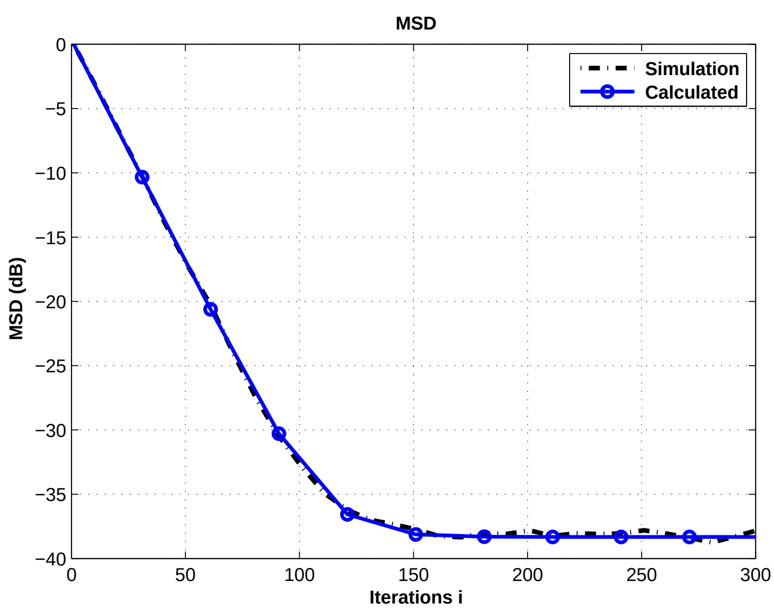

Figure 2.1: LMS learning curves

In steady state the recursion (2.69) will be

$$E\|\bar{\boldsymbol{w}}_\infty\|^2_{(\boldsymbol{I}-\bar{\boldsymbol{F}})\bar{\boldsymbol{\sigma}}} = \mu^2\sigma_v^2(\boldsymbol{\lambda}^T\bar{\boldsymbol{\sigma}}) \tag{2.73}$$

by setting the vector $\bar{\boldsymbol{\sigma}}$ in (2.75) to

$$\bar{\boldsymbol{\sigma}} = (\boldsymbol{I}-\bar{\boldsymbol{F}})^{-1}\boldsymbol{q} \tag{2.74}$$

and use this value for the vector $\bar{\boldsymbol{\sigma}}$ to make the weight in the left hand side of equation (2.75) equal to $\boldsymbol{q}$, i.e.,

$$E\|\bar{\boldsymbol{w}}_\infty\|^2_{\boldsymbol{q}} = \mu^2\sigma_v^2(\boldsymbol{\lambda}^T\boldsymbol{q}) \tag{2.75}$$

By setting the weight to $\boldsymbol{q}$, we obtain the MSD for the LMS filter, i.e.,

$$MSD = \frac{\sigma_v^2\mu\displaystyle\sum_{k=1}^{M}\frac{1}{2-\mu\lambda_k}}{1-\mu\displaystyle\sum_{k=1}^{M}\frac{\lambda_k}{2-\mu\lambda_k}} \tag{2.76}$$

Also by setting the vector $\bar{\boldsymbol{\sigma}}$ in (2.75) to

$$\bar{\boldsymbol{\sigma}} = (\boldsymbol{I}-\bar{\boldsymbol{F}})^{-1}\boldsymbol{\lambda} \tag{2.77}$$

We recover steady state EMSE of the LMS filter as

$$EMSE = \frac{\sigma_v^2 \mu \displaystyle\sum_{k=1}^{M} \frac{\lambda_k}{2 - \mu\lambda_k}}{1 - \mu \displaystyle\sum_{k=1}^{M} \frac{\lambda_k}{2 - \mu\lambda_k}} \tag{2.78}$$

From equations (2.76) and (2.78), the steady state errors MDS and EMSE will not go to zero anymore because the weight error vector $\tilde{\boldsymbol{w}}_i \nrightarrow 0$ ($\boldsymbol{w}_i \nrightarrow \boldsymbol{w}^o$) as $i \to \infty$ due to the gradient noise (using instantaneous values instead of the quantities $\{\boldsymbol{R_{du}}, \boldsymbol{R_u}\}$ as in (2.26)).

Figure 2.2 shows the MSE learning curves for the Steepest Descent and its stochastic approximation algorithm LMS. The difference between the EMSE for both filters are shown also.

Also, a comparison between the MSD learning curves for the Steepest Descent and its stochastic approximation algorithm LMS is shown in Figure 2.3.

In generating these curves, the following parameters are used: the input regressor is i.i.d. Gaussian random variable with covariance matrix $\boldsymbol{R_u} = diag\{5, 4, 3, 2, 1\}$, $\boldsymbol{w}^o = [0.1348, 0.2697, 0.4045, 0.5394, 0.6742]^T$, $\mu = 0.05$ and $\sigma_v^2 = 0.001$.

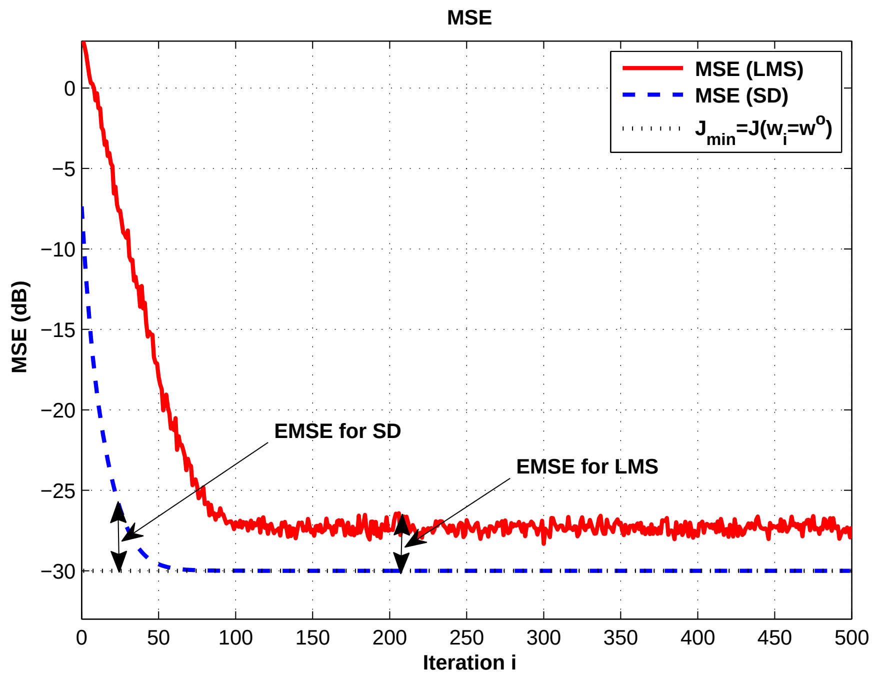

Figure 2.2: The MSE curves for the Steepest Decent and its stochastic approximation (LMS) , and the EMSE for the two algorithms

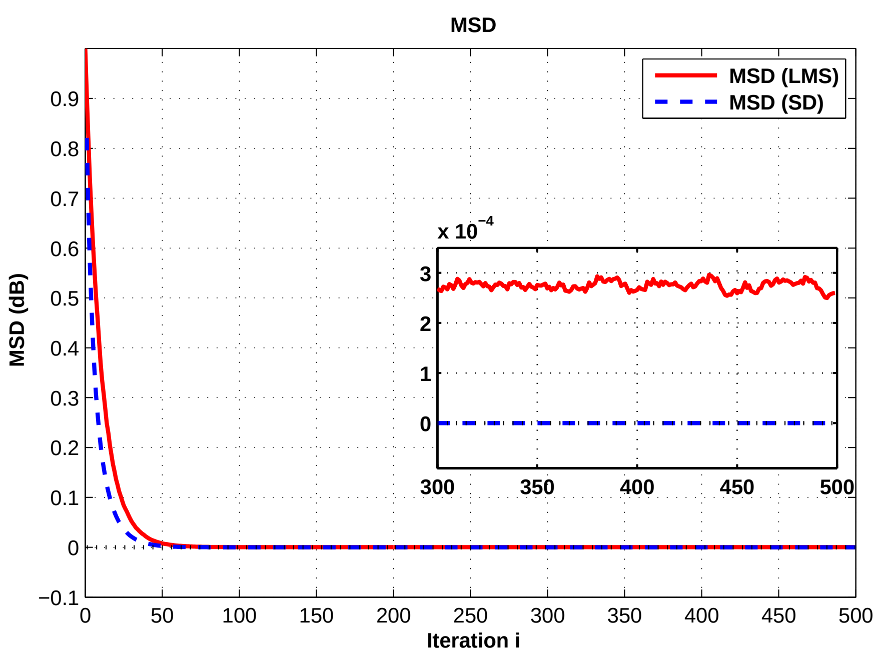

Figure 2.3: The MSD curves for the Steepest Decent and its stochastic approximation (LMS)

CHAPTER 3

MAJORIZATION THEORY

Majorization is a partial ordering on vectors which determines the degree of similarity between the vector elements. Functions that translate the ordering of vectors to a standard scalar ordering are known as Schur's functions (Schur-convex or Schur-concave functions).

Many problems arising in signal processing and communications involve comparing vector-valued strategies or solving optimization problems with vector-valued or matrix-valued variables. Majorization theory is a key tool that allows us to solve or simplify these problems. In this chapter, a brief introduction about the majorization theory and Schur's functions will be presented [17, 26, 27].

3.1 Basic Concepts

Definition 3.1 *(Majorization) For any two vectors* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^M$ *with descending order components* $x_1 \geq x_2 \geq \ldots \geq x_M \geq 0$ *and* $y_1 \geq y_2 \geq \ldots \geq y_M \geq 0$*, then the* *vector* $\boldsymbol{x}$ *majorizes the vector* $\boldsymbol{y}$ *written as* $\boldsymbol{x} \succ \boldsymbol{y}$ *if*

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i \ , \ \ k = 1, \ldots, M-1 \tag{3.1}$$

and

$$\sum_{i=1}^{M} x_i = \sum_{i=1}^{M} y_i \tag{3.2}$$

For example, it is easy to see from the definition that these $M \times 1$ vectors with sum S

$$\left(\frac{S}{M}, \frac{S}{M}, \ldots, \frac{S}{M}\right) \prec (S, 0, \ldots, 0) \tag{3.3}$$

In fact we can prove that these two vectors are upper/lower bounds in the sense that

$$\left(\frac{S}{M}, \frac{S}{M}, \ldots, \frac{S}{M}\right) \prec (a_1, a_2, \ldots, a_M) \prec (S, 0, \ldots, 0) \tag{3.4}$$

whenever $a_i \geq 0$, $\sum a_i = S$

Definition 3.2 *(Weak Majorization) For any two vectors* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^M$ *with descending order components* $x_1 \geq x_2 \geq \ldots \geq x_M \geq 0$ *and* $y_1 \geq y_2 \geq \ldots \geq y_M \geq 0$*, then the vector* $\boldsymbol{x}$ *weakly majorizes the vector* $\boldsymbol{y}$ *written as* $\boldsymbol{x} \,^{w}\!\!\succ \boldsymbol{y}$ *if*

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i \ , \ \ k = 1, \ldots, M-1 \tag{3.5}$$

From definitions 3.1 and 3.2 we can see that

$$\boldsymbol{x} \succ \boldsymbol{y} \Rightarrow \boldsymbol{x} \ {}^{w}\!\succ \boldsymbol{y} \tag{3.6}$$

Definition 3.3 *(Majorization Equivalents) The following conditions are equivalent to the majorization conditions in definition 3.1:*

1. *If $\boldsymbol{x} = \boldsymbol{y}\boldsymbol{P}$ for some doubly stochastic matrix $\boldsymbol{P}$, then $\boldsymbol{x} \succ \boldsymbol{y}$.*

2. *If $\sum \phi(x_i) \leq \sum \phi(y_i)$ for all continuous convex functions ϕ, then $\boldsymbol{x} \succ \boldsymbol{y}$.*

Note that the components of a vector $\boldsymbol{x}$ might not be order in a descending order. So, when we would like to compare two vectors, we first order their components in a descending order before actually comparing them.

3.2 Order-Preserving Functions: Schur Functions

Functions that preserve the ordering of majorization are known as Schur-convex (concave) functions. Next, we discuss some of their properties.

Definition 3.4 *(Schur-convex(concave) functions) A real function $\phi : \mathbb{R}^M \to \mathbb{R}$ is said to be Schur-convex (concave)*

$$\boldsymbol{x} \succ \boldsymbol{y} \Rightarrow \phi(\boldsymbol{x}) \geq (\leq)\ \phi(\boldsymbol{y}) \tag{3.7}$$

There are many characterizations of Schur-convex (concave) functions but the simplest one is described in the following Lemma.

Lemma 3.1 *The necessary and sufficient condition for a symmetric*[1] *function* $\phi(\boldsymbol{x})$ *to be Schur-convex(concave) is*

$$(x_1 - x_2)\left[\frac{\partial\phi(\boldsymbol{x})}{\partial x_1} - \frac{\partial\phi(\boldsymbol{x})}{\partial x_2}\right] \geq (\leq) 0 \tag{3.8}$$

Sometimes it will be convenient to use the following sufficient condition to test the Schur's convexity:

Lemma 3.2 *If* $g(x_k)$ *is convex then*

$$\phi(\boldsymbol{x}) = \sum_{i=1}^{M} g(x_k) \tag{3.9}$$

is Schur-convex. If $g(x_k)$ *is concave, then* $\phi(\boldsymbol{x})$ *is Schur-concave.*

Examples of Schur-convex (concave) functions:

1. The function $V(\boldsymbol{x}) = \sum_{k=1}^{M} x_k \ln(x_k)$ is Schur-convex function $\forall \boldsymbol{x} \in \mathbb{R}_+^M$. The condition in the Lemma (3.1) can be used to proof the Schur-convexity. Note first that $V(\boldsymbol{x})$ is symmetric under any arbitrary permutation of the input

[1]A function $\phi(x_1, x_2, \ldots, x_M)$ is symmetric if the argument vector $(x_1, x_2, \ldots, x_M)$ can be arbitrarily permuted without changing the value of the function.

vector $\boldsymbol{x}$. Now, the partial derivatives of $V(\boldsymbol{x})$ with respect to x_1 and x_2 are

$$\frac{\partial\, V(\boldsymbol{x})}{\partial x_1} = \ln(x_1) + 1 \quad , \quad \frac{\partial\, V(\boldsymbol{x})}{\partial x_2} = \ln(x_2) + 1$$

Thus, we can write

$$\begin{aligned}(x_1 - x_2)\left(\frac{\partial\, V(\boldsymbol{x})}{\partial x_1} - \frac{\partial\, V(\boldsymbol{x})}{\partial x_2}\right) &= (x_1 - x_2)\Big[\ln(x_1) + 1 - \ln(x_2) - 1\Big] \\ &= (x_1 - x_2)\Big[\ln(x_1) - \ln(x_2)\Big] \\ &= (x_1 - x_2)\ln\left(\frac{x_1}{x_2}\right) \geq 0\end{aligned} \tag{3.10}$$

This always true for $x_1 \geq x_2$ and the function $V(\boldsymbol{x})$ is a Schur's-convex function. For the following two vectors

$$\boldsymbol{x} = [10, 8, 4, 2, 1], \qquad \boldsymbol{y} = [10, 5, 5, 4, 1]$$

It can be easy to show that $\boldsymbol{x} \succ \boldsymbol{y}$. The value of $V(\boldsymbol{x})$ with these vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ will be

$$V(\boldsymbol{x}) = 46.5929, \qquad V(\boldsymbol{y}) = 44.6654$$

2. The function $F(\boldsymbol{x}) = \sin(x_1) + \sin(x_2) + \ldots + \sin(x_M)$ is Schur-concave on $[0, \pi]$. Here, the condition in Lemma (3.2) will be enough to proof the Schur-concavity of this function. This is because $F(\boldsymbol{x})$ is a sum of M identical and concave functions $g(x_k) = \sin(x_k)$ on $[0, \pi]$.

CHAPTER 4

MAJORIZATION PROPERTIES OF ADAPTIVE FILTERS

4.1 Motivation

As we saw in Section 2.3, the performance of adaptive filters can be described in the form of scalar measures (e.g., MSE, MSD and learning curves). These functions are affected by several parameters such as, the step size, the variance of the noise and the eigenvalues of the covariance matrix of the input signal $\boldsymbol{R_u}$.

In literature, there are many studies about the effect of the eigenvalues and the eigenvalues spread $\chi(\boldsymbol{R_u}) = \frac{\lambda_{max}}{\lambda_{min}}$ of the matrix $\boldsymbol{R_u}$ on the performance of adaptive filters, where λ_{max} and λ_{min} are the largest and the smallest eigenvalues of the matrix $\boldsymbol{R_u}$ respectively. In [8, p. 215], an experiment was done to test the effect of the eigenvalues spread on the performance of the Steepest Decent algorithm. They conclude from this experiment that the SD algorithm converges faster when $\chi(\boldsymbol{R_u})$

is small (λ_{max} and λ_{min} are almost equal). In contrast, when $\chi(\boldsymbol{R_u})$ increases (the input samples become more correlated) the convergence become slower. In [28–30], a study about the effect of normalization on the convergence rate and the eigenvalues spread for the NLMS by simplifying the matrices of moments. Analyzing these functions is the aim of this chapter. As explained in the last chapter, the majorization theory offers a method for analyzing the performance of adaptive filters with respect to the correlation scenario of the input signal. Any change in the correlation of the input signal will totally appear in the eigenvalues of the covariance matrix $\boldsymbol{R_u}$. In this work, each correlation scenario will be represented by $(1 \times M)$ vector as follows

$$\boldsymbol{\Lambda}_k = [\boldsymbol{\Lambda}_k(1)\ ,\ \boldsymbol{\Lambda}_k(2)\ ,\ \ldots\ ,\ \boldsymbol{\Lambda}_k(M)] \tag{4.1}$$

$$= [\lambda_1^k\ ,\ \lambda_2^k\ ,\ \ldots\ ,\ \lambda_M^k] \tag{4.2}$$

where all λ's are positive and arranged in a descending order $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$. For any two correlation scenarios, there is one vector for each one of them. In majorization, the comparison between them can be achieved. Moreover, these comparisons can be preserved through Schur's functions. As a result, it can be shown which correlation will lead to better performance.

For each adaptive filter, the performance measures such as the EMSE will be tested by Schur's conditions. In this chapter, the analysis for the following filters will be studied;

1. Steepest Descent Algorithm.

2. Newton's Algorithm.

3. Least-Mean-Square algorithm (LMS).

4.2 Majorization Properties of Steepest Descent

The learning curve for Steepest Descent from equation (2.52) is

$$
\begin{aligned}
J(\boldsymbol{w}_i) &= J_{min} + \sum_{k=1}^{M} \lambda_k (1 - \mu\lambda_k)^{2(i+1)} |\boldsymbol{x}_k(-1)|^2 \\
&= J_{min} + \mathcal{C} \sum_{k=1}^{M} g(\lambda_k)
\end{aligned}
\tag{4.3}
$$

where $g(\lambda_k) = \lambda_k(1 - \mu\lambda_k)^{2(i+1)}$. From (4.3), the learning curve for the SD is a sum of an M identical functions. According to the Schur's test in lemma (3.1), it will be enough if the convexity or concavity of $g(\lambda_k)$ is proved. The function $g(\lambda_k)$ will be convex if its second derivative is positive.

The second derivative of $g(\lambda_k)$ is

$$
g^{''}(\lambda_k) = 2\mu(i+1)(1 - \mu\lambda_k)^{2i}[(3 + 2i)\mu\lambda_k - 2] \tag{4.4}
$$

From (4.4) the function $g(\lambda_k)$ is convex if :

$$
\lambda_k \geq \frac{2}{(2i+3)\mu} \tag{4.5}
$$

or equivalently

$$\mu \geq \frac{2}{(2i+3)\lambda_k} \tag{4.6}$$

For $\mathbf{\Lambda}_k = [\lambda_1^k \geq \lambda_2^k \geq \ldots \geq \lambda_M^k]$ then :

$$\frac{2}{(2i+3)\lambda_M^k} \leq \mu \leq \frac{2}{\lambda_1^k} \tag{4.7}$$

The condition in (4.7) is sufficient but not a necessary condition for $J(w_i)$ to be Schur-convex function.

For $\mathbf{\Lambda}_1 \succeq \mathbf{\Lambda}_2 \succeq \ldots \succeq \mathbf{\Lambda}_N$ a sufficient condition for Schur-convexity is

$$\frac{2}{(2i+3)min\{\lambda_M^1, \lambda_M^2, \ldots, \lambda_M^N\}} \leq \mu \leq \frac{2}{\lambda_1^1} \tag{4.8}$$

It can be seen from (4.8) that the range of the step size μ at any time instant (i) is a subset of the range at the time instant $(i+1)$. Such as at $i = 0$ and at $i = 1$

$$\left\{\frac{2}{(3)min\{\lambda_M^1, \lambda_M^2, \ldots, \lambda_M^N\}} \leq \mu \leq \frac{2}{\lambda_1^1}\right\} \in \left\{\frac{2}{(5)min\{\lambda_M^1, \lambda_M^2, \ldots, \lambda_M^N\}} \leq \mu \leq \frac{2}{\lambda_1^1}\right\}$$

So it will be enough if the step size μ is selected from the range at time $i = 0$, but in some cases this condition will not be satisfied (due to the stability conditions of the SD algorithm), hence the need for going up to time $i = 1$ or more to satisfy this range or condition. If μ is selected from the suitable range (for example at i), the majorization will be satisfied for any time instant greater than or equal i.

Simulation results

Consider the following two vectors which represent two correlation scenarios:

$$\mathbf{\Lambda}_1 = [\mathbf{\Lambda}_1(1)\ ,\ \mathbf{\Lambda}_1(2)\ ,\ \mathbf{\Lambda}_1(3)\ ,\ \mathbf{\Lambda}_1(4)\ ,\ \mathbf{\Lambda}_1(5)] \tag{4.9}$$

$$= [37.7343\ ,\ 35.4682\ ,\ 32.3157\ ,\ 24.4882\ ,\ 22.2793] \tag{4.10}$$

$$\mathbf{\Lambda}_2 = [\mathbf{\Lambda}_2(1)\ ,\ \mathbf{\Lambda}_2(2)\ ,\ \mathbf{\Lambda}_2(3)\ ,\ \mathbf{\Lambda}_2(4)\ ,\ \mathbf{\Lambda}_2(5)] \tag{4.11}$$

$$= [34.0957\ ,\ 30.9876\ ,\ 29.6801\ ,\ 29.3734\ ,\ 28.1489] \tag{4.12}$$

It can be easily verified that $\mathbf{\Lambda}_1 \succ \mathbf{\Lambda}_2$, by applying the definition 3.1 which calculates the partial sums

$$SUM_{\mathbf{\Lambda}_1} = \left[\sum_{k=1}^{1}\mathbf{\Lambda}_1(k)\ ,\ \sum_{k=1}^{2}\mathbf{\Lambda}_1(k)\ ,\ \sum_{k=1}^{3}\mathbf{\Lambda}_1(k)\ ,\ \sum_{k=1}^{4}\mathbf{\Lambda}_1(k)\ ,\ \sum_{k=1}^{5}\mathbf{\Lambda}_1(k)\right] \tag{4.13}$$

$$= [37.7343\ ,\ 73.2026\ ,\ 105.5182\ ,\ 130.0064\ ,\ 152.2858] \tag{4.14}$$

$$SUM_{\mathbf{\Lambda}_2} = \left[\sum_{k=1}^{1}\mathbf{\Lambda}_2(k)\ ,\ \sum_{k=1}^{2}\mathbf{\Lambda}_2(k)\ ,\ \sum_{k=1}^{3}\mathbf{\Lambda}_2(k)\ ,\ \sum_{k=1}^{4}\mathbf{\Lambda}_2(k)\ ,\ \sum_{k=1}^{5}\mathbf{\Lambda}_2(k)\right] \tag{4.15}$$

$$= [34.0957\ ,\ 65.0833\ ,\ 94.7635\ ,\ 124.1369\ ,\ 152.2858] \tag{4.16}$$

By comparing the entries in (4.13) with (4.16) it is easy to see that $\mathbf{\Lambda}_1 \succ \mathbf{\Lambda}_2$.

If the step size μ is selected according to the condition in (4.8) with $i = 0$ as

$$\frac{2}{(3) \times 22.2793} \leq \mu \leq \frac{2}{37.7343}$$

$$0.0299 \leq \mu \leq 0.0530$$

Then $J(\boldsymbol{w}_i, \boldsymbol{\Lambda}_1) \geq J(\boldsymbol{w}_i, \boldsymbol{\Lambda}_2)\ \forall\ i$. This is demonstrated in Figure 4.1 which shows that the MSE learning curves for two scenarios with $\boldsymbol{\Lambda}_1 \succ \boldsymbol{\Lambda}_2$.

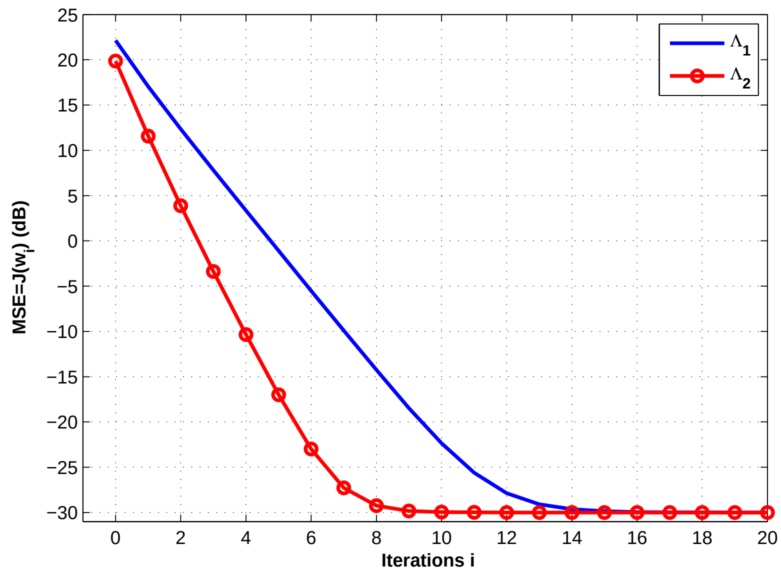

Figure 4.1: Learning curves for Steepest Descent with ($\boldsymbol{\Lambda}_1 \succ \boldsymbol{\Lambda}_2$)

4.3 Majorization Properties of Newton's Algorithm

The learning curve for Newton Method's Algorithm from equation (2.53) is

$$J(\boldsymbol{w}_i) = E|d - \boldsymbol{u}\boldsymbol{w}_i|^2 \tag{4.17}$$

$$= J_{min} + \sum_{k=1}^{M} \lambda_k |x_k(i)|^2 \tag{4.18}$$

$$= J_{min} + \mathcal{C}(1-\mu)^{2(i+1)} \sum_{k=1}^{M} \lambda_k \tag{4.19}$$

From (4.19) it can be shown that the learning curve depends on the sum of the whole vector ($\sum_{k=1}^{M} \lambda_k$), and it will be both Schur-convex and Schur-concave as concluded in [27] and [26]. Thus, $J(\boldsymbol{w}_i, \boldsymbol{\Lambda}_m) = J(\boldsymbol{w}_i, \boldsymbol{\Lambda}_n)$ for any $\boldsymbol{\Lambda}_m \succ \boldsymbol{\Lambda}_n$. This is demonstrated for the two extreme cases $\boldsymbol{\Lambda}_1 = [M\ ,\ 0\ ,\ 0\ ,\ \ldots\ ,\ 0]$ and $\boldsymbol{\Lambda}_1 = [1\ ,\ 1\ ,\ \ldots\ ,\ 1]$ in Figure 4.2.

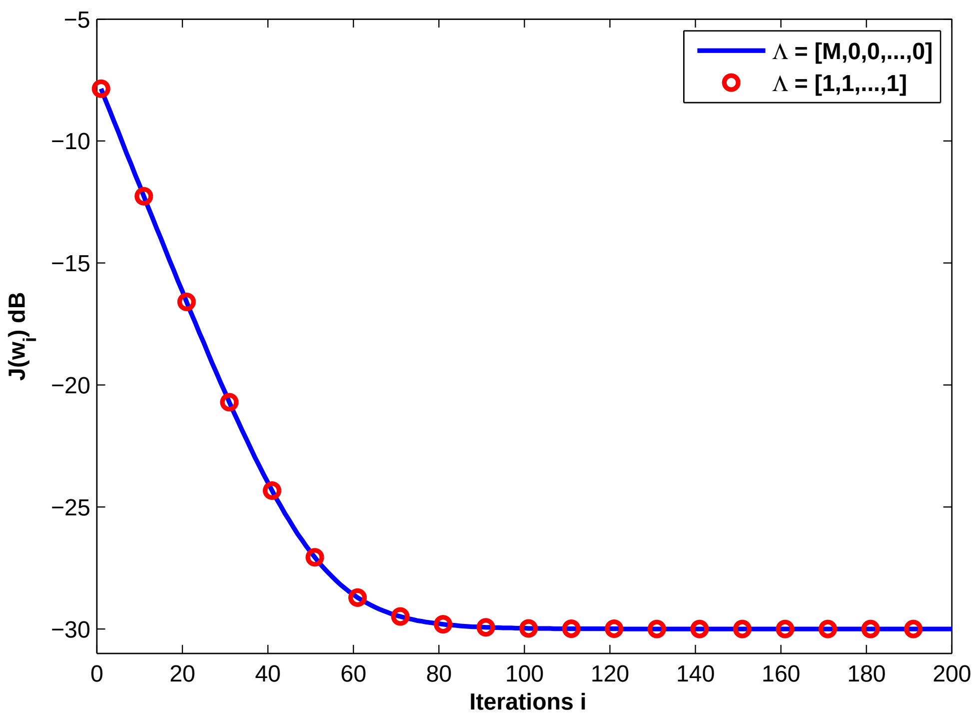

Figure 4.2: Learning curves for Newton's Method for ($\boldsymbol{\Lambda}_1 \succ \boldsymbol{\Lambda}_2$)

4.4 Majorization Properties of Least Mean Square (LMS)

In this section we study the majorization properties of the LMS filter in steady state.

4.4.1 Steady States (EMSE)

From equation (2.78) the steady state EMSE is fully described by the following expression [1] :

$$
\begin{aligned}
EMSE &= \frac{\sigma_v^2 \mu \sum_{k=1}^{M} \frac{\lambda_k}{2 - \mu\lambda_k}}{1 - \mu \sum_{k=1}^{M} \frac{\lambda_k}{2 - \mu\lambda_k}} \\
&= \frac{\sigma_v^2 \mu \ \mathcal{S}}{1 - \mu \ \mathcal{S}}
\end{aligned}
\tag{4.20}
$$

where $\mathcal{S} = \sum_{k=1}^{M} \frac{\lambda_k}{2 - \mu\lambda_k}$.

Applying the test in (3.8) on the EMSE to check the Schur's convexity as follows:

Partial derivative of the numerator of (4.20)

$$
\frac{\partial}{\partial \lambda_1} \left[\sigma_v^2 \mu \sum_{k=1}^{M} \frac{\lambda_k}{2 - \mu\lambda_k} \right] = \left[\frac{2\sigma_v^2 \mu}{(2 - \mu\lambda_1)^2} \right] \tag{4.21}
$$

Partial derivative of the denominator of (4.20)

$$\frac{\partial}{\partial \lambda_1}\left[1-\mu\sum_{k=1}^{M}\frac{\lambda_k}{2-\mu\lambda_k}\right]=\left[\frac{-2\mu}{\left(2-\mu\lambda_1\right)^2}\right] \tag{4.22}$$

Then:

$$\begin{aligned}\frac{\partial EMSE}{\partial \lambda_1} &= \left[\frac{\left(1-\mu\mathcal{S}\right)\frac{2\sigma_v^2\mu}{(2-\mu\lambda_1)^2}-\left(\sigma_v^2\mu\mathcal{S}\right)\frac{-2\mu}{(2-\mu\lambda_1)^2}}{(1-\mu\mathcal{S})^2}\right]\\ &= \frac{2\sigma_v^2\mu}{\left(2-\mu\lambda_1\right)^2}\left[\frac{\left(1-\mu\mathcal{S}+\mu\mathcal{S}\right)}{(1-\mu\mathcal{S})^2}\right]\\ &= \left[\frac{\frac{2\sigma_v^2\mu}{(2-\mu\lambda_1)^2}}{(1-\mu\mathcal{S})^2}\right]\end{aligned} \tag{4.23}$$

and

$$\frac{\partial EMSE}{\partial \lambda_2}=\left[\frac{\frac{2\sigma_v^2\mu}{(2-\mu\lambda_2)^2}}{(1-\mu\mathcal{S})^2}\right] \tag{4.24}$$

Using these results and applying the Schur's test in (3.8) with $\lambda_1 \geq \lambda_2$ yields:

$$\begin{aligned}(\lambda_1-\lambda_2)\Big(\frac{\partial EMSE}{\partial \lambda_1}-\frac{\partial EMSE}{\partial \lambda_2}\Big) &= (\lambda_1-\lambda_2)\left[\frac{\frac{2\sigma_v^2\mu}{(2-\mu\lambda_1)^2}}{\left(1-\mu\mathcal{S}\right)^2}-\frac{\frac{2\sigma_v^2\mu}{(2-\mu\lambda_2)^2}}{\left(1-\mu\mathcal{S}\right)^2}\right]\\ &= \frac{2\sigma_v^2\mu(\lambda_1-\lambda_2)}{\left(1-\mu\mathcal{S}\right)^2}\left[\frac{1}{\left(2-\mu\lambda_1\right)^2}-\frac{1}{\left(2-\mu\lambda_2\right)^2}\right]\end{aligned} \tag{4.25}$$

Now investigate sign of the RHS in equation (4.25) as

$$\frac{2\sigma_v^2\mu(\lambda_1-\lambda_2)}{\left(1-\mu\mathcal{S}\right)^2}\left[\frac{1}{\left(2-\mu\lambda_1\right)^2}-\frac{1}{\left(2-\mu\lambda_2\right)^2}\right]\gtreqless 0 \tag{4.26}$$

$$(\lambda_1 - \lambda_2)\left[\frac{1}{(2-\mu\lambda_1)^2} - \frac{1}{(2-\mu\lambda_2)^2}\right] \gtreqless 0 \tag{4.27}$$

Since $\lambda_1 \geq \lambda_2$ the LHS in (4.27) is always positive. From this result and by utilizing the condition in (3.8), then the $EMSE$ of the LMS is a Schur-convex function for any step size μ. Thus, for $\mathbf{\Lambda}_1 \succeq \mathbf{\Lambda}_2$ the performance can be ordered as

$$EMSE(\mathbf{\Lambda}_1) \geq EMSE(\mathbf{\Lambda}_2)$$

4.4.2 Steady States (MSD)

From equation (2.76) the steady state MSD is given by:

$$\begin{aligned} MSD &= \frac{\sigma_v^2 \mu \sum_{k=1}^{M} \frac{1}{2-\mu\lambda_k}}{1 - \mu \displaystyle\sum_{k=1}^{M} \frac{\lambda_k}{2-\mu\lambda_k}} \\ &= \frac{\sigma_v^2 \mu \sum_{k=1}^{M} \frac{1}{2-\mu\lambda_k}}{1 - \mu\ \mathcal{S}} \end{aligned} \tag{4.28}$$

where $\mathcal{S} = \displaystyle\sum_{k=1}^{M} \frac{\lambda_k}{2-\mu\lambda_k}$.
Following the same steps that used in proving the Schur's convexity for the EMSE.

Partial derivative of the numerator of (4.28)

$$\frac{\partial}{\partial\lambda_1}\left[\sigma_v^2 \mu \sum_{k=1}^{M} \frac{1}{2-\mu\lambda_k}\right] = \left[\frac{\sigma_v^2 \mu^2}{(2-\mu\lambda_1)^2}\right] \tag{4.29}$$

Partial derivative of the denominator of (4.28)

$$\frac{\partial}{\partial \lambda_1}\left[1-\mu\sum_{k=1}^{M}\frac{\lambda_k}{2-\mu\lambda_k}\right]=\left[\frac{-2\mu}{(2-\mu\lambda_1)^2}\right] \tag{4.30}$$

Then:

$$\begin{aligned}
\frac{\partial MSD}{\partial \lambda_1} &= \left[\frac{\frac{\sigma_v^2\mu^2}{(2-\mu\lambda_1)^2}\big(1-\mu\mathcal{S}\big)+\frac{2\sigma_v^2\mu^2}{(2-\mu\lambda_1)^2}\sum_{k=1}^{M}\frac{1}{2-\mu\lambda_k}}{(1-\mu\mathcal{S})^2}\right] \\
&= \frac{\sigma_v^2\mu^2}{(2-\mu\lambda_1)^2}\left[\frac{1-\mu\sum_{k=1}^{M}\frac{\lambda_k}{2-\mu\lambda_k}+2\sum_{k=1}^{M}\frac{1}{2-\mu\lambda_k}}{(1-\mu\mathcal{S})^2}\right] \\
&= \frac{\sigma_v^2\mu^2}{(2-\mu\lambda_1)^2(1-\mu\mathcal{S})^2}\left[1+\sum_{k=1}^{M}\frac{2-\mu\lambda_k}{2-\mu\lambda_k}\right] \\
&= \frac{\sigma_v^2\mu^2(1+M)}{(2-\mu\lambda_1)^2(1-\mu\mathcal{S})^2}
\end{aligned} \tag{4.31}$$

and

$$\frac{\partial MSD}{\partial \lambda_2}=\frac{\sigma_v^2\mu^2(1+M)}{(2-\mu\lambda_2)^2(1-\mu\mathcal{S})^2} \tag{4.32}$$

Using these results and applying the Schur's test in (3.8) yields:

$$(\lambda_1-\lambda_2)\Big(\frac{\partial MSD}{\partial \lambda_1}-\frac{\partial MSD}{\partial \lambda_2}\Big)=(\lambda_1-\lambda_2)\left[\frac{\sigma_v^2\mu^2(1+M)}{(2-\mu\lambda_1)^2(1-\mu\mathcal{S})^2}-\frac{\sigma_v^2\mu^2(1+M)}{(2-\mu\lambda_2)^2(1-\mu\mathcal{S})^2}\right] \tag{4.33}$$

$$=\frac{\sigma_v^2\mu^2(1+M)(\lambda_1-\lambda_2)}{\big(1-\mu\mathcal{S})^2}\left[\frac{1}{(2-\mu\lambda_1)^2}-\frac{1}{(2-\mu\lambda_2)^2}\right] \tag{4.34}$$

Now investigate sign of the RHS in equation (4.33) as

$$\frac{\sigma_v^2\mu^2(1+M)(\lambda_1-\lambda_2)}{(1-\mu\mathcal{S})^2}\left[\frac{1}{(2-\mu\lambda_1)^2}-\frac{1}{(2-\mu\lambda_2)^2}\right]\gtreqless 0 \tag{4.35}$$

$$(\lambda_1-\lambda_2)\left[\frac{1}{(2-\mu\lambda_1)^2}-\frac{1}{(2-\mu\lambda_2)^2}\right]\gtreqless 0 \tag{4.36}$$

Since $\lambda_1 \geq \lambda_2$ the LHS in (4.36) is always positive. The same conclusion of the EMSE case can be drawn for the MSD case. The MSD of the LMS is a Schur-convex function for any step size μ. For $\mathbf{\Lambda}_1 \succeq \mathbf{\Lambda}_2$ the performance can be ordered as

$$MSD(\mathbf{\Lambda}_1) \geq MSD(\mathbf{\Lambda}_2)$$

Example 1 (Numerical illustration): The following six vectors represent the sets of the eigenvalues for six different inputs to the LMS filter with different correlations scenarios:

$$\mathbf{\Lambda}_1 = \left[\ 5\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ 0\ \right] \tag{4.37}$$

$$\mathbf{\Lambda}_2 = \left[\ 4.8889\ ,\ 0.1000\ ,\ 0.0100\ ,\ 0.0010\ ,\ 0.0001\ \right] \tag{4.38}$$

$$\mathbf{\Lambda}_3 = \left[\ 3.6000\ ,\ 0.8900\ ,\ 0.3000\ ,\ 0.2000\ ,\ 0.0100\ \right] \tag{4.39}$$

$$\mathbf{\Lambda}_4 = \left[\ 2.8000\ ,\ 1.1000\ ,\ 0.6000\ ,\ 0.4500\ ,\ 0.0500\ \right] \tag{4.40}$$

$$\mathbf{\Lambda}_5 = \left[\ 1.2090\ ,\ 1.0910\ ,\ 1.0000\ ,\ 0.9000\ ,\ 0.8000\ \right] \tag{4.41}$$

$$\mathbf{\Lambda}_6 = \left[\ 1\ ,\ 1\ ,\ 1\ ,\ 1\ ,\ 1\ \right] \tag{4.42}$$

In $\mathbf{\Lambda}_1$ we have the worst case (highly correlated signal), while $\mathbf{\Lambda}_6$ represents the best case (white signal). It can be shown that

$$\mathbf{\Lambda}_1 \succ \mathbf{\Lambda}_2 \succ \mathbf{\Lambda}_3 \succ \mathbf{\Lambda}_4 \succ \mathbf{\Lambda}_5 \succ \mathbf{\Lambda}_6$$

In Table 4.1 the values of the steady state EMSE and MSD of the LMS filter for each scenario with two values of the step size μ. These two values of μ are chosed acoording to the following condition for the LMS filter [1].

$$\sum_{k=1}^{M} \frac{\lambda_k \mu}{2 - \lambda_k \mu} < 1 \tag{4.43}$$

This table shows that the order of majorization is preserved as

$$EMSE(\mathbf{\Lambda}_1) \geq EMSE(\mathbf{\Lambda}_2) \geq EMSE(\mathbf{\Lambda}_3) \geq EMSE(\mathbf{\Lambda}_4) \geq EMSE(\mathbf{\Lambda}_5) \geq EMSE(\mathbf{\Lambda}_6)$$

and

$$MSD(\mathbf{\Lambda}_1) \geq MSD(\mathbf{\Lambda}_2) \geq MSD(\mathbf{\Lambda}_3) \geq MSD(\mathbf{\Lambda}_4) \geq MSD(\mathbf{\Lambda}_5) \geq MSD(\mathbf{\Lambda}_6)$$

Table 4.1: EMSE and MSD for the LMS with different scenarios

	$EMSE(dB)$		$MSD(dB)$	
$\boldsymbol{\Lambda}_k$	$\mu = 0.005$	$\mu = 0.1$	$\mu = 0.005$	$\mu = 0.1$
$\boldsymbol{\Lambda}_1$	-48.9209	-33.0103	-48.9209	-33.0103
$\boldsymbol{\Lambda}_2$	-48.9234	-33.0931	-48.9209	-33.0103
$\boldsymbol{\Lambda}_3$	-48.9456	-33.8483	-48.9209	-33.0103
$\boldsymbol{\Lambda}_4$	-48.9550	-34.1570	-48.9209	-33.0103
$\boldsymbol{\Lambda}_5$	-48.9650	-34.4649	-48.9209	-33.0103
$\boldsymbol{\Lambda}_6$	-48.9653	-34.4716	-48.9209	-33.0103

Example 2 (Eigenvalues Spread): The following five vectors represent the sets of the eigenvalues for five different scenarios. Here in this example the five vectors have the same eigenvalues spread $\chi(\boldsymbol{R_u}) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{30}{1} = 30$ and $\boldsymbol{\Lambda}_1 \succ \boldsymbol{\Lambda}_2 \succ \boldsymbol{\Lambda}_3 \succ \boldsymbol{\Lambda}_4 \succ \boldsymbol{\Lambda}_5$:

$$\boldsymbol{\Lambda}_1 = \left[\ 30.0\ ,\ 30.0\ ,\ 17.0\ ,\ 1.00\ ,\ 1.00\ ,\ 1.00\ ,\ 1.00\ \right] \tag{4.44}$$

$$\boldsymbol{\Lambda}_2 = \left[\ 30.0\ ,\ 30.0\ ,\ 5.00\ ,\ 5.00\ ,\ 5.00\ ,\ 5.00\ ,\ 1.00\ \right] \tag{4.45}$$

$$\boldsymbol{\Lambda}_3 = \left[\ 30.0\ ,\ 20.0\ ,\ 15.0\ ,\ 5.00\ ,\ 5.00\ ,\ 5.00\ ,\ 1.00\ \right] \tag{4.46}$$

$$\boldsymbol{\Lambda}_4 = \left[\ 30.0\ ,\ 15.0\ ,\ 15.0\ ,\ 10.0\ ,\ 5.00\ ,\ 5.00\ ,\ 1.00\ \right] \tag{4.47}$$

$$\boldsymbol{\Lambda}_5 = \left[\ 30.0\ ,\ 10.0\ ,\ 10.0\ ,\ 10.0\ ,\ 10.0\ ,\ 10.0\ ,\ 1.00\ \right] \tag{4.48}$$

In Table 4.2 the values of the steady state EMSE and MSD of the LMS filter for each scenario.

	$EMSE(dB)$		$MSD(dB)$	
$\boldsymbol{\Lambda}_k$	$\mu = 0.005$	$\mu = 0.01$	$\mu = 0.005$	$\mu = 0.01$
$\boldsymbol{\Lambda}_1$	-35.1751	-29.1077	-46.1379	-40.4449
$\boldsymbol{\Lambda}_2$	-35.2411	-29.3231	-46.1563	-40.5755
$\boldsymbol{\Lambda}_3$	-35.3636	-29.7999	-46.1900	-40.8554
$\boldsymbol{\Lambda}_4$	-35.4015	-29.9306	-46.2003	-40.9299
$\boldsymbol{\Lambda}_5$	-35.4381	-30.0501	-46.2101	-40.9972

Table 4.2: EMSE and MSD with different scenarios with same eigenvalues spread $\chi(\boldsymbol{R_u})$

From this example, all $\boldsymbol{\Lambda}'$s with the same eigenvalues spread but with different performances. Real indication of which input to the adaptive filter result in better performance is which vector is majorizes the other.

4.5 How general is the relation between Majorization and adaptive filters ?

In this work, we investigate the majorization properties for the Steepest Decent algorithm, Newton's algorithm and steady state performance for the LMS filter. But how general is this?

In the following subsections we will investigate the majorization properties for the transient performance of the LMS filter and steady state NLMS by simulation only. The work on the mathematical proof will be part of the future work in this area.

4.5.1 Transient behavior of LMS

Conducting simulation on the LMS filter yields the MSD learning curves in Figure 4.3. From this figure, the ordering between the learning curves of the LMS filter coincides with the majorization order between the input vectors, i.e., if

$$\mathbf{\Lambda}_1 \succ \mathbf{\Lambda}_2 \succ \mathbf{\Lambda}_3 \succ \mathbf{\Lambda}_4$$

then

$$MSD(i, \mathbf{\Lambda}_1) \geq MSD(i, \mathbf{\Lambda}_2) \geq MSD(i, \mathbf{\Lambda}_3) \geq MSD(i, \mathbf{\Lambda}_4)$$

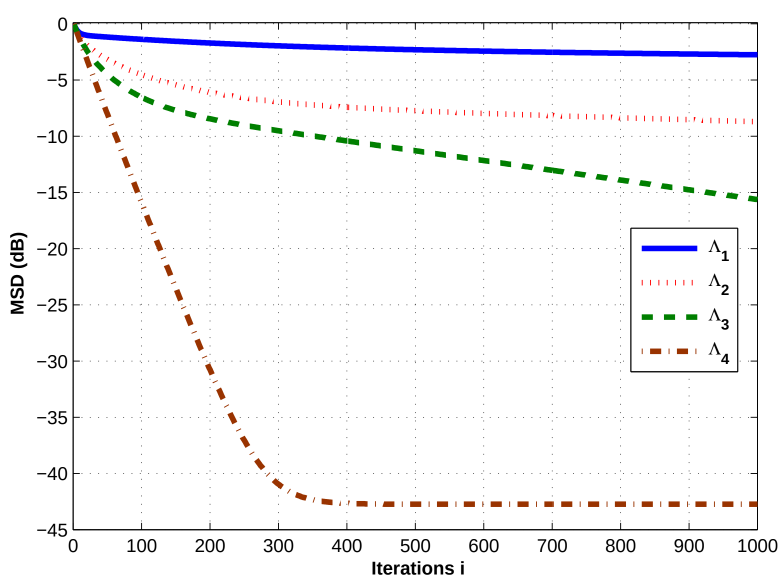

Figure 4.3: Learning Curve for LMS filter with ($\mathbf{\Lambda}_1 \succ \mathbf{\Lambda}_2 \succ \mathbf{\Lambda}_3 \succ \mathbf{\Lambda}_4$ and $\mu = 0.005$)

4.5.2 Steady state EMSE of the NLMS

Studying the majorization properties of the NLMS is not an easy task because applying the majorization tests on the expression of this filter is difficult even for the approximated versions.

In [1] an approximation of the EMSE of the NLMS filter is given by

$$EMSE = \frac{\mu\sigma_v^2}{2-\mu} Tr(\boldsymbol{R_u}) E\left(\frac{1}{\|\boldsymbol{u}\|^2}\right) \tag{4.49}$$

$$= \frac{\mu\sigma_v^2}{2-\mu} \sum_{k=1}^{M} \lambda_k E\left(\frac{1}{\|\boldsymbol{u}\|^2}\right) \tag{4.50}$$

where the moment $E\left(\frac{1}{\|\boldsymbol{u}\|^2}\right)$ is calculated in [31] and given by

$$E\left(\frac{1}{\|\boldsymbol{u}\|^2}\right) = \sum_{m=1}^{M} \frac{\lambda_m^{M-1}\ln(\lambda_m)}{|\boldsymbol{\Lambda}| \prod_{i=1, i\neq m}^{M} (\frac{\lambda_m}{\lambda_i} - 1)} \tag{4.51}$$

using the value of this moment in (4.49) yields

$$EMSE = \frac{\mu\sigma_v^2}{2-\mu} \sum_{k=1}^{M} \lambda_k \sum_{m=1}^{M} \frac{\lambda_m^{M-1}\ln(\lambda_m)}{|\boldsymbol{\Lambda}| \prod_{i=1, i\neq m}^{M} (\frac{\lambda_m}{\lambda_i} - 1)} \tag{4.52}$$

Calculating the EMSE of NLMS in (4.52) with the following five vectors

$$\boldsymbol{\Lambda}_1 = \left[\, 30.0\ ,\ 30.0\ ,\ 17.0\ ,\ 1.00\ ,\ 1.00\ ,\ 1.00\ ,\ 1.00\, \right] \tag{4.53}$$

$$\boldsymbol{\Lambda}_2 = \left[\, 30.0\ ,\ 30.0\ ,\ 5.00\ ,\ 5.00\ ,\ 5.00\ ,\ 5.00\ ,\ 1.00\, \right] \tag{4.54}$$

$$\boldsymbol{\Lambda}_3 = \left[\, 30.0\ ,\ 20.0\ ,\ 15.0\ ,\ 5.00\ ,\ 5.00\ ,\ 5.00\ ,\ 1.00\, \right] \tag{4.55}$$

$$\boldsymbol{\Lambda}_4 = \left[\, 30.0\ ,\ 15.0\ ,\ 15.0\ ,\ 10.0\ ,\ 5.00\ ,\ 5.00\ ,\ 1.00\, \right] \tag{4.56}$$

$$\boldsymbol{\Lambda}_5 = \left[\, 30.0\ ,\ 10.0\ ,\ 10.0\ ,\ 10.0\ ,\ 10.0\ ,\ 10.0\ ,\ 1.00\, \right] \tag{4.57}$$

The values of the EMSE are listed in Table 4.3. The same observation

$\boldsymbol{\Lambda}_k$	$EMSE(dB)$
$\boldsymbol{\Lambda}_1$	-42.7632
$\boldsymbol{\Lambda}_1$	-43.3351
$\boldsymbol{\Lambda}_1$	-43.7685
$\boldsymbol{\Lambda}_1$	-43.8716
$\boldsymbol{\Lambda}_1$	-43.9471

Table 4.3: EMSE of NLMS with $\boldsymbol{\Lambda}_1 \succ \boldsymbol{\Lambda}_2 \succ \boldsymbol{\Lambda}_3 \succ \boldsymbol{\Lambda}_4 \succ \boldsymbol{\Lambda}_5$

From this table we can see that the majorization theory is applicable also for the NLMS.

CHAPTER 5

PERFORMANCE ANALYSIS OF THE RECURSIVE LEAST SQUARES (RLS) FILTER

The RLS is one of the important algorithms from the adaptive filter's family. Motivation for the RLS adaptive filters relies on the fact that it provides a solution to the least square error minimization problem in (5.1). The RLS algorithms are more costly than the basic families such as the Leat mean squares family (LMS) but with much faster convergence speed. Analyzing the performance of the RLS algorithm is not an easy task due to the presence of the input covariance matrix and its inverse which depends on current and past input regressors. As such, only a few works considered the performance of the RLS and its variants [1, 8, 32–40]. The simplest approach is based on the energy relation [1, 41] which (with the aid of separation principle [1]) can be used to state that the EMSE is a function of the moment $E[||\boldsymbol{u}_i||^2_{\boldsymbol{P}_i}]$, where $\boldsymbol{u}_i$ is the input regressor and $\mathbf{P}_i$ is the estimate of the inverse of input covariance matrix. Another separation principle is then used to write $E[||\boldsymbol{u}_i||^2_{\boldsymbol{P}_i}] = Tr(E[\boldsymbol{u}_i^* \boldsymbol{u}_i]\boldsymbol{P}_i)$. But this approach is not rigorous as $\boldsymbol{u}_i$ and $\boldsymbol{P}_i$ are dependent. Other approaches use the idea of random matrix to study the performance of the RLS [38,39]. In addition to requiring much more sophisticated machinery, these approaches are valid for filters relatively larger sizes.

In this work, a new approach for studying the steady state performance of the Recursive Least Square (RLS) adaptive filter for a circularly correlated Gaussian input is presented. The mean-square analysis of the RLS relies on the moment of the random variable $\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}$, where $\boldsymbol{P}_i$ is the estimate of the inverse of input covariance matrix. Earlier approaches evaluate this moment by assuming that the $\boldsymbol{u}_i$ and $\boldsymbol{P}_i$ are independent which could result in negative value of the steady state Excess Mean Square Error (EMSE). In this work, this assumption is avoided and a closed form expression for this moment is derived. This derivation is based on finding the cumulative distribution function (CDF) of the random variable of the form $\frac{1}{\gamma+||\boldsymbol{u}||^2_D}$, where $\boldsymbol{u}$ is circular correlated Gaussian input and $\mathbf{D}$ is a diagonal matrix. As a result, more accurate estimation of the EMSE of the RLS filter is obtained. Simulation results corroborate the analytical findings.

5.1 Least Squares Algorithms

Given an $(i+1 \times 1)$ measurements vector $\boldsymbol{y}_i$, and an $(i+1 \times M)$ data matrix $\boldsymbol{H}_i$ where

$$\boldsymbol{y}_i = \begin{bmatrix} d(0) \\ d(1) \\ \vdots \\ d(i) \end{bmatrix} \quad , \quad \boldsymbol{H}_i = \begin{bmatrix} \boldsymbol{u}_0 \\ \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_i \end{bmatrix}$$

as mentioned in table 2.1 the idea of the least squares algorithms is to minimize the following estimation problem

$$\underset{w}{min} \|\boldsymbol{y}_i - \boldsymbol{H}_i \boldsymbol{w}\|^2 \tag{5.1}$$

This minimization problem can be considered as a minimization of the average of the error signal $e(i) = d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}$, this idea is explained in the following relation

$$E|\boldsymbol{d} - \boldsymbol{u}\boldsymbol{w}|^2 \approx \frac{1}{N} \sum_{j=0}^{i-1} |d(j) - \boldsymbol{u}_j \boldsymbol{w}_{j-1}|^2 = \|\boldsymbol{y}_i - \boldsymbol{H}_i \boldsymbol{w}\|^2 \tag{5.2}$$

5.2 Exponentially Weighted Recursive Least Squares (RLS)

Instead of the problem in (5.1), the exponentially weighted RLS attempts to solve the following problem

$$\min_{\boldsymbol{w}}\left[\gamma^{(i+1)}\boldsymbol{w}^{*}\Pi\boldsymbol{w}+(\boldsymbol{y}_i-\boldsymbol{H}_i\boldsymbol{w})^{*}\Gamma_i(\boldsymbol{y}_i-\boldsymbol{H}_i\boldsymbol{w})\right] \tag{5.3}$$

where $\Gamma_i = diag\{\gamma^i\ ,\ \gamma^{i-1}\ ,\ \ldots\ ,\ \gamma\ ,\ 1\}$ is an $(i+1 \times i+1)$ is a diagonal weighting matrix defined in terms of the forgetting factor γ $(0 \ll \gamma \leq 1)$, and Π is an $(M \times M)$ positive definite matrix. It can be shown that the solution of (5.3) is given by

$$\boldsymbol{w}_i = \boldsymbol{P}_i\boldsymbol{H}_i^{*}\Gamma_i\boldsymbol{y}_i \tag{5.4}$$

The RLS allows to obtain the solution of (5.3) in a recursive manner. Specifically,

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1}+\boldsymbol{P}_i\boldsymbol{u}_i^{*}[d(i)-\boldsymbol{u}_i\boldsymbol{w}_{i-1}] \tag{5.5}$$

where

$$\boldsymbol{P}_i = \left[\gamma^{(i+1)}\Pi+\boldsymbol{H}_i^{*}\Gamma_i\boldsymbol{H}_i\right]^{-1} \tag{5.6}$$

inverting both sides in (5.6) yields

$$\boldsymbol{P}_i^{-1} = \gamma^{(i+1)}\Pi + \boldsymbol{H}_i^*\Gamma_i\boldsymbol{H}_i \tag{5.7}$$

$$= \gamma^{i+1}\Pi + \gamma\boldsymbol{H}_{i-1}^*\Gamma_{i-1}\boldsymbol{H}_{i-1} + \boldsymbol{u}_i^*\boldsymbol{u}_i \tag{5.8}$$

$$= \gamma\big(\gamma^i\Pi + \boldsymbol{H}_{i-1}^*\Gamma_{i-1}\boldsymbol{H}_{i-1}\big) + \boldsymbol{u}_i^*\boldsymbol{u}_i \tag{5.9}$$

$$= \gamma\boldsymbol{P}_{i-1}^{-1} + \boldsymbol{u}_i^*\boldsymbol{u}_i \tag{5.10}$$

by applying the following matrix inversion identity on (5.10) to find $\boldsymbol{P}_i$

$$(\boldsymbol{A} + \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{DA}^{-1} \tag{5.11}$$

the result will be

$$\boldsymbol{P}_i = \gamma^{-1}\left[\boldsymbol{P}_{i-1} - \frac{\boldsymbol{P}_{i-1}\boldsymbol{u}_i^*\boldsymbol{u}_i\boldsymbol{P}_{i-1}}{\gamma + \boldsymbol{u}_i\boldsymbol{P}_{i-1}\boldsymbol{u}_i^*}\right] \tag{5.12}$$

5.3 Steady State Using The Energy Relation

In the system identification model, the measurement $d(i)$ takes the form

$$d(i) = \boldsymbol{u}_i^*\boldsymbol{w}^o + v(i) \tag{5.13}$$

where $v(i)$ is an additive noise and $\boldsymbol{w}^o$ is the system coefficients. The update recursion in (5.5) can be rewritten in terms of weight error vector $\tilde{\boldsymbol{w}}_i = \boldsymbol{w}^o - \boldsymbol{w}_i$

as

$$\tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i-1} - \boldsymbol{P}_i \boldsymbol{u}_i^* e(i) \tag{5.14}$$

where $e(i) = d(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}$ which is called the estimation error, to other errors are the a priori and a posteriori estimation errors defined by, respectively, $e_a(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1}$ and $e_p(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_i$. Multiply both sides of (5.14) by $\boldsymbol{u}_i$ from the left the result will be

$$e_p(i) = e_a(i) - \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \tag{5.15}$$

Next, combining (5.14) and (5.15) will led to

$$\tilde{\boldsymbol{w}}_i + \frac{\boldsymbol{P}_i \boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} e_a(i) = \tilde{\boldsymbol{w}}_{i-1} + \frac{\boldsymbol{P}_i \boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} e_p(i) \tag{5.16}$$

evaluating the energies of both sides will led to the well known **The Energy Conservation Relation** [1] given by

$$\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} + \frac{|e_a(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}} + \frac{|e_p(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.17}$$

By taking the expectation of both sides in (5.17), yields the following result in steady state (i.e. as $i \to \infty$)

$$E\frac{|e_a(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} = E\frac{|e_p(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.18}$$

where the following fact in steady state is used $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}}$.

Using (5.15) and the fact that $e(i) = e_a(i) + v(i)$ yields to **the variance relation** (as $i \to \infty$)

$$\sigma_v^2 E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} + E\Big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \;\cdot\; |e_a(i)|^2\Big) = 2E|e_a(i)|^2 \tag{5.19}$$

This relation can be used to evaluate $E|e_a(i)|^2$. To do so however, the use of the separation condition is necessary to separate the moment $E\Big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \;\cdot\; |e_a(i)|^2\Big)$. This is usually done as

$$E\Big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \;\cdot\; |e_a(i)|^2\Big) \approx E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}\; E|e_a(i)|^2 \tag{5.20}$$

Substituting this in (5.19) and solving for $E|e_a(i)|^2$ yields

$$\lim_{i\to\infty} E|e_a(i)|^2 = \frac{\sigma_v^2 \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}}{2 - \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.21}$$

From (5.21), it is clear that to evaluate the EMSE, the limit $\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}$ must be evaluated.

The common approach in literature [1,8] is to calculate the limit $\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}$ is done by assuming that $\boldsymbol{P}_i$ and $\boldsymbol{u}_i$ are independent and the value of this limit will be

$$\lim_{i\to\infty} E[||\boldsymbol{u}_i||^2_{\boldsymbol{P}_i}] = Tr(E[\boldsymbol{u}_i^*\boldsymbol{u}_i]\boldsymbol{P}_i) \tag{5.22}$$

$$= Tr(\boldsymbol{R_u P}) \tag{5.23}$$

and the EMSE will be

$$\lim_{i\to\infty} E|e_a(i)|^2 = \frac{\sigma_v^2 Tr(\boldsymbol{R_u P})}{2 - Tr(\boldsymbol{R_u P})} \tag{5.24}$$

this expression of the EMSE will give negative values for wide range of the forgetting factor γ as we can see from Figure for $M = 5$ and $M = 10$

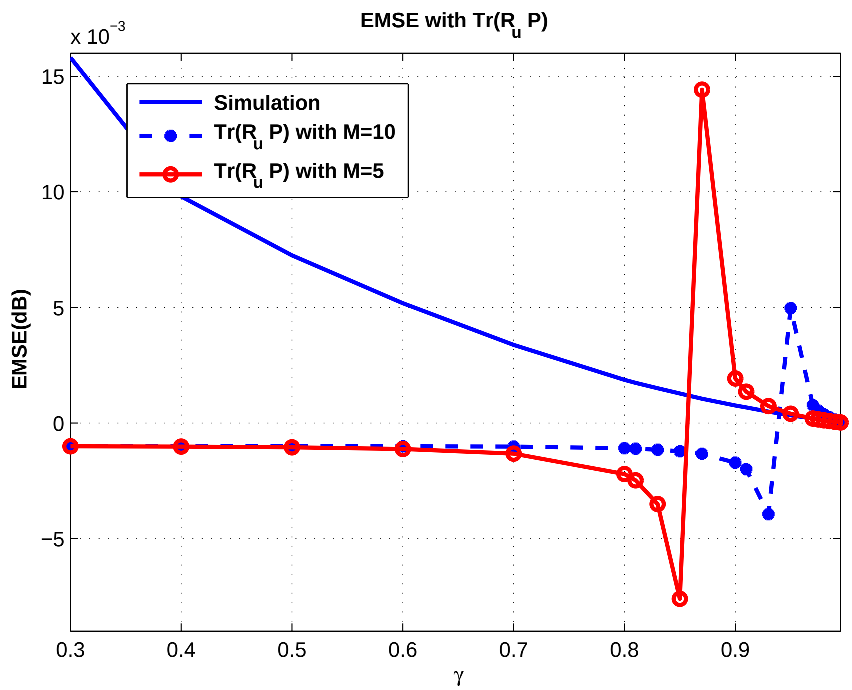

Figure 5.1: EMSE using $Tr(\boldsymbol{R_u P})$ with $M = 10, 5$ Vs γ

To overcome this unrealistic values for the EMSE we can start by multiplying equation (5.12) from left and right by $\boldsymbol{u}_i$ and $\boldsymbol{u}_i^*$ respectively. This yields after taking the expectations of both sides

$$E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} = \gamma^{-1}E\left[\boldsymbol{u}_i\boldsymbol{P}_{i-1}\boldsymbol{u}_i^* - \frac{\boldsymbol{u}_i\boldsymbol{P}_{i-1}\boldsymbol{u}_i^*\boldsymbol{u}_i\boldsymbol{P}_{i-1}\boldsymbol{u}_i^*}{\gamma + \boldsymbol{u}_i\boldsymbol{P}_{i-1}\boldsymbol{u}_i^*}\right] \tag{5.25}$$

$$= \gamma^{-1}E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}} - \gamma^{-1}E\left[\frac{\left(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}}\right)^2}{\gamma + \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}}}\right] \tag{5.26}$$

$$= \gamma^{-1}E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}} - \gamma^{-1}E\left[\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}} - \frac{\gamma\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}}}{\gamma + \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}}}\right] \tag{5.27}$$

$$= E\left[\frac{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}}}{\gamma + \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}}}\right] \tag{5.28}$$

$$= 1 - \gamma E\left[\frac{1}{\gamma + \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_{i-1}}}\right] \tag{5.29}$$

Now, at steady state (5.25) can be rewritten as

$$\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} = 1 - \gamma E\left[\frac{1}{\gamma + \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}}\right] \tag{5.30}$$

where $\boldsymbol{P} = \lim_{i\to\infty}\boldsymbol{P}_{i-1}$ which will be assumed in this work to be known. Note that the right hand side of (5.30) follows the fact that $\boldsymbol{P}_{i-1}$ and $\boldsymbol{u}_i$ are independent.

Substituting (5.30) in (5.21) yields

$$\lim_{i\to\infty} E|e_a(i)|^2 = \frac{\sigma_v^2\left(1 - E\left[\frac{\gamma}{\gamma+\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}}\right]\right)}{1 + E\left[\frac{\gamma}{\gamma+\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}}\right]} \tag{5.31}$$

Now the RHS of (5.31) is always positive as the both numerator and denominator are always positive in contrast to the result of [1] which gives negative value of

EMSE for $(1-\gamma)M$ greater than 2. Assuming that $\mathbf{P}$ is available, the EMSE calculation boils down to evaluating the moment of random variable Z defined as

$$Z \triangleq \frac{1}{\gamma + \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}} \tag{5.32}$$

5.4 Evaluating the CDF and Moment of Z

In this work, a correlated circular complex Gaussian input is consider, that is, $\boldsymbol{u}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{R})$ such that $\{\boldsymbol{u}_i\}$ are i.i.d. Now, since $\boldsymbol{P}_{i-1}$ is a function of $\{\boldsymbol{u}_j\}_{j=0}^{j=i-1}$, it follows that $\boldsymbol{u}_i$ and $\boldsymbol{P}_{i-1}$ are independent. Thus, as explained above, it is possible to write

$$\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} = 1 - E\left[\frac{\gamma}{\gamma + \|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}}\right] \tag{5.33}$$

Let $\bar{\boldsymbol{u}}_i$ be the whitened version of $\boldsymbol{u}_i$, that is[1], $\bar{\boldsymbol{u}}_i = \boldsymbol{u}_i\boldsymbol{R}^{-\frac{1}{2}}$. The random variable Z can be written as

$$Z = \frac{1}{\gamma + \|\bar{\boldsymbol{u}}_i\|^2_{\boldsymbol{R}^{\frac{1}{2}}\boldsymbol{P}\boldsymbol{R}^{\frac{H}{2}}}} = \frac{1}{\gamma + \|\bar{\boldsymbol{u}}_i\|^2_{\boldsymbol{A}}} \tag{5.34}$$

[1] where $\boldsymbol{R}^{\frac{H}{2}}$ and $\boldsymbol{R}^{-\frac{1}{2}}$ are short notations for $(\boldsymbol{R}^{\frac{1}{2}})^H$ and $(\boldsymbol{R}^{\frac{1}{2}})^{-1}$, respectively

where $\boldsymbol{A} = \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{P}\boldsymbol{R}^{\frac{H}{2}}$. Now, if $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^H$ denote the eigenvalues decomposition of $\boldsymbol{A}$ and $\boldsymbol{F} = diag\{f_1, f_2, \ldots, f_M\}$, then

$$Z = \frac{1}{\gamma + \|\bar{\boldsymbol{u}}_i\|^2_{\boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^H}} = \frac{1}{\gamma + \|\tilde{\boldsymbol{u}}_i\|^2_{\boldsymbol{F}}} \tag{5.35}$$

where $\tilde{\boldsymbol{u}}_i = \bar{\boldsymbol{u}}_i\boldsymbol{U}$ which is a white Gaussian vector and $\boldsymbol{F} = diag\{f_1, f_2, \ldots, f_M\}$. In this approach, the moment of Z can be evaluated from its CDF which will be derived in the following section.

5.4.1 CDF of the Random Variable Z

By using the definition of random variable Z, its CDF can be formulated as

$$\begin{aligned} F_Z(z) &= Pr\{Z \leq z\} \\ &= Pr(z\gamma + z\|\tilde{\boldsymbol{u}}_i\|^2_{\boldsymbol{F}} - 1 \geq 0) \end{aligned} \tag{5.36}$$

which can be set up as

$$F_Z(z) = \int_{-\infty}^{\infty} p(\tilde{\boldsymbol{u}})\ \mathbf{step}(z\gamma + z\|\tilde{\boldsymbol{u}}_i\|^2_{\boldsymbol{F}} - 1)\mathbf{d}\tilde{\boldsymbol{u}} \tag{5.37}$$

where$p(\tilde{\boldsymbol{u}})$ is the pdf of $\tilde{\boldsymbol{u}}$ and for M-dimensional circular Gaussian regressor with an identity covariance matrix it will be

$$p(\tilde{\boldsymbol{u}}) = \frac{1}{\pi^M} e^{-\|\tilde{\boldsymbol{u}}\|^2} \tag{5.38}$$

and $\mathbf{step}(x)$ is the unit step function defined as

$$\mathbf{step}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{x(jw+\beta)}}{jw+\beta} \mathbf{d}w \tag{5.39}$$

Substituting (5.38) and (5.39) into equation (5.37) yields the following integral

$$\begin{aligned} F_Z(z) &= \frac{1}{2\pi^{M+1}} \int_{-\infty}^{\infty} e^{-\|\tilde{\boldsymbol{u}}\|^2} \\ &\qquad \times \int_{-\infty}^{\infty} \frac{e^{\left(z\gamma + z\|\tilde{\mathbf{u}}_i\|_{\boldsymbol{F}}^2 - 1\right)(jw+\beta)}}{(jw+\beta)} dw \; d\tilde{\mathbf{u}} \\ &= \frac{1}{2\pi^{M+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\tilde{\mathbf{u}}\left(I - z\boldsymbol{F}(jw+\beta)\right)\tilde{\mathbf{u}}^*} d\tilde{\mathbf{u}} \\ &\qquad\qquad \times \frac{e^{(z\gamma-1)(jw+\beta)}}{(jw+\beta)} dw \end{aligned} \tag{5.40}$$

The inner integral is nothing but the Gaussian integral. Thus, intuition suggest that (see [42] for a formal proof)

$$\frac{1}{\pi^M} \int_{-\infty}^{\infty} e^{-\tilde{\mathbf{u}}\left(I - z\boldsymbol{F}(jw+\beta)\right)\tilde{\mathbf{u}}^*} \mathbf{d}\tilde{\mathbf{u}} = \frac{1}{I - z\boldsymbol{F}(jw+\beta)|} \tag{5.41}$$

Eventually the CDF of Z is reduced to following integral

$$F_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{(z\gamma-1)(jw+\beta)}}{|I - y\boldsymbol{F}(jw+\beta)|(jw+\beta)} \mathbf{d}w \tag{5.42}$$

To evaluate this integral, the fraction that appears above must be expanded as a partial fraction expansion as the follows

$$\begin{aligned} \frac{1}{|\boldsymbol{I} - z\boldsymbol{F}(jw+\beta)|(jw+\beta)} \\ &= \frac{A_0}{(jw+\beta)} \\ &+ \sum_{m=1}^{M} \frac{A_m}{[1 - zf_m(jw+\beta)]} \end{aligned} \tag{5.43}$$

where the constants A_0, A_k are given by

$$A_0 = 1 \tag{5.44}$$

$$A_m = \frac{f_m z}{\prod_{\substack{i=1 \\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} \tag{5.45}$$

By substituting (5.43) into (5.42), the integral in (5.42) is decomposed into the sum of $M+1$ integrals as

$$\begin{aligned} F_Z(z) = \frac{A_0}{2\pi} \int_{-\infty}^{\infty} \frac{e^{(z\gamma-1)(jw+\beta)}}{(jw+\beta)} dw \\ + \sum_{m=1}^{M} \frac{A_m}{2\pi f_m} \int_{-\infty}^{\infty} \frac{e^{(z\gamma-1)(jw+\beta)}}{[\frac{1}{zf_m} - (jw+\beta)]} dw \end{aligned} \tag{5.46}$$

In evaluating these integrals, the following two formulas from [43] will be used

$$\int_{-\infty}^{\infty} (\beta + ix)^{-\nu} e^{-iyx} \mathbf{d}x = \frac{2\pi(-y)^{\nu-1} e^{\beta y}}{\Gamma(\nu)} \mathbf{step}(-y) \tag{5.47}$$

and

$$\int_{-\infty}^{\infty} (\beta - ix)^{-\nu} e^{-iyx} \mathbf{d}x = \frac{2\pi y^{\nu-1} e^{-\beta y}}{\Gamma(\nu)} \mathbf{step}(y) \tag{5.48}$$

Then the CDF of Z can be expressed in closed form as

$$F_Z(z) = \sum_{m=1}^{M} \frac{e^{\frac{-(1-z\gamma)}{zf_m}}}{\prod_{\substack{i=1\\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} [\mathbf{step}(z) - \mathbf{step}(z\gamma - 1)] \tag{5.49}$$

Figure 5.2 shows the empirical and analytical CDF of the random variable Z. The figure shows excellent match between the analytical expression and the simulated CDF.

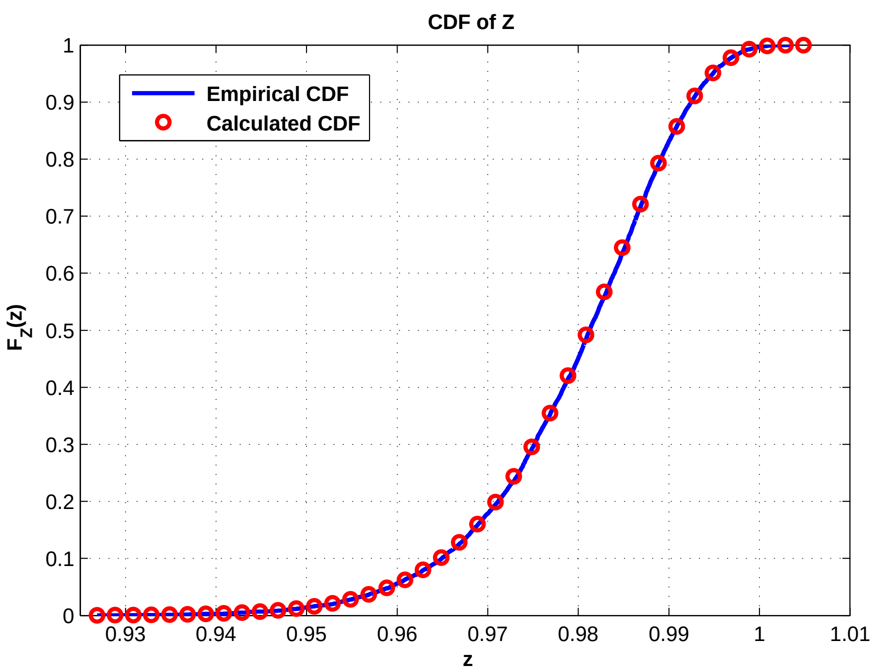

Figure 5.2: Empirical CDF Vs calculated CDF of Z

5.4.2 Moment of the Random Variable Z

Since the random variable Z is always positive, its first moment can be expressed in terms of its CDF as

$$E[Z] = \int_{-\infty}^{\infty} (1 - F_Z(z))dy \tag{5.50}$$

Now, from the definition of Z in (5.32), it can be shown that the actual support of this random variable is $0 \leq z \leq \frac{1}{\gamma}$. Thus, the integration in (5.53) can be

$$\begin{aligned} E[Z] &= \int_{0}^{\frac{1}{\gamma}} (1 - F_Z(z))dz && (5.51) \\ &= \int_{0}^{\frac{1}{\gamma}} \left[1 - \sum_{m=1}^{M} \frac{e^{\frac{-(1-z\gamma)}{zf_m}}}{\prod_{\substack{i=1 \\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} \right] dz && (5.52) \\ &= \frac{1}{\gamma} - \sum_{m=1}^{M} \left[\frac{1}{\prod_{\substack{i=1 \\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} \int_{0}^{\frac{1}{\gamma}} e^{\frac{-(1-z\gamma)}{zf_m}} dz \right] && (5.53) \\ &= \frac{1}{\gamma} - \sum_{m=1}^{M} \left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m}) e^{\frac{\gamma}{f_m}}}{\gamma \prod_{\substack{i=1 \\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} \right] && (5.54) \end{aligned}$$

where $\mathbb{E}_n(x) = \int_{1}^{\infty} \frac{e^{-xt}}{t^n} dt$ is the exponential integral function [43].

Figure 5.3 shows a comparison between the mean of the random variable Z obtained in (5.54) and the simulated one.

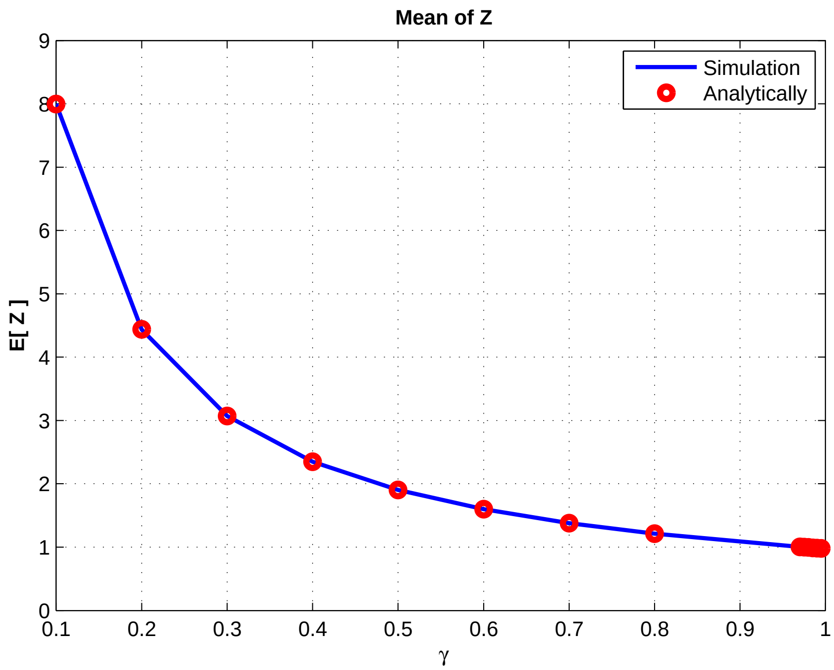

Figure 5.3: Calculated and simulated mean of Z Vs γ

5.4.3 The Steady State Value of the Moment $E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}$ and EMSE of the RLS

After substituting the $E[Z]$ from (5.54) into (5.30), the required moment $\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}$ will be

$$\begin{aligned}\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} &= 1-\gamma\left[\frac{1}{\gamma}-\sum_{m=1}^{M}\left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m})e^{\frac{\gamma}{f_m}}}{\gamma\prod_{\substack{i=1\\ \neq m}}^{M}[1-\frac{f_i}{f_m}]}\right]\right] \\ &= \gamma\sum_{m=1}^{M}\left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m})e^{\frac{\gamma}{f_m}}}{\gamma\prod_{\substack{i=1\\ \neq m}}^{M}[1-\frac{f_i}{f_m}]}\right]\end{aligned} \tag{5.55}$$

Finally, the steady-state value of EMSE for the RLS algorithm can be evaluated after substituting the above moment in (5.21).

$$\lim_{i\to\infty} E|e_a(i)|^2 = \frac{\sigma_v^2\gamma \sum_{m=1}^{M}\left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m})e^{\frac{\gamma}{f_m}}}{\gamma\prod_{\substack{i=1\\ \neq m}}^{M}[1-\frac{f_i}{f_m}]}\right]}{2-\gamma\sum_{m=1}^{M}\left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m})e^{\frac{\gamma}{f_m}}}{\gamma\prod_{\substack{i=1\\ \neq m}}^{M}[1-\frac{f_i}{f_m}]}\right]} \tag{5.56}$$

5.5 Steady State Value of $\boldsymbol{P}_i$

At steady state (i.e. $i \to \infty$)

$$\boldsymbol{P}_i = \boldsymbol{P}_{i-1} = \boldsymbol{P} \tag{5.57}$$

using this in equation (5.12) yields

$$P = \gamma^{-1}\left[P - \frac{P\boldsymbol{u}_i^*\boldsymbol{u}_iP}{\gamma + \boldsymbol{u}_iP\boldsymbol{u}_i^*}\right] \tag{5.58}$$

$$= \gamma^{-1}\left[P - \frac{P\boldsymbol{u}_i^*\boldsymbol{u}_iP}{\gamma + \|\boldsymbol{u}_i\|_P}\right] \tag{5.59}$$

Since the matrix $\boldsymbol{P}_i$ is positive definite matrix [1, p. 287] also its steady sate $\boldsymbol{P}$ will be a positive definite. Then, $\boldsymbol{P}$ can be written as $\boldsymbol{P} = \boldsymbol{P}^{\frac{H}{2}}\boldsymbol{P}^{\frac{1}{2}} = \boldsymbol{P}^{\frac{1}{2}}\boldsymbol{P}^{\frac{H}{2}}$ and equation (5.59) with this decomposition will be

$$(1-\gamma)\boldsymbol{P}^{\frac{H}{2}}\boldsymbol{P}^{\frac{1}{2}} = \mathrm{E}\left[\frac{\boldsymbol{P}^{\frac{H}{2}}\boldsymbol{P}^{\frac{1}{2}}\boldsymbol{u}_i^*\boldsymbol{u}_i\boldsymbol{P}^{\frac{H}{2}}\boldsymbol{P}^{\frac{1}{2}}}{\gamma+\|\boldsymbol{u}_i\|_{\boldsymbol{P}}^2}\right] \tag{5.60}$$

$$(1-\gamma)I = \mathrm{E}\left[\frac{\boldsymbol{P}^{\frac{1}{2}}\boldsymbol{u}_i^*\boldsymbol{u}_i\boldsymbol{P}^{\frac{H}{2}}}{\gamma+\|\boldsymbol{u}_i\|_{\boldsymbol{P}}^2}\right] \tag{5.61}$$

$$= \mathrm{E}\left[\frac{\boldsymbol{F}^{\frac{1}{2}}U^H\bar{\boldsymbol{u}}_i^*\bar{\boldsymbol{u}}_iU\boldsymbol{F}^{\frac{H}{2}}}{\gamma+\|\bar{\boldsymbol{u}}_i\|_{U\boldsymbol{F}U^H}^2}\right] \tag{5.62}$$

$$= \mathrm{E}\left[\frac{\boldsymbol{F}^{\frac{1}{2}}\tilde{\boldsymbol{u}}_i^*\tilde{\boldsymbol{u}}_i\boldsymbol{F}^{\frac{H}{2}}}{\gamma+\|\tilde{\boldsymbol{u}}_i\|_{\boldsymbol{F}}^2}\right] \tag{5.63}$$

where $\tilde{\boldsymbol{u}}_i \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$. Going from (5.61) to (5.62) done by employing the following decomposition

$$R^{\frac{1}{2}}\boldsymbol{P}_{i-1}R^{\frac{H}{2}} = U_{i-1}\boldsymbol{F}_{i-1}U_{i-1}^* \tag{5.64}$$

$$\boldsymbol{P}^{\frac{1}{2}}R^{\frac{H}{2}} = \boldsymbol{F}^{\frac{1}{2}}U^H \tag{5.65}$$

$$R^{\frac{1}{2}}\boldsymbol{P}^{\frac{H}{2}} = U\boldsymbol{F}^{\frac{H}{2}} \tag{5.66}$$

Multiplying equation (5.63) by $\boldsymbol{F}^{\frac{H}{2}}$ from left and $\boldsymbol{F}^{\frac{1}{2}}$ from right, the result will be

$$(1-\gamma)\boldsymbol{F} = \mathrm{E}\left[\frac{\boldsymbol{F}\tilde{\boldsymbol{u}}_i^*\tilde{\boldsymbol{u}}_i\boldsymbol{F}}{\gamma+\|\tilde{\boldsymbol{u}}_i\|_{\boldsymbol{F}}^2}\right] \tag{5.67}$$

or

$$\mathrm{E}\left[\frac{\boldsymbol{F}\tilde{\boldsymbol{u}}_i^*\tilde{\boldsymbol{u}}_i}{\gamma+\|\tilde{\boldsymbol{u}}_i\|_{\boldsymbol{F}}^2}\right] = (1-\gamma)\boldsymbol{I} \tag{5.68}$$

Now defining the following moment matrix $\mathbb{E} = \mathrm{E}\left[\frac{\boldsymbol{F}\tilde{\boldsymbol{u}}_i^*\tilde{\boldsymbol{u}}_i}{\gamma+\|\tilde{\boldsymbol{u}}_i\|_{\boldsymbol{F}}^2}\right]$. For this moment matrix the off-diagonal entries given by $\mathbb{E}_{kk'} = \mathbb{E}\left[\frac{f_k\tilde{\boldsymbol{u}}_i(k)^*\tilde{\boldsymbol{u}}_i(k')}{\gamma+\|\tilde{\boldsymbol{u}}_i\|_{\boldsymbol{F}}^2}\right]$ are zeros because $\mathbb{E}_{kk'}$ is an odd function of $\tilde{\boldsymbol{u}}_i(k)$ which has a symmetric probability density function (pdf) and independent of the rest of the elements of $\tilde{\boldsymbol{u}}_i$. So the moment matrix $\mathbb{E}$ is diagonal matrix.

The k^{th} entry in the main diagonal of $\mathbb{E}$ is

$$E\left[\frac{f_k|\tilde{\boldsymbol{u}}(k)|^2}{\gamma+\|\tilde{\boldsymbol{u}}_i\|_{\boldsymbol{F}}^2}\right] = E\left[\frac{f_k|\tilde{\boldsymbol{u}}(k)|^2}{\gamma+\sum_{j=1}^{M} f_j|\tilde{\boldsymbol{u}}(j)|^2}\right] = E\left[Y_k\right] \tag{5.69}$$

where Y_k is given by

$$Y_k = \frac{f_k|\tilde{\boldsymbol{u}}(k)|^2}{\gamma+\|\tilde{\boldsymbol{u}}\|_F^2} \tag{5.70}$$

It is clear from (5.70) that the first moment of Y_k is a function of the diagonal entries of the matrix $\boldsymbol{F}$. These entries of $\boldsymbol{F}$ can be found by solving an M nonlinear equations as

$$\begin{aligned} E[Y_1] &= (1-\gamma) \\ E[Y_2] &= (1-\gamma) \\ &\vdots \\ E[Y_M] &= (1-\gamma) \end{aligned} \tag{5.71}$$

5.5.1 CDF of the Random Variable Y_k

Following the same steps that are used for the random variable Z in the previous subsection. The CDF of the random variable Y_k is given by

$$\begin{aligned} F_{Y_k}(y) &= Pr(Y_k \leq y) \\ &= \int_{-\infty}^{\infty} p(\tilde{\boldsymbol{u}})\ \mathbf{step}(y\gamma + y\|\tilde{u}\|_F^2 - f_k|\tilde{\boldsymbol{u}}(k)|^2)\mathbf{d}\tilde{\boldsymbol{u}} \\ &= \int_{-\infty}^{\infty} p(\tilde{\boldsymbol{u}})\ \mathbf{step}(y\gamma + y\|\tilde{\boldsymbol{u}}\|_F^2 - \|\tilde{\boldsymbol{u}}\|_{B_k}^2)\mathbf{d}\tilde{\boldsymbol{u}} \end{aligned} \tag{5.72}$$

where $B_k = diag\{0\ ,\ \ldots\ ,\ f_k\ ,\ \ldots\ ,\ 0\}$.

Replacing $p(\tilde{u})$ and $\mathbf{step}(x)$ by their values in (5.38) and (5.39) respectively yields to the following integral

$$\begin{aligned} F_{Y_k}(y) &= \frac{1}{2\pi^{M+1}} \int_{-\infty}^{\infty} e^{-\|\tilde{\boldsymbol{u}}\|^2} \\ &\quad \times \int_{-\infty}^{\infty} \frac{e^{\left(y\gamma + y\|\tilde{\boldsymbol{u}}\|_F^2 - \|\tilde{\boldsymbol{u}}\|_{B_k}^2\right)(jw+\beta)}}{(jw+\beta)} dw\ d\tilde{\boldsymbol{u}} \\ &= \frac{1}{2\pi^{M+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\tilde{\boldsymbol{u}}\left(I-(yF-B_k)(jw+\beta)\right)\tilde{\boldsymbol{u}}^*} d\tilde{\boldsymbol{u}} \\ &\quad \times \frac{e^{y\gamma(jw+\beta)}}{(jw+\beta)} dw \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{y\gamma(jw+\beta)}}{|I-(yF-B_k)(jw+\beta)|(jw+\beta)} \mathbf{d}w \end{aligned} \tag{5.73}$$

The partial expansion of the fraction that appears above is

$$\begin{aligned}
\frac{1}{|I-(yF-B_k)(jw+\beta)|(jw+\beta)} \\
&= \frac{A_0}{(jw+\beta)} \\
&+ \frac{A_k}{[1-f_k(y-1)(jw+\beta)]} \\
&+ \sum_{\substack{m=1,\\ \neq k}}^{M} \frac{A_m}{[1-yf_m(jw+\beta)]}
\end{aligned} \tag{5.74}$$

where the constants A_0, A_k and A_m ($m = 1\ ,\ 2\ ,\ \ldots\ ,\ M,\ m \neq k$) are given by

$$A_0 = 1 \tag{5.75}$$

$$A_k = \frac{f_k(1-y)}{\prod_{\substack{m=1\\ \neq k}}^{M}[1-\frac{yf_m}{f_k(1-y)}]} \tag{5.76}$$

$$A_m = \frac{f_m y}{\prod_{\substack{i=1\\ \neq k \neq m}}^{M}[1-\frac{f_i}{f_m}]} \tag{5.77}$$

By substituting (5.74) into (5.73), the integral in (5.73) is decomposed into the sum of $M+1$ integrals as

$$\begin{aligned}
F_{Y_k}(y) = \frac{A_0}{2\pi} \int_{-\infty}^{\infty} \frac{e^{y\gamma(jw+\beta)}}{(jw+\beta)} dw \\
+ \frac{A_k}{2\pi f_k(1-y)} \int_{-\infty}^{\infty} \frac{e^{y\gamma(jw+\beta)}}{[\frac{1}{f_k(1-y)}+(jw+\beta)]} dw \\
+ \sum_{\substack{m=1,\\ \neq k}}^{M} \frac{A_m}{2\pi f_m} \int_{-\infty}^{\infty} \frac{e^{y\gamma(jw+\beta)}}{[\frac{1}{yf_m}-(jw+\beta)]} dw
\end{aligned} \tag{5.78}$$

Ultimately the CDF of Y_k can be expressed in closed form as

$$\begin{aligned} F_{Y_k}(y) &= \mathbf{step}(y) + \frac{A_k e^{\frac{-y\gamma}{f_k(1-y)}}}{f_k(y-1)}[\mathbf{step}(y) - \mathbf{step}(1-y)] \\ &= \mathbf{step}(y) + \frac{e^{\frac{-y\gamma}{f_k(1-y)}}}{\prod_{\substack{m=1\\ \neq k}}^{M}[1 - \frac{yf_m}{f_k(1-y)}]}[\mathbf{step}(y) - \mathbf{step}(y-1)] \end{aligned} \tag{5.79}$$

5.5.2 Moment of the Random Variable Y_k

The random variable Y_k is positive, its first moment can be expressed in terms of the CDF using integration by parts as

$$E[Y_k] = \int_{-\infty}^{\infty} (1 - F_{Y_k}(y))dy \tag{5.80}$$

The support of the Y_k is $0 \leq y \leq 1$ and the integration in (5.80) can be rewritten as

$$E[Y_k] = \int_{-\infty}^{\infty} \frac{e^{\frac{-y\gamma}{f_k(1-y1)}}}{\prod_{\substack{m=1\\ \neq k}}^{M}[1 - \frac{yf_m}{f_k(1-y)}]}[\mathbf{step}(y) - \mathbf{step}(y-1)]dy \tag{5.81}$$

$$= \int_{0}^{1} \frac{e^{\frac{-y\gamma}{f_k(1-y)}}}{\prod_{\substack{m=1\\ \neq k}}^{M}[1 - \frac{yf_m}{f_k(1-y)}]}dy \tag{5.82}$$

Let $\nu = \frac{y}{y-1}$, then the above integration will be

$$E[Y_k] = \int_{0}^{\infty} \frac{e^{\frac{-\nu\gamma}{f_k}}}{\prod_{\substack{m=1\\ \neq k}}^{M}[1 + \frac{\nu f_m}{f_k}]} \frac{d\nu}{(\nu+1)^2} \tag{5.83}$$

To proceed the inner fraction will written in partial fraction form as

$$\begin{aligned}\frac{1}{(\nu+1)^2\prod_{\substack{m=1\\ \neq k}}^{M}[1+\frac{\nu f_m}{f_k}]} &= \frac{C_1}{(v+1)}+\frac{C_2}{(v+1)^2} \\ &+\sum_{\substack{m=1\\ \neq k}}^{M}\frac{C_m}{[1+\frac{v f_m}{f_k}]}\end{aligned} \tag{5.84}$$

where the coefficients C_1, C_2 and C_m are given by

$$\begin{aligned} C_1 &= \frac{-\left[\frac{d}{dv}\prod_{\substack{m=1\\ \neq k}}^{M}(1+v\frac{f_m}{f_k})\right]_{v=-1}}{\left[\prod_{\substack{m=1\\ \neq k}}^{M}(1-\frac{f_m}{f_k})\right]^2} \\ C_2 &= \frac{1}{\prod_{\substack{m=1\\ \neq k}}^{M}(1-\frac{f_m}{f_k})} \\ C_m &= \frac{1}{(1-\frac{f_k}{f_m})^2\prod_{\substack{l=1\\ \neq m,k}}^{M}(1-\frac{f_l}{f_m})}\end{aligned} \tag{5.85}$$

Using these coefficients in (5.83) to find the moment $E[Y_k]$ as

$$\begin{aligned} E[Y_k] &= C_1\int_0^1\frac{e^{-\frac{v\gamma}{f_k}}}{(v+1)}dv + C_2\int_0^1\frac{e^{-\frac{v\gamma}{f_k}}}{(v+1)^2}dv \\ &+\sum_{\substack{m=1\\ \neq k}}^{M}\int_0^1\frac{C_m\ e^{-\frac{v\gamma}{f_k}}}{[1+v\frac{f_m}{f_k}]}dv \\ &= C_1\ e^{\frac{\gamma}{f_k}}\ \mathbb{E}_1\left(\frac{\gamma}{f_k}\right)+C_2\ e^{\frac{\gamma}{f_k}}\ \mathbb{E}_2\left(\frac{\gamma}{f_k}\right) \\ &+\sum_{\substack{m=1\\ \neq k}}^{M}\frac{f_k}{f_m}\ C_M\ e^{\frac{\gamma}{f_m}}\ \ \mathbb{E}_1\left(\frac{\gamma}{f_m}\right)\end{aligned} \tag{5.86}$$

We need to solve the M nonlinear equations in (5.71) to find the entries of the matrix $\boldsymbol{F}$.

5.6 Tracking Analysis

In non-stationary environment, the weight vector $\boldsymbol{w}^o$ will be time dependent and it will have the following model

$$\boldsymbol{w}_i^o = \boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i \tag{5.87}$$

where $\boldsymbol{q}_i$ is an i.i.d., zero mean random sequence with covariance matrix

$$E[\boldsymbol{q}_i \boldsymbol{q}_i^*] = \boldsymbol{Q} \tag{5.88}$$

also it can be shown that

$$E\boldsymbol{w}_i^o = E\boldsymbol{w}_{i-1}^o = \boldsymbol{w}^o \tag{5.89}$$

The update recursion in (5.5) can be rewritten in terms of weight error vector $\tilde{\boldsymbol{w}}_i = \boldsymbol{w}_i^o - \boldsymbol{w}_i$ as

$$\boldsymbol{w}_i^o - \boldsymbol{w}_i = (\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}) - \boldsymbol{P}_i \boldsymbol{u}_i^* e(i) \tag{5.90}$$

Following the same arguments that presented in Section 5.3 will lead to the following **The Energy Conservation Relation** [1] given by

$$\|\boldsymbol{w}_i^o - \boldsymbol{w}_i\|^2_{\boldsymbol{P}_i^{-1}} + \frac{|e_a(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} = \|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}} + \frac{|e_p(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.91}$$

or

$$E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} + E\frac{|e_a(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} = E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}} + E\|\boldsymbol{q}_i\|^2_{\boldsymbol{P}_i^{-1}} + E\frac{|e_p(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.92}$$

where $e_a(i) = \boldsymbol{u}_i(\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1})$, $e_p(i) = \boldsymbol{u}_i(\boldsymbol{w}_i^o - \boldsymbol{w}_i)$.

The variance relation for the RLS filter in the non-stationary environment will be

$$\sigma_v^2 E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} + E\big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \cdot |e_a(i)|^2\big) + E\|\boldsymbol{q}_i\|^2_{\boldsymbol{P}_i^{-1}} = 2E|e_a(i)|^2 \tag{5.93}$$

Employing the separation condition to separate the following moment $E\big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \cdot |e_a(i)|^2\big)$ as

$$E\big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \cdot |e_a(i)|^2\big) \approx E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \; E|e_a(i)|^2 \tag{5.94}$$

substituting this in (5.93) and solving for the $E|e_a(i)|^2$ yields

$$\lim_{i\to\infty} E|e_a(i)|^2 = \frac{\sigma_v^2 \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} + E\|\boldsymbol{q}_i\|^2_{\boldsymbol{P}_i^{-1}}}{2 - \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.95}$$

the sequence $\boldsymbol{q}_i$ is independent of all regressors and it will be also independent of $\boldsymbol{P}_i^{-1}$, thus

$$E\|\boldsymbol{q}_i\|^2_{\boldsymbol{P}_i^{-1}} = E\|\boldsymbol{q}_i\|^2_{\boldsymbol{P}^{-1}} = Tr(\mathbf{Q}\boldsymbol{P}^{-1}) = \frac{1}{(1-\gamma)} Tr(\mathbf{Q}\boldsymbol{R}_{\boldsymbol{u}}) \tag{5.96}$$

using this in (5.95) yields

$$\lim_{i\to\infty} E|e_a(i)|^2 = \frac{\sigma_v^2 \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} + \frac{1}{(1-\gamma)} Tr(\mathbf{Q}\boldsymbol{R}_{\boldsymbol{u}})}{2 - \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.97}$$

It is known from (5.55) that the value of $\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}}$ is

$$\lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}} = \gamma \sum_{m=1}^{M} \left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m}) e^{\frac{\gamma}{f_m}}}{\gamma \prod_{\substack{i=1\\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} \right] \tag{5.98}$$

Substituting this in (5.97) yields

$$\lim_{i\to\infty} E|e_a(i)|^2 = \frac{\sigma_v^2 \gamma \displaystyle\sum_{m=1}^{M} \left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m}) e^{\frac{\gamma}{f_m}}}{\gamma \prod_{\substack{i=1\\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} \right] + \dfrac{1}{(1-\gamma)} Tr(\mathbf{Q}\boldsymbol{R}_{\boldsymbol{u}})}{2 - \gamma \displaystyle\sum_{m=1}^{M} \left[\frac{\mathbb{E}_2(\frac{\gamma}{f_m}) e^{\frac{\gamma}{f_m}}}{\gamma \prod_{\substack{i=1\\ \neq m}}^{M} [1 - \frac{f_i}{f_m}]} \right]} \tag{5.99}$$

5.7 Simulation Results

In simulations, the steady-state performance of the RLS algorithm is investigated for an unknown system identification with $w^o = [0.13484\ ,\ 0.26968\ ,\ 0.40452\ ,\ 0.53936\ ,\ 0.67420]^T$. The noise is zero mean i.i.d with variance $\sigma_v^2 = 0.001$. Input to the adaptive filter and to the unknown system is correlated circular complex Gaussian having correlation $\mathbf{R}(i,j) = \alpha_c^{|i-j|}$ $(0 < \alpha_c < 1)$. First, we study the steady-state performance of the RLS algorithm by evaluating the required moment $E[\|\mathbf{u}_i\|^2_{\mathbf{P}}]$ and the steady-state EMSE. This study assumes that the steady-state value of the matrix $\mathbf{P}$ is available from the simulation of actual RLS recursion. Using this $\mathbf{P}$, we first evaluate the moment $E[\|\mathbf{u}_i\|^2_{\mathbf{P}}]$ using (5.18) and compare the result with the one from simulation and the analytical one proposed in [1] (i.e., using $\mathrm{Tr}(\mathbf{RP})$) in Figure 5.4. It can be seen that the proposed moment calculation has a very good match with the simulation one as compared to the one proposed in [1]. Next, in analyzing the EMSE, we evaluate the EMSE using (5.56) with available $\mathbf{P}$ and compare its result from the EMSE results via actual RLS recursion, via analytical EMSE using moment from simulation. This comparison is reported in Figure 5.5. In Figure 5.6 we have the same curves in Figure 5.5 plus the EMSE using the moment proposed in [1]. It can be depicted from the figures that the proposed EMSE result has a good match with the simulation for larger values of γ (say $\gamma > 0.8$) but it gives a poor estimate for smaller values of γ. Same behavior is observed for the EMSE obtained via the moment from simulation. Reasons for this deviation will be reported in the next section. On the other hand, the EMSE using the moment proposed in [1] gives positive values only for larger forgetting factor $i.e., \gamma \geq 9$. This is because of the fact that the EMSE expression given in [1] becomes unrealistic (negative or infinity) for $(1-\gamma)M \geq 2$. In contrast, our approach is valid for all values of γ and M.

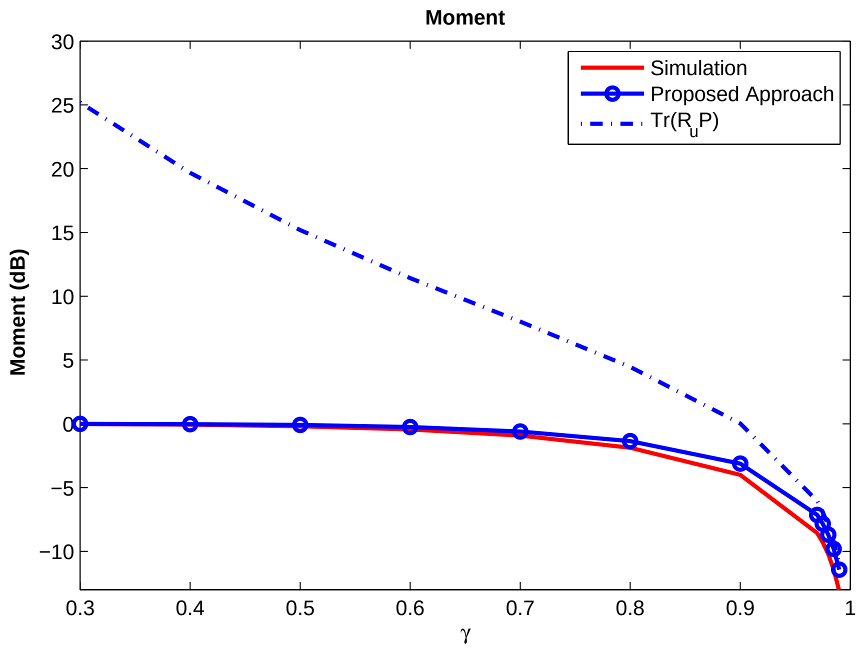

Figure 5.4: Moment value Vs γ

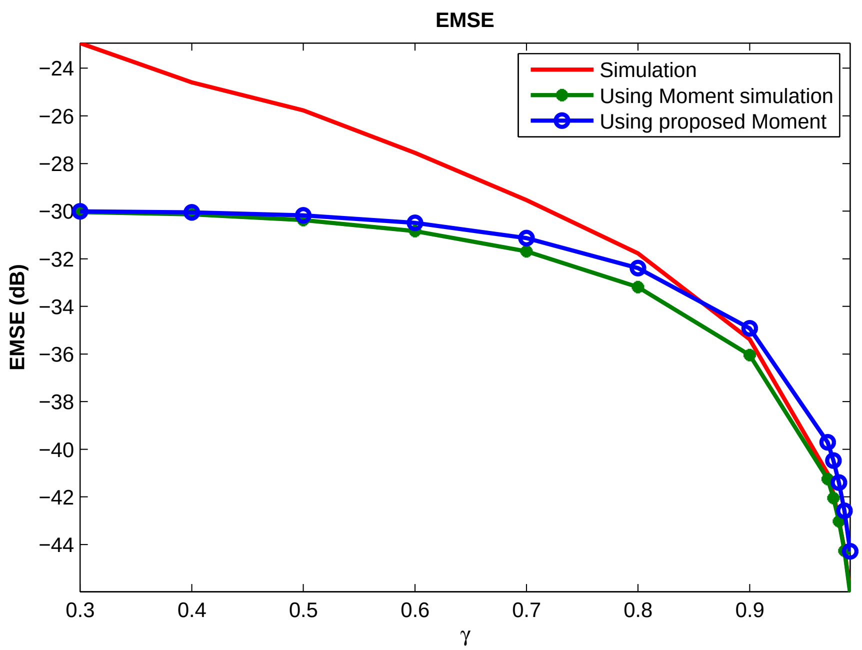

Figure 5.5: EMSE Vs γ

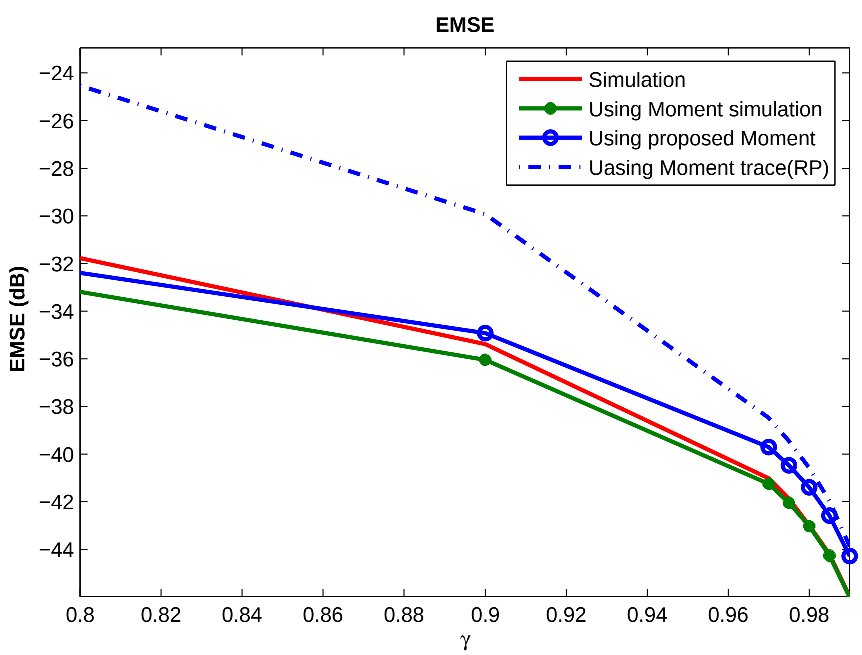

Figure 5.6: EMSE Vs γ

5.8 Improved Steady State Analysis

Recalling equation (5.17) in chapter (5)

$$\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} + \frac{|e_a(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}} + \frac{|e_p(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.100}$$

Using the following assumption in steady state in equation (5.100)

$$E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} \approx E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}} \tag{5.101}$$

will give the following variance relation.

$$\sigma_v^2 E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} + E\big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \ \cdot \ |e_a(i)|^2\big) = 2E|e_a(i)|^2 \tag{5.102}$$

But, from Figure 5.7 it can be seen that this assumption is not accurate for a wide range of the forgetting factor γ. This disparity between the RHS and LFS of (5.102) appears because of that $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}}$ is not equal to $E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}}$ in the steady state, as shown in Figure 5.8.

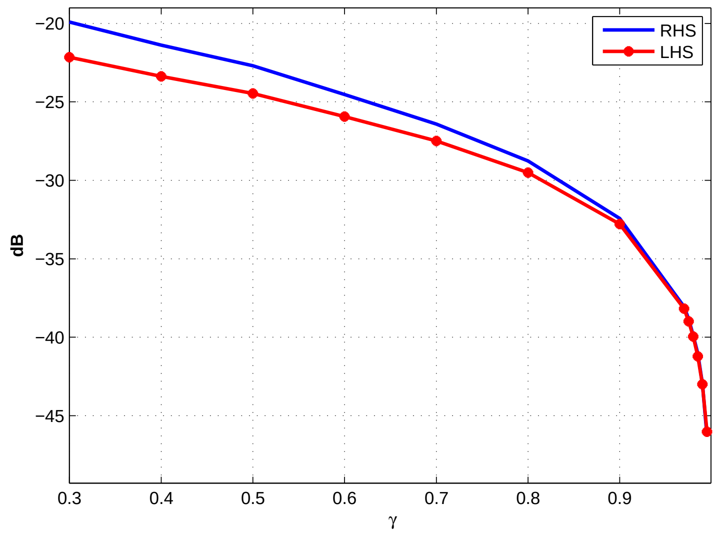

Figure 5.7: RHS and LHS of (5.102) Vs γ

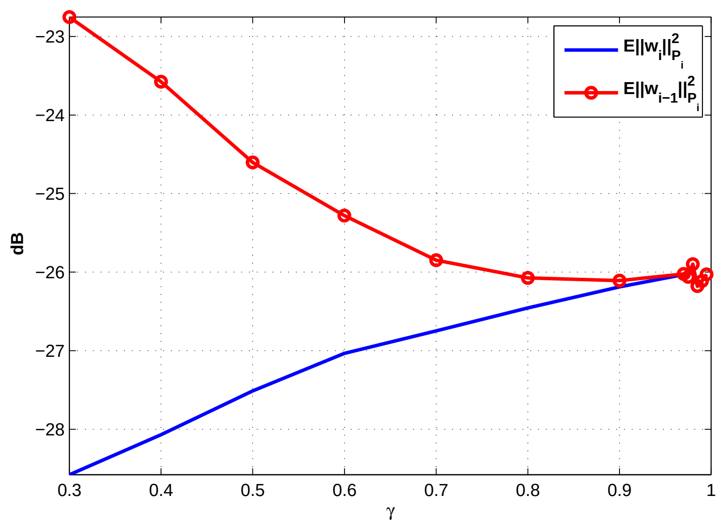

Figure 5.8: $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}}$ and $E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}}$ Vs γ

To overcome this issue it is better if the following two quantities are assumed to be equal in steady state, as we can see from Figure 5.9

$$E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_{i-1}^{-1}} \tag{5.103}$$

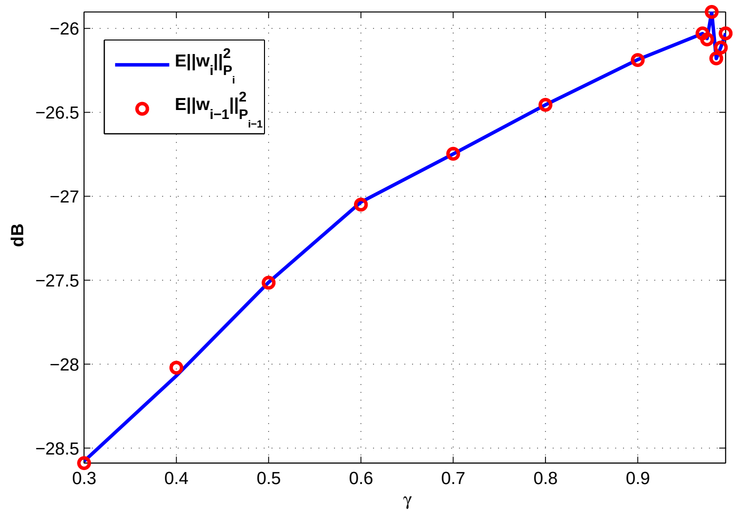

Figure 5.9: $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}}$ and $\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_{i-1}^{-1}}$ Vs γ

If the time index of the matrix $\boldsymbol{P}_i^{-1}$ in the RHS of (5.101) is changed from i to $i-1$ it will be coincide with the time index of $\tilde{\boldsymbol{w}}_{i-1}$.

From equation (5.10), the relation between $\boldsymbol{P}_i^{-1}$ and $\boldsymbol{P}_{i-1}^{-1}$ is

$$\boldsymbol{P}_i^{-1} = \gamma \boldsymbol{P}_{i-1}^{-1} + \boldsymbol{u}_i^* \boldsymbol{u}_i \tag{5.104}$$

By using this value of $\boldsymbol{P}_i^{-1}$, the moment $E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}}$ can be rewritten as

$$E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}} = E\tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{P}_i^{-1} \tilde{\boldsymbol{w}}_{i-1} \tag{5.105}$$

$$= E\tilde{\boldsymbol{w}}_{i-1}^* (\gamma \boldsymbol{P}_{i-1}^{-1} + \boldsymbol{u}_i^* \boldsymbol{u}_i) \tilde{\boldsymbol{w}}_{i-1} \tag{5.106}$$

$$= \gamma E\tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{P}_{i-1}^{-1} \tilde{\boldsymbol{w}}_{i-1} + E\tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{u}_i^* \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1} \tag{5.107}$$

$$= \gamma E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_{i-1}^{-1}} + E|e_a(i)|^2 \tag{5.108}$$

Taking the expectation of both sides in (5.100) and using this equivalent expression for the moment $E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_i^{-1}}$, yields the following result in steady state (i.e. as $i \to \infty$)

$$(1-\gamma)E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} + E\frac{|e_a(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} = E|e_a(i)|^2 + E\frac{|e_p(i)|^2}{\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.109}$$

where the fact that $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} = E\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{P}_{i-1}^{-1}}$ is steady state is used. Using (5.15) and the fact that $e(i) = e_a(i) + v(i)$ yields (as $i \to \infty$)

$$\sigma_v^2 E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} + E\big(\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} \cdot |e_a(i)|^2\big) = E|e_a(i)|^2 + (1-\gamma)E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}} \tag{5.110}$$

The plots of the RHS and the LHS of equation (5.110) are plotted in Figure 5.10.

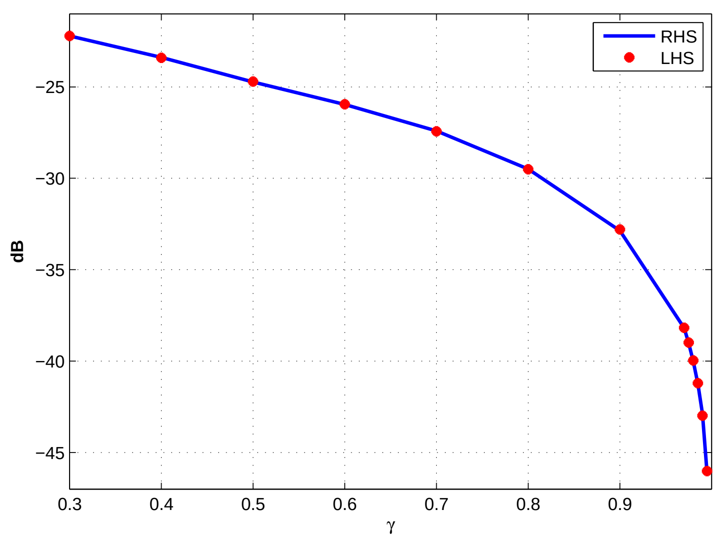

Figure 5.10: RHS and LHS of (5.110) Vs γ

In contrast to Figure 5.7 which plots the two sides of (5.102) the two sides of (5.110) coincide as we can see from the above figure.

The EMSE derived from an old variance relatione (5.102) is given by

$$E|e_a(i)|^2 = \frac{\sigma_v^2 \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}}{2 - \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.111}$$

and the EMSE derived from the new one (5.110) will be

$$E|e_a(i)|^2 = \frac{\sigma_v^2 \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i} - (1-\gamma)E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}}}{1 - \lim_{i\to\infty} E\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}} \tag{5.112}$$

In Figure 5.11, we have plots of the EMSE obtained from simulation and from equations (5.111) and (5.112).

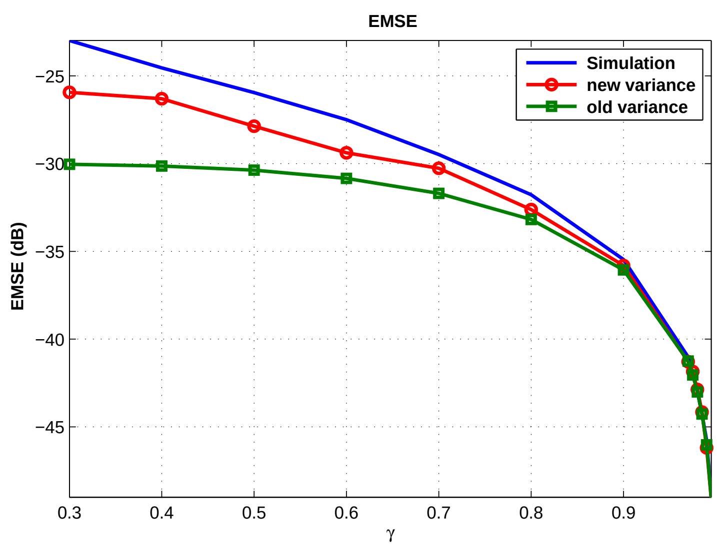

Figure 5.11: EMSE Vs γ

Using the new variance relation to study the steady state performance of the RLS filter will give better results than using an old one. This study relies on calculating the new moment $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}}$ that appears in equation (5.110). This calculation will be a part of the future work in this topic.

CHAPTER 6

THESIS CONTRIBUTIONS, CONCLUSION AND FUTURE WORK

6.1 Thesis Contributions

This work has successfully presented a majorization theory as mathematical tool to study the performance of adaptive filters. To our knowledge, this use of majorization theory in adaptive filters is the first connection between these two fields. For the Recursive Least Squares (RLS) adaptive filter, the proposed idea for calculating its steady state performance gives good results for the whole range of the forgetting factor γ, in contrast to the available results in literature.

6.2 Conclusion

In the first part of this thesis, we investigate the effect of input correlation on the performance of some well known adaptive algorithms; specifically, Steepest Decent, Newton's Method and LMS. This investigation is done by employing the majorization theory and its techniques. In majorization theory we can order between vectors and preserve this order through Schur's functions. In adaptive filters, the correlation of the input repressor can be totally described by the eigenvalues of the covariance matrix $\boldsymbol{R_u}$. By describing each input correlation scenario by the eigenvalues of its matrix (each eigenvalue $\lambda_k \in \boldsymbol{\Lambda} = \{\lambda_1\ ,\ \lambda_2\ ,\ \ldots\ ,\ \lambda_M\}$), we can order theses scenarios between the best and the worst scenarios. By testing the response of any adaptive filter by Schur's conditions, we can say which scenario will give better transient or steady state performance.

For the Steepest Descent method, we derived a condition on the step size μ for its learning curve to be Schur-convex. Also, for the Newton's method we saw that its learning curve is neither Schur-convex nor concave, because it is a function of the sum of the eigenvalues. In other words, by testing its learning curve with the best or with the worst scenarios will give the same behaviors. Finally, for the LMS filter the Schur' convexity is shown for its MSD learning curve with small step size as well as its proved also for the steady state EMSE.

In the second part, we analyze the RLS algorithm at steady state for correlated complex Gaussian input and we evaluate its EMSE by calculating the moment $E[\|\boldsymbol{u}_i\|^2_{\boldsymbol{P}_i}]$. The novelty of the work resides in the evaluation of this moment which is based on the derivation of a closed form expression for the CDF and moment of random variable of the form $\frac{1}{\gamma+\|\boldsymbol{u}\|_{\boldsymbol{D}^2}}$. Moreover, our approach employs the independence between $\boldsymbol{u}_i$ and $\boldsymbol{P}_{i-1}$ (which comes from i.i.d. nature of $\{\boldsymbol{u}_i\}$) in contrast to the existing approaches which use independence between $\boldsymbol{u}_i$ and $\boldsymbol{P}_i$ (which is not true and could give negative value of EMSE). Hence, unlike the previous work, our approach is valid for a wide range of forgetting factor γ and filter's length M. Theoretical results are validated by simulations.

6.3 Future Work

In Chapter 4, the majorization theory was applied to the Steepest Descent, Newton's Method and steady state measures of the LMS filter. Many directions of future research in this way could be investigated, such these directions are

- Applying the majorization theory to the transient behaviour of the LMS filter.
- Applying the majorization theory to the class of adaptive filters with general data non-linearity. As a special case, applying this technique for the well known NLMS algorithm.
- Applying the majorization study to a large class of adaptive filters.

The performance of the RLS filter in steady state has been improved in Chapter 5. Future work can be done in the following directions:

- Calculating the EMSE of the RLS filter from the new variance relation (5.110) by calculating the new moment $E\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{P}_i^{-1}}$.

- Performing the improved analysis for the steady state MSD of the RLS filter.

- How can this improved analysis be used to enhance the study of transient analysis for the RLS filter?

REFERENCES

[1] A. H. Sayed, *Adaptive Filters*, 1st ed. Wiley-IEEE Press, Apr. 2008.

[2] S. Qureshi, "Adaptive equalization," *IEEE Communications Magazine*, vol. 20, no. 2, p. 916, 1982.

[3] B. Widrow, J. Glover, J.R., J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, J. Eugene Dong, and R. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, p. 16921716, 1975.

[4] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Transactions on Automatic Control*, vol. 12, no. 3, p. 282287, 1967.

[5] P. M. Grant, C. F. N. C. F. N. Cowan, and P. F. P. F. Adams, *Adaptive filters / edited by C.F.N. Cowan and P.M. Grant ; with contributions from P.F. Adams ... [et al.]*. Englewood Cliffs, N.J. : Prentice-Hall, 1985, includes index.

[6] R. D. Gitlin, H. C. Meadors, and S. B. Weinstein, "The tap-leakage algorithm: An algorithm for the stable operation of a

digitally implemented, fractionally spaced adaptive equalizer," *Bell System Technical Journal*, vol. 61, no. 8, p. 18171839, 1982. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1982.tb03086.x/abstract

[7] R. Kwong and E. Johnston, "A variable step size LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, p. 16331642, 1992.

[8] S. S. Haykin, *Adaptive filter theory.* Upper Saddle River, N.J.: Prentice Hall, 2002.

[9] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, p. 561580, 1975.

[10] A. H. Sayed and M. Rupp, "Time-domain feedback analysis of adaptive gradient algorithms via the small gain theorem," vol. 2563, 1995, p. 458469. [Online]. Available: http://dx.doi.org/10.1117/12.211422

[11] N. Yousef and A. Sayed, "A feedback analysis of the tracking performance of blind adaptive equalization algorithms," in *Proceedings of the 38th IEEE Conference on Decision and Control, 1999*, vol. 1, 1999, pp. 174–179 vol.1.

[12] ——, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314–324, 2001.

[13] S. Jones, R. Cavin, and W. Reed, "Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 318–329, 1982.

[14] W. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique," *Signal Processing*, vol. 6, no. 2, pp. 113–133, Apr. 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0165168484900136

[15] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2811–2825, 1993.

[16] T. Y. Al-Naffouri and A. Sayed, "Transient analysis of adaptive filters," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01)*, vol. 6, 2001, p. 38693872 vol.6.

[17] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. Springer, Jan. 2009.

[18] J. Nocedal and S. J. Wright, *Numerical optimization.* New York: Springer, 1999.

[19] D. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 30, no. 4, p. 578586, 1982.

[20] E. Walach and B. Widrow, “The least mean fourth (LMF) adaptive algorithm and its family,” *IEEE Transactions on Information Theory*, vol. 30, no. 2, p. 275283, 1984.

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

[22] B. WIDROW, M. E. HOFF *et al.*, “Adaptive switching circuits.” 1960.

[23] A. Zerguine, C. F. N. Cowan, and M. Bettayeb, “Adaptive echo cancellation using least mean mixed-norm algorithm,” *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1340–1343, 1997.

[24] A. Zerguine, “Convergence behavior of the normalized least mean fourth algorithm,” in *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, 2000*, vol. 1, 2000, pp. 275–278 vol.1.

[25] E. Eweda, “Analysis and design of a signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated gaussian data,” *IEEE Transactions on Circuits and Systems*, vol. 37, no. 11, pp. 1367–1374, 1990.

[26] D. P. Palomar and Y. Jiang, *MIMO Transceiver Design via Majorization Theory (Foundations and Trend*. Now Publishers Inc, Jun. 2007.

[27] E. Jorswieck and H. Boche, *Majorization and Matrix-monotone Functions in Wireless Communications*. Now Publishers Inc, Jul. 2007.

[28] N. Bershad, "Behavior of the #949;-normalized LMS algorithm with gaussian inputs," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 5, pp. 636–644, 1987.

[29] V. Nascimento, "A simple model for the effect of normalization on the convergence rate of adaptive filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*, vol. 2, 2004, pp. ii–453–6 vol.2.

[30] P. An, M. Brown, and C. Harris, "On the convergence rate performance of the normalized least-mean-square adaptation," *IEEE Transactions on Neural Networks*, vol. 8, no. 5, pp. 1211–1214, 1997.

[31] T. Al-Naffouri and M. Moinuddin, "Exact performance analysis of the -NLMS algorithm for colored circular gaussian inputs," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, p. 5080 5090, Oct. 2010.

[32] E. Eleftheriou and D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 5, pp. 1097–1110, 1986.

[33] L. Guo, L. Ljung, and P. Priouret, "Tracking performance analysis of the forgetting factor RLS algorithm," in *, Proceedings of the 31st IEEE Conference on Decision and Control, 1992*, 1992, pp. 688–693 vol.1.

[34] P. Diniz and M. Siqueira, "Analysis of the QR-RLS algorithm for colored-input signals," in *, 1995 International Conference on Acoustics, Speech, and Signal Processing, 1995. ICASSP-95*, vol. 2, 1995, pp. 1428–1431 vol.2.

[35] G. Moustakides, "Study of the transient phase of the forgetting factor RLS," *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2468–2476, 1997.

[36] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, 2004.

[37] M. Chebolu and S. Jayaweera, "Energy efficiency analysis of an RLS-based adaptive signal processing algorithm for energy aware sensor networks," in *Proceedings of 2005 International Conference on Intelligent Sensing and Information Processing, 2005*, 2005, pp. 13–18.

[38] A. Vakili and B. Hassibi, "A stieltjes transform approach for analyzing the RLS adaptive filter," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, 2008, pp. 432–437.

[39] M. Pajovic and J. Preisig, "Performance analysis of the least squares based LTI channel identification algorithm using random matrix methods," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2011, pp. 516–523.

[40] G. Mateos and G. Giannakis, “Distributed recursive least-squares: Stability and performance analysis,” *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3740–3754, 2012.

[41] T. Al-Naffouri and A. Sayed, “Transient analysis of data-normalized adaptive filters,” *IEEE Transactions on Signal Processing*, vol. 51, no. 3, p. 639 652, Mar. 2003.

[42] T. Al-Naffouri and B. Hassibi, “On the distribution of indefinite quadratic forms in gaussian random variables,” in *IEEE International Symposium on Information Theory, 2009. ISIT 2009*, 2009, pp. 1744–1748.

[43] I. S. Gradshten, I. M. Ryzhik, and A. Jeffrey, *Table of integrals, series, and products.* New York: Academic Press, 1980.

Vitae

- Name: Khaled Abudlaziz Al-Hujaili
- Nationality: Saudi
- Date of Birth: 28, September 1980
- Khaled got his bachelors degree in Electrical Engineering with honors from King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia in January 2004.
- Worked as an electrical engineer at Saudi Electricity Company, Madinah from 2004 to 2006.
- Worked as graduate assistance at Al-Madinah College of Technology, Madinah from 2006 to 2010.
- Worked as graduate assistance at Taibah University, Madinah from 2010 to now.
- Completed masters degree requirement in December 2013 with a CGPA of 3.8/4.0.
- Email:*kh_alhujaili@hotmail.com* or *khujaili@taibahu.edu.sa*
- Permanent Address: P.O.Box 15400, Al-Madina Al-Monawrah 41444, Saudi Arabia.